Titre: Title:	Représentations et techniques d'analyse des traces d'écoute vidéo dans un MOOC
Auteur: Author:	Boniface Mbouzao
Date:	2020
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Mbouzao, B. (2020). Représentations et techniques d'analyse des traces d'écoute vidéo dans un MOOC [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/5537/

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Représentations et techniques d'analyse des traces d'écoute vidéo dans un MOOC

BONIFACE MBOUZAO
Département de génie informatique et génie logiciel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie informatique

Décembre 2020

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Représentations et techniques d'analyse des traces d'écoute vidéo dans un MOOC

présentée par **Boniface MBOUZAO**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Daniel ALOISE, président
Michel DESMARAIS, membre et directeur de recherche
Sébastien BÉLAND, membre
Luc PAQUETTE, membre externe

DÉDICACE

À tous mes compagnons de ma communauté Jésuite de Montréal,
je garde en mémoire votre soutien. À mes amis du labo,
vous me manquerez...

REMERCIEMENTS

Nous exprimons notre reconnaissance et notre gratitude à toutes les personnes qui de près ou de loin ont participé à l'être que nous sommes.

Nous voulons exprimer notre reconnaissance à nos parents qui nous ont donné la vie et nous ont accompagnés dans nos premiers pas sur cette terre.

Nous exprimons également notre gratitude envers la Compagnie de Jésus qui nous a soutenus moralement et financièrement tout au long de nos années d'études universitaires.

Nous voulons exprimer particulièrement nos remerciements au Dr. Ian Ian Shrier qui a généreusement mis à notre disposition les données de ses cours qui ont servi pour les recherches menées dans le cadre de cette thèse.

Nous exprimons notre reconnaissance et notre gratitude au Professeur Michel C. Desmarais qui nous a guidés tout au long de ces longues années de recherche. Nous exprimons notre gratitude pour sa patience et sa diligence.

Enfin, à toutes les personnes de bonne volonté qui œuvrent partout dans le monde pour l'avancement de la science et de la technologie au service du bien-être de tout le cosmos, nous exprimons nos sincères remerciements et reconnaissance.

RÉSUMÉ

Les cours ouverts en ligne de grande envergure (*"Massive open on-line courses"* : MOOC) s'appuient souvent sur la vidéo comme premier choix de contenu médiatique. Compte tenu de leur importance, il n'est pas surprenant que de nombreuses études portant sur la manière dont les étudiants utilisent les vidéos dans le cadre des MOOC aient vu le jour ces dernières années. La nécessité de développer des méthodes d'analyse des interactions vidéo pour améliorer les systèmes MOOC s'impose.

Dans le cadre de cette thèse, des techniques d'encodage des interactions sont présentées pour l'analyse détaillée des habitudes de visionnement des vidéos des étudiants et sont comparées à des approches dominantes (approches généralement utilisées dans la littérature).

La première technique d'encodage introduit SIVS (*Sequence of Interaction in Vector Space*) encode les séquences d'interaction vidéo dans un modèle d'espace vectoriel euclidien qui définit les mesures de distance entre elles. Un modèle est défini comme un centroïde de séquences. Nous appliquons la méthode d'encodage et menons une étude sur la façon dont les étudiants interagissent avec les vidéos en analysant les modèles d'interaction entre les différentes vidéos. Dans le cadre de cette méthode, l'analyse de l'influence de la vidéo sur les motifs est encadrée comme une tâche de classification basée sur les approches de la machine à vecteur de soutien (*"Support Vector Machine"* : SVM), de l'arbre de décision (*"Gradian Boosted Machine"* : GBM) et du plus proche voisin (*"K-Nearest Neighbors"* : KNN).

Les résultats montrent que la méthode basée sur l'encodage proposée reposant sur les séquences d'interactions a l'avantage de fournir un encodage plus précis d'un ensemble d'interactions individuelles des étudiants en un "modèle" que l'agrégation plus simple de variables des attributs d'écoute vidéo qui est souvent utilisée dans l'étude des traces d'interaction des étudiants avec les vidéos, ou des traces des étudiants en général. Cette méthode permet de définir des regroupements étiquetés, les vidéos dans notre cas, qui contrastent avec les regroupements non étiquetés standard qui nécessitent une interprétation souvent difficile des regroupements. Ces résultats révèlent qu'il existe une différence significative dans les modèles d'interaction entre les vidéos. Cela démontre l'utilité de ce nouvel encodage d'interaction vidéo par rapport à des approches plus simples et communément utilisées.

Dans une seconde recherche, nous supposons que les étudiants interagissent avec les vidéos en faisant une pause, en cherchant à avancer ou à reculer, en rejouant des segments, etc. Nous pouvons raisonnablement supposer que les étudiants ont des modèles d'interactions vidéo différents, mais il reste difficile de comparer les interactions vidéo des étudiants. Certaines méthodes ont été développées, telles que la chaîne de Markov et le calcul de la distance entre les séquences d'événements. Cependant, ces méthodes comportent des réserves, comme nous le montrons dans les exemples de prototypes. Dans cette thèse, nous proposons une nouvelle méthodologie de comparaison des séquences d'interaction vidéo basée à la fois sur le temps passé dans chaque état et sur la succession des états en calculant la distance entre les matrices de transition des séquences d'interaction vidéo à travers une représentation particulière que nous nommons "TMED" (*Transition Matrix Based on Edit Distance*).

La méthodologie de similarité proposée vise à combler une lacune méthodologique sur la représentation et la comparaison des séquences vidéo des méthodes d'interaction. La méthode proposée surmonte les limites des méthodes précédentes basées sur la chaîne de Markov et les séquences d'interactions connues sous le nom de "Edit Distance based" (ED) que nous appellerons la méthode de distance d'édition. La principale contribution de cette méthode est le fait qu'elle prend en compte le temps passé dans chaque état et le style général de succession des états. Elle offre une nouvelle technique aux chercheurs qui souhaitent comparer les interactions des utilisateurs de vidéo et trouver éventuellement un style d'interaction vidéo.

TMED combine deux styles de représentation de la séquence vidéo d'interaction et calcule la similarité en fonction de l'avantage de chaque style de représentation. La similarité basée sur la distance d'édition est généralement bonne sur une même plage de longueur des séquences d'interaction alors que la matrice de représentation basée sur l'interaction est meilleure sur des séquences de différentes plages de longueur.

La représentation TMED est également mieux en mesure de représenter une séquence d'interaction lors de tâches de classification, comme le montrent nos résultats. En fait, cette représentation est plus performante pour prédire la séquence d'interaction de l'étudiant et la vidéo avec laquelle l'étudiant a interagi à travers l'analyse de la représentation de leurs interactions vidéo. Les résultats obtenus de ces analyses montrent finalement que la représentation TMED proposée permet de mieux caractériser l'interaction vidéo dans une tâche de classification et de mieux discriminer quel étudiant interagit avec une vidéo, ou avec quelle vidéo un étudiant interagit.

La représentation TMED proposée ouvre de nouvelles voies de recherche. Elle pourrait aider à déterminer s'il existe une cohérence dans l'interaction vidéo pour chaque utilisateur en fonction du niveau de similarité de leurs séquences. On peut également utiliser cette représentation pour savoir si chaque style vidéo impose un style d'interaction spécifique aux utilisateurs de la vidéo. Une autre piste est le lien que l'on peut trouver entre le style vidéo d'interaction et le succès des étudiants. Les différents styles d'interaction peuvent-ils conduire à l'échec ou à la réussite dans un cours en ligne ? Nos investigations dans le cadre de cette thèse nous ont conduits à répondre à ces questions.

Dans certains systèmes de collectes de données des étudiants, l'accès aux informations des interactions à l'intérieur des vidéos peut être limité. Quelquefois il est simplement possible d'avoir accès à l'information d'écoute vidéo. Notre troisième recherche se focalise sur la prédiction du succès des étudiants basée sur les interactions vidéo à travers des mesures agglomératives spécifiques des mesures d'écoute des vidéos. La prédiction des succès des étudiants qui pourrait alimenter les tableaux de bord des instructeurs et les aider à adapter leur cours dans sa structure et son contenu, et pouvoir ainsi adapter des interventions à des groupes spécifiques d'étudiants, est l'objectif de cette recherche. À cette fin, la recherche de He, Zheng et al. 2018 a introduit trois mesures cumulatives (le taux d'assiduité : AR (*"Attendance Rate"*), le taux d'utilisation : UR (*"Utilization Rate"*), et le taux de visionnage : WR (*"Watching Ratio"*)) pour prédire les succès des étudiants. La limite de leur méthodologie est qu'elle dépend des facteurs graphiques qui peuvent être laissés à l'appréciation subjective du chercheur. Pour dépasser cette limite, nous introduisons une nouvelle mesure que nous appellons l'indice de visionnage WI (*"Watching Index"*)) basée sur AR et UR des étudiants qui interagissent avec les vidéos du MOOC afin de prédire quel groupe d'étudiants réussira ou échouera le cours.

Grâce à l'analyse quantitative qui comprend des mesures agglomératives telles que le taux d'assiduité (AR), le taux d'utilisation (UR) et l'indice de visionnage (WI), tels que définis dans cette thèse, il est possible d'identifier à partir des mesures agglomératives d'écoute jusqu'à 60 % des étudiants qui échoueront le cours en se basant sur l'interaction des étudiants de la première semaine d'un cours d'une durée totale de treize (13) semaines. On est même capable d'identifier 78 % des étudiants qui réussissent à partir de ces données. En utilisant les mesures agglomératives définies, les établissements d'enseignement peuvent signaler les étudiants à risque d'échec du MOOC en fonction des interactions de l'étudiant avec les vi-

déos d'apprentissage au début du cours (première semaine de cours). Notre étude montre une meilleure classification par rapport aux résultats des études précédentes de He, Zheng et al. 2018 qui ont utilisé les mêmes types de mesure. Cette recherche devrait aider les développeurs de MOOC à mieux identifier les étudiants susceptibles d'échouer le cours et donc à prendre des mesures pour l'éviter. Les résultats montrent que ces mesures, prises après la première semaine et à mi-parcours, peuvent être efficaces pour prédire les étudiants qui réussiront ou échoueront le cours.

Une quatrième recherche vise à prédire de façon précoce les succès des étudiants à partir de leur façon d'interagir avec les vidéos. Il s'agit de pouvoir introduire une autre méthode de prédiction des succès dans les systèmes d'apprentissage en ligne. L'analyse de la façon dont les étudiants interagissent avec ces vidéos devient essentielle pour comprendre comment les étudiants apprennent dans de tels environnements. Nous analysons des séquences vidéo interactions de 4 800 étudiants avec 9 vidéos différentes correspondant à l'interaction de la première semaine d'un cours en ligne de treize semaines pour prédire les succès des étudiants à la fin du cours. Nous utilisons la représentation TMED précédemment introduite qui est en fait une matrice de transition modifiée des interactions vidéo des étudiants et nous utilisons trois classificateurs tels que le support de vecteur machine (SVM), l'arbre de décision (GBM), le plus proche voisin (KNN) et la forêt aléatoire (RF) pour analyser les interactions vidéo des étudiants. Pour valider la méthode proposée, nous nous appuyons sur la capacité de la représentation TMED à discriminer les interactions vidéo individuelles des étudiants montre dans la deuxième recherche présentée ci-haut pour, ensuite, vérifier si l'on peut prédire les succès des étudiants à la fin du cours en se basant uniquement sur la première semaine d'interaction vidéo en utilisant cette représentation TMED.

Les résultats de cette recherche sont ensuite comparés aux résultats obtenus dans les recherches précédentes publiées dans deux conférences différentes. Cette comparaison montre que les résultats de prédictions obtenus par l'utilisation de la représentation TMED sont plus performantes que celles obtenues par ces deux autres méthodes basées sur des mesures agglomératives d'interaction vidéo des étudiants. Notre étude montre ainsi que la représentation proposée TMED de l'interaction vidéo peut être utilisée pour les problèmes de classification.

Une autre conclusion de cette recherche est que, en se basant uniquement sur l'interaction vidéo de la première semaine d'un cours en ligne de treize semaines, on peut prédire avec une précision raisonnable la note finale de l'étudiant en termes de réussite ou d'échec. Cette

conclusion pourrait répondre à l'un des besoins des instructeurs et des développeurs en matière d'identification précoce des étudiants qui risquent d'échouer, et pourrait fournir une aide supplémentaire pour leur éviter l'échec.

Dans nos travaux futurs, nous combinerons cette source d'information uniquement vidéo avec les résultats d'autres études qui s'alimentent de données universitaires et d'autres informations des données d'interaction des étudiants avec le système d'apprentissage pour améliorer la prédiction de la réussite ou de l'échec des étudiants. Il s'agit d'une première tentative de prédiction des succès des étudiants basée uniquement sur les interactions vidéo des étudiants. Les résultats sont prometteurs et doivent être consolidés dans les futures investigations.

ABSTRACT

Massive open online courses (MOOC) often rely on video as the first choice of media content. Given their importance, it is not surprising that many studies on how students use video in MOOC have emerged in recent years. There is a need to develop methods for analysing video interactions to improve MOOC systems.

In this thesis, interaction encoding techniques are introduced for the detailed analysis of students' video viewing habits and are compared to dominant approaches (approaches generally used in the literature).

The first encoding technique introduces SIVS (*Sequence of Interaction in Vector Space*) which encodes video interaction sequences into a vector spatial model that defines the distance measurements between them. A model is defined as a centroid of sequences. We apply the encoding method and conduct a study on how students interact with videos by analyzing the interaction patterns between different videos. In this method, the analysis of the influence of the video on the patterns is framed as a classification task based on the Support Vector Machine (SVM), the Gradient Boosting Machine (GBM: typically decision trees) and the K-nearest neighbor (KNN) approaches.

The results show that the proposed interaction sequence-based encoding method has the advantage of providing a more accurate encoding of a set of individual student interactions into a "model" than the simpler aggregation of video listening attribute variables that is often used in the study of student traces of interaction with videos, or student traces in general. This method allows to define labeled clusters, video styles in our case, which contrast with standard unlabeled clusters that require often difficult cluster interpretation. These results reveal that there is a significant difference in the interaction patterns between the videos and demonstrates the usefulness of this new video interaction encoding compared to simpler, commonly used approaches.

In a second study, we propose a new methodology for comparing video interaction sequences based both on the time spent in each state and on the succession of states by calculating the distance between the transition matrices of the video interaction sequences through a particular representation that we call "TMED" (*Transition Matrix Based on Edit Distance*).

The proposed similarity methodology is intended to fill a methodological gap in the representation and comparison of video footage of interaction methods. The proposed method overcomes the limitations of previous methods based on Markov chain and interaction sequences known as "Edit Distance based" (ED). The main contribution of this method is the fact that it takes into account the time spent in each state and the general style of state succession. It offers a new technique to researchers who wish to compare the interactions of video users and possibly find a style of video interaction.

TMED combines two styles of representation of the interaction video sequence and offers a measure of similarity according to the advantage of each style of representation. The similarity based on the editing distance is generally good over same length range of interaction sequences and the interaction-based representation matrix is better over sequences of different length ranges.

Results show that TMED representation is also capable of better representing a sequence of interaction during classification tasks. In fact, this representation is better at predicting the student's interaction sequence and the video with which the student interacted through the analysis of the representation of their video interaction. The results obtained from these analysis finally show that the proposed TMED representation allows to better characterize the video interaction in a classification task and to better discriminate which student interacts with a video, or with which video a student interacts.

The proposed TMED representation opens up new avenues of research. It could help to determine whether there is consistency in the video interaction for each user based on the level of similarity of their sequences. It can also be used to determine whether each video style imposes a specific style of interaction on the users of the video. Another track is the link that can be found between the video style of interaction and student success. Can different interaction styles lead to failure or success in an online course? Our investigations in the context of this thesis led us to answer these questions.

In some student data collection systems, access to interaction information within videos may be limited. Sometimes it is simply possible to access video viewing information. Our third research focuses on the prediction of student success based on video interactions through specific cumulative measures of video listening measures. The prediction of student success,

which could feed into instructors' dashboards and help them tailor their course structure and content, and thus be able to tailor interventions to specific groups of students, is the goal of this research. To this end, the research by He, Zheng, et al. 2018 introduced three cumulative measures (Attendance Rate: AR, Utilization Rate: UR, and Watching Ratio: WR) to predict student success. The limitation of their methodology is that it depends on graphical factors that can be left to the subjective judgement of the researcher. To overcome this limitation, we introduce a new measure we call the *Watching Index* (WI) based on the *Attendance Rate* (AR) and *Utilization Rate* (UR) of students who interact with MOOC videos to predict which group of students will pass or fail the course.

Through quantitative analysis that includes cumulative measures such as Attendance Rate (AR), Utilization Rate (UR), and Watching Index (WI), as defined in this thesis, it is possible to identify from the cumulative listening measures up to 60 % of the students who will drop out or fail the course based on student interaction in the first week of a thirteen (13) week course. One can even identify 78 % of successful students from this data. Using the defined cumulative measures, educational institutions can flag students at risk of failure, or dropping out, of MOOC based on student interactions with learning videos at the beginning of the course. Our study shows a better classification compared to the results of previous studies by He, Zheng, et al. 2018 who used the same types of measures. This research should help MOOC developers to better identify students who are likely to drop out or fail the course and thus take steps to avoid it. The results show that these measurements, taken after the first week and at mid-course, can be very effective in predicting which students will pass or fail the course.

A fourth research project aims to predict student success early on based on the way they interact with the videos. The aim is to be able to introduce an alternative methodology for predicting success in e-learning systems. Analysis of how students interact with these videos becomes essential to understand how students learn in such environments. We analyze video footage of 4,800 students' interactions with 9 different videos corresponding to the interaction of the first week of a thirteen-week online course to predict student success at the end of the course. We use the previously introduced TMED representation which is actually a modified transition matrix of student video interactions and use three classifiers, Machine Vector Support (SVM), Gradian Boosted Machine (GBM),K-Nearest Neighbors (KNN) and Random Forest (RF), to analyze student video interactions. To validate the proposed method, we rely on the ability of the TMED representation to discriminate the individual student video interactions shown in the second research presented above, and then verify whether student

success at the end of the course can be predicted based only on the first week of video interaction using TMED representation.

The results of this research are then compared to the results of previous research published in two different conferences. This comparison shows that the prediction results obtained by using the TMED representation outperform those obtained by these two other methods based on cumulative measures of student video interaction. The proposed TMED representation of video interaction can be used for success prediction.

Our study thus shows that, based solely on the video interaction of the first week of a thirteen-week online course, the student's final grade can be predicted with reasonable accuracy in terms of pass or fail. This finding could address one of the needs of instructors and developers for early identification of students at risk of failure, and could provide additional assistance in preventing failure.

In our future work, we will combine this video-only source of information with the results of other studies that draw on academic data and other information from student interaction data with the learning system to improve the prediction of student success or failure. This is a first attempt to predict student success based solely on student video interactions. The results are promising and need to be consolidated in future investigations.

TABLE DES MATIÈRES

DÉDICACE iii

REMERCIEMENTS iv

RÉSUMÉ v

ABSTRACT x

TABLE DES MATIÈRES xiv

LISTE DES TABLEAUX xviii

LISTE DES FIGURES xxi

LISTE DES SIGLES ET ABRÉVIATIONS xxiii

LISTE DES ANNEXES xxiv

CHAPITRE 1 INTRODUCTION 1

CHAPITRE 2 QUESTIONS DE RECHERCHE 10

2.1 Contexte général 10

2.2 Représentations et algorithmes d'analyse des séquences vidéo 12

2.2.1 Représentation d'écoute vidéo sensible à la durée de la vidéo SIVS. . 13

2.2.2 QR.1 : La représentation SIVS est-elle plus performante que la représentation cumulative pour discriminer des écoutes de durée semblables ? 14

2.2.3 Représentation d'écoute vidéo insensible à la durée de la vidéo TMED 14

2.2.4 QR. 2 : La présentation TMED est-elle plus performante que les approches séquentielles et de chaîne de transition pour discriminer entre différents types d'écoutes ? 15

2.3 Caractère unique des traces d'interaction d'un étudiant et d'une vidéo 16

2.3.1 QR. 3 : Un étudiant possède-t-il un style d'écoute qui lui est propre ? Une vidéo possède-t-elle aussi une signature d'écoute ? 16

2.4 Prédiction du succès à partir des traces d'écoute vidéo 17

2.4.1 Les mesures agglomératives d'écoutes vidéo 18

2.4.2 QR. 4 : Les mesures agglomératives permettent-elles de prédire avec précision les chances du succès des étudiants ? 18
2.4.3 Utilisation de la représentation TMED pour la prédiction précoce du succès 19
2.4.4 QR. 5 : Comment la représentation TMED se compare-t-elle aux autres méthodes de prédiction de succès ? 19

CHAPITRE 3 REVUE DE LA LITTÉRATURE 21
3.1 Introduction 21
3.2 Analyse de l'apprentissage utilisant des interactions vidéo des étudiants . . . 22
3.2.1 L'évaluation de l'efficacité par l'analyse des traces vidéo 22
3.2.2 Les formes d'enseignement par l'analyse des traces vidéo 23
3.3 Les analyses des traces vidéo liées à nos investigations. 25
3.4 Séquence d'activités vidéo des étudiants et représentations (QR.1 et QR.2) . 26
3.4.1 Représentation d'interaction vidéo basée sur les séquences d'événements 27
3.4.2 Représentations basées sur les transitions d'écoute vidéo 28
3.4.3 Représentations basées sur les mesures cumulatives d'écoute vidéo . . 29
3.5 Codage d'interaction vidéo basé sur les modèles de visionnement (QR.1 et QR.2) 32
3.6 Mesure de la similarité des séquences des étudiants (QR.3) 34
3.7 Prédiction des succès d'étudiants basée sur les mesures agglomératives d'interaction vidéo (QR.4) 38
3.8 La représentation d'interaction vidéo en vue de la prédiction des succès des étudiants (QR. 5) 42

CHAPITRE 4 ENSEMBLE DE DONNÉES ET TRAITEMENT 45
4.1 Introduction 45
4.2 Statistiques générales du cours Body 101x 45
4.3 Organisation interne du cours 49
4.4 L'utilisation des traces vidéo 50

CHAPITRE 5 ENCODAGE SIVS DES INTERACTIONS VIDÉO 54
5.1 Introduction 54
5.2 Méthodologie pour coder et classifier les interactions vidéo 55
5.2.1 Encodage vidéo basé sur les mesures cumulatives d'écoute vidéo. . . . 55
5.2.2 Passage de séquence d'activité à l'encodage SIVS 57
5.2.3 Utilisation des centroïdes pour définir les classes 59
5.2.4 Classification des interactions vidéo 61

5.3 Expérimentations 61
5.3.1 Expérimentation 1 : données synthétiques 62
5.3.2 Expérimentation 2 : données réelles 63
5.4 Résultats 64
5.4.1 Résultats des données synthétiques 65
5.4.2 Résultats des données réelles 65
5.5 Conclusions 66

CHAPITRE 6 ENCODAGE TMED POUR SIMILARITÉ D'ÉCOUTE VIDÉO . . 68
6.1 Introduction 68
6.2 Contexte 69
6.2.1 D'événements à une séquence d'activités 69
6.2.2 La représentation sous la forme de la chaîne de Markov 70
6.2.3 Distance de séquence d'édition : OM Distance 71
6.2.4 Performance de classification multi-classe 72
6.3 Représentation proposée, TMED 75
6.3.1 Construction de la matrice de transition 75
6.3.2 Distance entre deux matrices : distance matricielle 78
6.3.3 Similarité TMED (S_{mat}) entre deux matrices 78
6.3.4 Similarité OM Distance (S_{om}) entre deux séquences 80
6.3.5 Similarité (S_{mark}) entre deux matrices de Markov 81
6.4 Étapes de la mesure de la similarité matricielle proposée 81
6.5 Validation de la méthodologie de similarité proposée 85
6.5.1 Cas prototypiques 86
6.5.2 Validation de la sensibilité de la représentation TMED proposée . . . 87
6.5.3 Identification des séquences d'interaction des étudiants 91
6.5.4 Identification des vidéos à partir des interactions des étudiants 92
6.5.5 Application de la méthodologie de similarité dans l'analyse textuelle . 92
6.6 Résultats 96
6.7 Comparaison entre les représentations SIVS et TMED 98
6.8 Conclusion 99

CHAPITRE 7 PRÉDICTION DE L'ÉCHEC ET DU SUCCÈS DES ÉTUDIANTS BASÉE SUR LES MESURES AGGLOMÉRATIVES D'UTILISATION DES VIDÉOS 101
7.1 Introduction 101
7.2 Les mesures agglomératives d'utilisation des vidéos 103

7.2.1 Taux d'assiduité (*"Attendance Rate"*) 103
7.2.2 Taux d'utilisation (*"Utilization Rate"*) 103
7.2.3 Taux de visionnage (*"Watch Ratio"*) 104
7.2.4 L'index de visionnage (*"Watch Index"*) 105
7.3 Méthodologie de classification 105
7.3.1 Classification en utilisant le taux de visionnage (WR) 105
7.3.2 Classification proposée des groupes 108
7.3.3 Préparation des données 109
7.4 Conclusion 113

CHAPITRE 8 PRÉDICTION PRÉCOCE DU SUCCÈS DES ÉTUDIANTS BASÉE SUR LA REPRÉSENTATION D'INTERACTIONS VIDÉO TMED 115
8.1 Introduction 115
8.2 Méthodologie 116
8.2.1 Classificateurs 119
8.2.2 Prédiction des succès des étudiants 119
8.3 Résultats 121
8.4 Comparaison des performances de classification de TMED aux études précédentes 124
8.5 Conclusion et travaux futurs 127

CHAPITRE 9 CONCLUSIONS ET TRAVAUX FUTURS 129
RÉFÉRENCES 137

ANNEXES 156

LISTE DES TABLEAUX

Tableau 4.1 Informations générales sur le cours. 46
Tableau 4.2 Répartition des notes finales des étudiants. 46
Tableau 4.3 Le nombre des vidéos par style vidéo du cours Body 101x. . 52
Tableau 5.1 Fréquence des styles vidéo (entre parenthèses, les fréquences retenues dans l'expérience) et nombre moyen d'étudiants qui ont interagi avec les vidéos. 64
Tableau 5.2 Précision des séries de validation croisée de l'identification de la vidéo décuplée de quatre ensembles d'interactions des données synthétiques de même longueur en utilisant l'approche basée sur les mesures cumulatives d'écoute vidéo et la représentation SIVS. 65
Tableau 5.3 Précision des séries de validation croisée l'identification de la vidéo décuplée de quatre vidéos de même longueur en utilisant l'approche basée sur les mesures cumulatives d'écoute vidéo et la méthode proposée basée sur SIVS. Ceux qui sont rapportés ici sont les précisions moyennes des quatre ensembles différents des vidéos de même longueur dans chaque ensemble. 66
Tableau 6.1 Résultats de performance de classification par classe. 74
Tableau 6.2 Résultats de la validation croisée vingt fois 400 séries de prédictions d'étudiants de 5 et 12 étudiants utilisant trois méthodes différentes de représentation de l'interaction des étudiants avec des vidéos montrant que la technique de représentation proposée est plus performante que les autres. ED (Distance d'édition), MC (Chaîne de Markov, distance de Frobenius), TMED (Combinaison, distance de Frobenius). Les valeurs de F_1 sont des valeurs moyennes des F_1 de toutes les classes (voir section 6.2.4). 94
Tableau 6.3 Résultats de la validation croisée de 400 séries de vidéos de prédiction des vidéos de 45 et 108 enregistrements d'interactions d'étudiants en utilisant trois méthodes différentes de représentation de l'interaction des étudiants avec les vidéos, ED (Distance d'édition), MC (Chaîne de Markov), TMED (Combinaison). Les valeurs de F_1 sont des valeurs moyennes des F_1 de toutes les classes (voir section 6.2.4). . . . 95
Tableau 7.1 Statistiques descriptives de AR, UR et WR après la première semaine et entre parenthèses à la moitié du cours. 107

Tableau 7.2 Répartition par groupe du nombre des étudiants qui ont réussi ou échoué après la première semaine et à mi-parcours (sixième semaine), basée sur WR et telle que proposée par He, Zheng et al. 2018 108
Tableau 7.3 Répartition par groupe du nombre des étudiants qui ont réussi ou échoué après la première semaine et à mi-parcours (sixième semaine) basée sur WI. 109
Tableau 7.4 Comparaison des résultats de He, Zheng et al. 2018 et de la méthode proposée (Proposition) présentée à des fins de comparaison. 110
Tableau 7.5 Probabilités de classification dans chaque groupe en utilisant les deux méthodes. Ici $P(E/G_1)$= Probabilité d'échouer lorsqu'on est classé dans le groupe I et $P(S/G_2)$ = Probabilité de succès (de réussir) quand on est classé dans le groupe II. 111
Tableau 8.1 Résultats de la validation croisée avec 5 replis de prédiction des succès ou non des étudiants pour chaque vidéo. Les données sont non balancées avec 15% des étudiants qui ont réussi et 85% qui ont échoué par vidéo (soit 722 et 4078 étudiants) et entre parenthèses, les données sont balancées avec 50% d'étudiants qui ont réussi et 50% d'étudiants qui ont échoué par vidéo (soit 722 et 722 étudiants). 123
Tableau 8.2 Résultats de la prédiction des succès des étudiants en combinant les résultats des données de plusieurs vidéos pour prédire le succès ou non des étudiants. Notez une augmentation de précision à mesure que les données augmentent. Le résultat de la première semaine est celle de la combinaison des neuf (9) vidéos et à mi- parcourt du cours (six semaines) celles des soixante-dix-neuf (79) vidéos. Les résultats entre parenthèses sont des résultats des données balancées avec 722 étudiants de chaque classe. En dehors des parenthèses sont les résultats des données non balancées soit de 4800 étudiants au complet. En gras sont les valeurs de performances de prédiction après la première semaine d'interaction et à la mi-parcours du cours (après six semaines). 124

Tableau 8.3 Comparaison des performances de trois méthodes de prédiction de la réussite des étudiants à la fin du cours en termes de succès ou d'échec après la première semaine d'interaction vidéo et après six semaines pour les données non équilibrées et entre parenthèses les résultats des données équilibrées. Nous comparons ici les performances de deux méthodes différentes de prédiction de la réussite ou de l'échec des étudiants : l'une basée sur des règles de classification et l'autre utilisant des classificateurs. Méthode 1 = méthode proposée dans He, Zheng et al. 2018 basée sur WR (utilisation de règles), Méthode 2 = méthode proposée dans Mbouzao, Michel C Desmarais et Shrier 2020 basée sur WI (utilisation de règles), Méthode 3 = méthode proposée dans cette étude basée sur TMED (utilisation de classificateurs) En gras les performances de prédiction de la méthode proposée basée sur la représentation TMED. Il faut noter ici que pour les données balancées les deux méthodes précédentes performent en déca du hasard après la première semaine mais dans la réalité, les données ne sont toujours pas balancées en général. 126
Tableau A.1 Vidéos et activités du cours Body101x (Partie 1) 156
Tableau B.1 Vidéos et activités du cours Body101x (Partie 1 suite) 157
Tableau C.1 Vidéos et activités du cours Body101x (Partie 2) 158
Tableau D.1 Vidéos et activités du cours Body101x (Partie 3) 159
Tableau H.1 Composition de la taille du vocabulaire des 15 migrants de l'échantillon 169
Tableau H.2 La distance euclidienne entre les textes. 170
Tableau H.3 La similarité entre les discours exprimée en termes de pourcentage de ressemblance. Par exemple 0.55 signifie 55% de ressemblance et 1.0 signifie 100% de ressemblance. 171

LISTE DES FIGURES

Figure 4.1 Plate-forme avec vidéo montrant les divers niveaux d'écoute de la vidéo. 48
Figure 4.2 Plate-forme avec vidéo montrant la transcription de la vidéo. . 48
Figure 4.3 Structure des traces brutes extrait du serveur. 53
Figure 5.1 Représentation de l'interaction entre l'étudiant i et une vidéo d'une durée de n transitions. 60
Figure 6.1 Exemple de matrice de confusion d'une classification en vue de déterminer les performances du classificateur dans le cas de 3 classes. 73
Figure 6.2 Flux de la méthode proposée pour calculer les similarités entre les séquences vidéo des étudiants. Nous obtenons les trois similarités (ED, Markov et TMED) en suivant le flux des opérations. 82
Figure 6.3 Même séquence cyclique de transitions. Le cycle peut commencer à différentes étapes mais suit le même schéma de transition. . . . 88
Figure 6.4 Résultat de la similarité : (a) La similarité basée sur la distance d'édition (ED) ne peut pas reconnaître la similarité des séquences cycliques. (b) La similarité basée sur la chaîne de Markov peut reconnaître la similarité, comme on s'y attendait, les similarités avoisinent 100%. (c) La similarité basée sur la représentation TMED proposée peut reconnaître les séquences cycliques. Également les similarités avoisinent 100% 88
Figure 6.5 Même séquence d'états avec des longueurs différentes. 89

Figure 6.6 (a) la similarité basée sur la séquence d'Édition montre une certaine progression mais très déséquilibrée par rapport à ce qui est attendu lorsqu'on observe les séquences. Les degrés de similarité des diagonales ne sont pas les mêmes.(b) la similarité basée sur la chaîne de Markov ne peut pas reconnaître la durée dans chaque état car les probabilités de transition entre les états sont préservées sur la longueur de chaque séquence. Toutes les séquences sont considérées comme identiques, même si elles ne sont pas tout à fait identiques en fonction de la durée dans chaque état. (c) la similarité proposée par TMED peut reconnaître le fait que ces séquences soient identiques mais le niveau de similarité est basé sur le temps passé dans chaque état. Le résultat montre l'augmentation progressive du niveau de similarité. Les diagonales montrent le même degré de similarité, c'est ce qui était attendu vue la structure des interactions. 90

Figure 6.7 Résultats de la comparaison entre la méthode de représentation proposée et d'autres méthodes de prédiction de la vidéo (première ligne) et de l'étudiant (deuxième ligne) à partir des interactions des étudiants. markov= représentation en chaîne de Markov, optimal matching = représentation séquentielle et proposed = représentation TMED. . . . 93

Figure 6.8 La similarité entre les discours exprimée en terme de pourcentage de ressemblance. Par exemple 0.55 signifie 55% de ressemblance et 1.0 signifie 100% de ressemblance. 96

Figure 6.9 La similarité entre les discours de manière visuelle. 97

Figure 7.1 Distribution de AR et UR de la première semaine 106

Figure 7.2 Distribution de WR 107

Figure 7.3 Distributions d'indicateurs à la moitié de la durée du cours. . 110

Figure 8.1 Flux de prédiction de la réussite ou de l'échec d'un étudiant. Le classificateur peut être SVM, GBM, KNN ou RF. La sélection de la prédiction finale pour chaque étudiant se fait en choisissant la majorité des prédictions des 3, 5, 7, 9 ou 79 vidéos selon l'étape de prédiction. Nous prédisons ici pour les données balancées (A) et les données non balancées (B). 118

Figure 8.2 Exemple de la variation de la valeur de K pour déterminer la valeur de K avec la meilleure précision de la prédiction en utilisant KNN. Ici c'est l'exemple de la variation de K pour la vidéo 1. Ici k=11 est retenu pour les prédictions. 120

Figure H.1 La similarité entre les discours de manière visuelle. 171

LISTE DES SIGLES ET ABRÉVIATIONS

AIED	International Conference on Artificial Intelligence in Education
AR	Attendance Rate
EC TEL	European Conference on Technology Enhanced Learning
ED	Edit Distance
EDM	Educational Data Mining
EIAH	Environnement Informatique d'Apprentissage Humain
GBM	Gradient Boosting Machine (Boosted Trees)
KNN	K-Nearest Neighbors
MC	Markov Chain
MOOC	Massive Open Online Courses
QR.	Question de Recherche
SIVS	Sequence of Interaction in Vector Space
SVM	Support Vector Machine
TMED	Transition Matrix Edit Distance
UR	Utilization Rate
WI	Watching Index
WR	Watching Ratio

LISTE DES ANNEXES

Annexe A Vidéos et activités du cours Body101x Partie 1 156

Annexe B Vidéos et activités du cours Body101x Partie 1 suite 157

Annexe C Vidéos et activités du cours Body101x Partie 2 158

Annexe D Vidéos et activités du cours Body101x Partie 3 159

Annexe E Description des styles vidéo 160

Annexe F Description des Classificateurs 162

Annexe G Méthodes de calcul des distances entre les Séquences 166

Annexe H Méthodologie de similarité pour analyse de texte 169

CHAPITRE 1 INTRODUCTION

Étant donné l'importance grandissante des vidéos dans un système MOOC, notre travail va consister en un développement d'une approche d'analyse des traces vidéo. L'interaction des utilisateurs avec les vidéos dans un MOOC peut éventuellement révéler leur style d'apprentissage et leur succès. L'essentiel de la matière du cours dans les systèmes MOOC est donné de plus en plus de nos jours dans les vidéos. Les problèmes, les quiz, et les exemples donnés dans le cadre du cours sont essentiellement liés aux vidéos. Pouvoir déterminer la manière dont les étudiants interagissent avec les vidéos serait alors d'une importance particulière pour l'amélioration du système par les développeurs et aider les instructeurs à mieux accompagner les étudiants pour qu'ils puissent continuer leur apprentissage.

Le développement des techniques d'analyse pour l'utilisation des vidéos et les interactions des apprenants avec ces dernières est une problématique importante dans un tel contexte. Ainsi pourrait-on analyser ce que nous disent les traces d'interaction des étudiants avec les vidéos. Dans cette perspective, le développement des techniques qui permettraient de mieux comprendre et analyser l'utilisation des vidéos, dans le cadre des MOOC, aiderait grandement les développeurs et les instructeurs de la plate-forme à prendre des décisions appropriées concernant l'implémentation et l'utilisation des techniques pédagogiques. Cette démarche serait essentielle pour améliorer l'expérience d'apprentissage en ligne.

Le recours à l'enseignement en ligne devient crucial dans la mesure où en situation de crise, l'enseignement en salle est momentanément impossible. Avoir une solution de rechange en ligne peut aider à l'expérience d'apprentissage et devient important. Avoir des techniques d'évaluation de l'expérience de l'apprenant aide à pouvoir aider à une meilleure implémentation des techniques d'apprentissage en ligne, mais également à améliorer les techniques pédagogiques d'apprentissage en ligne.

Dans le cadre de cette thèse, le développement des techniques d'analyse s'articule autour de cinq contributions principales d'évaluation de l'interaction de l'apprenant avec les vidéos. Nous utiliserons le terme "mesures cumulatives" pour désigner les mesures d'écoutes vidéo au niveau d'une vidéo à savoir le pourcentage de temps qu'un étudiant a consacré à jouer une vidéo, puis à la position pause d'une vidéo par rapport au temps total passé à interagir avec la vidéo ensuite le nombre de fois de recherche en arrière et en avant dans la vidéo consi-

dérée. Nous appellerons "mesures agglomératives", les mesures d'interactions qui concerne un ensemble des vidéos (par exemple les vidéos de la première semaine) pour déterminer divers aspects de l'utilisation de ces vidéos par un étudiant en particulier. Ces contributions peuvent se résumer en cinq axes qui touchent la méthodologie d'analyse d'écoute vidéo à travers l'utilisation des traces. Ces contributions peuvent être formulées comme suit :

(i) **Proposition d'une représentation d'interaction vidéo SIVS :** Boniface Mbouzao, Michel Desmarais and Ian Shrier,*A Methodology for Student Video Interaction Patterns Analysis and Classification*, In Proceedings of the International Conference on Educational Data Mining (EDM) (12th, Montreal, Canada, July 2-5, 2019).

SIVS (*"Student Interactions in Vector Space"*) est une représentation des interactions vidéo des étudiants dans un espace vectoriel euclidien dans le but d'une analyse détaillée des tendances des étudiants en matière de visionnement de vidéos et la comparaison avec une approche jusqu'ici utilisée basée sur les mesures cumulatives d'écoutes vidéo. La méthode que nous proposons et qui utilise cette représentation des interactions vidéo, encode les séquences d'interaction vidéo dans un modèle d'espace vectoriel euclidien. Les séquences d'interaction sont encodées sous forme de matrice. Pour définir la distance entre les matrices en faisant correspondre le premier indice à un indice de ligne et le second à un indice de colonne et l'on utilise la notation dans le cas des coordonnées purement covariantes. Ainsi, la distance entre deux matrices qui encode deux séquences d'interaction vidéo est défini entre elles, comme la norme de Frobenius. On définit alors un prototype d'écoute d'une vidéo comme un centroïde de séquences en prenant la moyenne de tous les éléments des matrices de toutes les écoutes de cette vidéo pour obtenir une représentation moyenne de toutes les écoutes. La matrice de cette moyenne de toutes les écoutes constitu le centroïde de l'écoute de la vidéo en question. Nous appliquons cet encodage et menons une étude sur la façon dont les étudiants interagissent avec les vidéos en analysant les différences d'interaction entre les différentes vidéos à travers leur centroïdes. Dans le cadre de la validation de notre encodage, l'analyse de l'influence de la vidéo sur l'interaction des étudiants avec elle est présentée comme une tâche de classification. L'hypothèse ici est que chaque vidéo impose une certaine façon particulière d'interagir avec elle. En effet, dans certains types de vidéos ou des quizs sont intégrés dans la vidéo ou simplement des "pauses" sont imposées obligatoirement à un moment de l'écoute, la vidéo impose clairement une façon d'interagir avec elle. Il est question alors de savoir si cela peut être généralisé et l'on peut identifier une vidéo à partir de l'encodage de l'interaction d'un étudiant avec cette vidéo. Nous utilisons pour nos tests les approches bien connues pour les tâches de clas-

sification à savoir la machine à vecteur de soutien (SVM), de l'arbre de décision (GBM) et du plus proche voisin (KNN). Nous comparons les résultats de l'encodage proposé à ceux obtenus avec un encodage souvent utilisé dans la littérature basée sur les mesures agglomératives des écoutes vidéo. Les résultats révèlent que l'encodage que nous proposons est meilleur dans la reconnaissance à travers l'encodage de l'interaction des étudiants, la vidéo avec laquelle les étudiants interagissent. Une contribution de cette recherche est la proposition d'une représentation d'interaction vidéo plus appropriée dans les tâches de classifications des séquences d'interaction par rapport à celle existante. Une autre contribution non négligeable est le fait que cet encodage inclut dans sa structure les parties de la vidéo réécoutées par l'étudiant et des parties non écoutés qui peuvent servir dans l'analyse des écoutes alors que cela ne serait pas possible avec un encodage basé sur les mesures agglomératives des écoutes vidéo. Le chapitre 5 de cette thèse présente les détails de ces contributions.

(ii) **Proposition d'une représentation des séquences d'interactions TMED en vue de la comparaison de niveau de similarité entre les séquences d'interaction vidéo :** Boniface Mbouzao, Michel Desmarais and Ian Shrier, *Methodology to measure of similarity in student video sequence of interactions*, In Proceedings of the International Conference on Educational Data Mining (EDM), 13th, July 10-13, 2020.

Nous pouvons raisonnablement supposer que les étudiants ont des modèles d'interactions vidéo différents. Mais il reste difficile de comparer les séquences d'interactions vidéo des étudiants entre elles de façon efficace. Dans la littérature, deux représentations sont utilisées à cet effet. La première est basée sur la représentation de chaîne de Markov. On représente l'interaction vidéo de l'étudiant sous forme de matrice où chaque colonne et chaque ligne représentent les états de transition d'interaction. On exprime en termes de probabilité de passage d'un état à l'autre au cours de l'interaction. Ces probabilités nous proviennent des calculs cumulatifs des diverses transitions faites par l'étudiant. La limite d'une telle représentation dans la comparaison entre deux chaînes de Markov représentant deux interactions peut conduire à la conclusion que ces deux interactions sont semblables alors qu'elles sont différentes. La similarité de deux représentations d'écoute vidéo utilisant la représentation de chaîne de Markov a pour limite de se concentrer sur la succession des transitions laissant tomber le temps passé dans chaque état. Il suffit que les probabilités de transitions soient semblables pour que deux écoutes soient déclarées similaires.

Une autre représentation d'interaction vidéo utilisée pour la comparaison de la façon dont les étudiants interagissent avec les vidéos est la séquence d'événements prenant en compte le temps mis dans chaque événement dans une ligne de temps. La distance d'édition entre les séquences d'événements permet de mesurer la similarité entre les séquences. La limite d'une telle représentation dans la comparaison des séquences d'interactions est qu'il suffit d'un décalage d'événement pour que deux séquences pourtant similaires soient déclarées très distantes.

Pour dépasser les limites de ces deux représentations dans le cadre de la comparaison des interactions vidéo, une représentation plus efficace pour cette tâche est nécessaire. La représentation TMED (*"Transcition Matrix based on Edit Distance"*) est celle qui bénéficie des avantages de ces deux premières représentations. Il s'agit d'une matrice de transition qui respecte le nombre de temps passé dans chaque état par rapport aux autres états. Pour valider l'efficacité de cette représentation, une méthodologie de comparaison entre les représentations des interactions vidéo est proposée. Le premier niveau de validation est obtenu à travers des cas prototypes dont les résultats sont connus d'avance. On compare donc les résultats de comparaison obtenus par chacune des trois représentations. Avec la représentation TMED nous obtenons toujours le résultat attendu ce qui n'est pas le cas avec les deux autres représentations. Le second niveau de validation de cette représentation est fait sur des données réelles pour voir si cette nouvelle représentation est capable de mieux discriminer les étudiants et les vidéos. Prenant l'hypothèse de la première contribution à savoir chaque vidéo impose une façon d'interagir avec elle, il nous faut voir si la représentation TMED est capable d'identifier mieux que les deux autres représentations les vidéos. Il s'agit donc de voir comment chaque représentation est capable de reconnaitre la vidéo avec laquelle interagit un étudiant à partir de sa représentation d'interaction. Une seconde hypothèse dans cette étude est qu'un étudiant en particulier a sa façon d'interagir avec les vidéos. Pour vérifier cette hypothèse, nous avons fait une expérimentation pour déterminer lequel des trois représentations est capable de reconnaitre l'étudiant à partir de son interaction avec une vidéo quelconque. Les résultats montrent qu'avec la représentation TMED on est capable mieux qu'avec les deux autres représentations de reconnaitre l'étudiant à partir de sa représentation d'interaction vidéo.

Ainsi la représentation TMED proposée permet de mieux caractériser l'interaction vidéo dans une tâche de discriminer quel étudiant interagit avec une vidéo, ou chercher avec

quelle vidéo un étudiant interagit. En résumé, la contribution de cette recherche est une proposition de représentation d'interaction vidéo pour les tâches de comparaison. Une autre contribution est une méthodologie de comparaison des séquences d'interaction vidéo pour des tâches d'identification des patterns d'écoute semblables. Une dernière contribution est le fait qu'on soit capable grâce à la représentation TMED d'identifier l'étudiant. Ces contributions sont présentées dans le chapitre 6 de cette thèse.

(iii) **Application de la méthodologie de niveau de similarité dans un contexte d'analyse textuelle :** Accepté au XXIIème congrès de la SFSIC, Martial Sylvain Marie Abega Eloundou, Boniface Mbouzao, *Interactions « discours de migrants, flux migratoires, espaces géographiques »*, SFSIC 2020, 27-29 janvier 2021, Échirolles (France).

Dans l'étude de vidéo et des transcrits vidéo, une problématique courante est celle de savoir la similarité de discours entre les intervenants. Dans cette étude nous utilisons les vidéos des interviews des migrants africains arrivés illégalement en Europe pour déterminer les similitudes dans leur discours. La représentation utilisée ici est celle similaire à l'analyse de textes. Nous avons constitué les mots clés du corpus à partir des mots les plus fréquents dans les discours de tous les migrants qui ne sont pas des conjonctions ni des ponctuations. Nous acceptons un mot faisant partie du corpus lorsqu'il est utilisé au moins par trois intervenants. Ces mots vont être des attributs du vecteur de chaque intervenant. Pour chaque discours, la valeur correspondante à la colonne d'un mot correspond au nombre de l'occurrence du mot dans ce discours. A partir de cette représentation vectorielle nous utilisons la méthode proposée précédemment pour évaluer la similarité entre les discours de tous les intervenants. Ensuite, on extrait les parties des discours semblables pour en faire un récit atypique du corpus d'intervention. A partir de ce discours atypique l'on est capable de faire des analyses sociologiques plus intéressantes qui ont le mérite d'avoir pour source d'information les migrants eux-mêmes parlant des raisons et motivations de leurs aventures. La contribution de cette recherche est qu'on peut utiliser la méthodologie de similarité proposée précédement dans un cadre d'une étude textuelle. Cette contribution est également présentée dans le chapitre 6 de cette thèse.

(iv) **Proposition de prédiction des succès des étudiants à partir des mesures cumulatives d'interaction vidéo :** Boniface Mbouzao, Michel Desmarais and Ian

Shrier, *Early prediction of success in MOOC from video interaction features*, In Proceedings of the 21th International Conference on Artificial Intelligence in Education, AIED 2020, July 6th-10th 2020.

La prédiction des succès ou non des étudiants en vue d'alimenter les tableaux de bord des instructeurs dans les premières semaines du cours a pour but d'aider les développeurs des MOOC à adapter la structure et le matériel de leurs cours pour prévenir les échecs. Cette prédiction pourrait être utile également pour adapter les interventions à des groupes d'étudiants spécifiques ; c'est là un objectif de recherche précieux. Cette contribution peut être vue comme une prédiction précoce des succès des étudiants à partir des mesures agglomératives de l'utilisation des vidéos. Nous introduisons une nouvelle mesure que nous avons appelé WI (Index de Visionnement). Il s'agit de pouvoir prédire le succès ou non d'un étudiant dans un cours en ligne en termes de réussite ou d'échec basé sur des mesures agglomératives de la façon dont l'étudiant a interagi avec les vidéos du cours dans la première semaine. Les résultats obtenus de cette méthode proposée basée sur cette nouvelle mesure sont comparés à une autre étude qui utilise la plupart des mesures agglomératives d'écoute vidéo pour prédire le succès ou non des étudiants dans la littérature (He, Zheng et al. 2018). Les résultats des prédictions basées sur la nouvelle mesure proposée est plus performante que celle de l'étude précédente et reste une méthode de prédiction basée sur des mesures purement objectives pour la séparation des groupes des étudiants.

En somme, la mesure d'écoute vidéo introduite (Index vidéo : WI) est la combinaison du taux d'assiduité et du taux d'utilisation définit dans cette thèse en vue d'une prédiction précoce des succès ou non des étudiants en termes de réussite ou d'échec. Pour la validation, la prédiction basée sur cette mesure est comparée à une méthode semblable dans la littérature (He, Zheng et al. 2018) basée sur des mesures agglomératives semblables. Cette troisième contribution tout en simplifiant l'étude précédente en se basant uniquement sur des mesures agglomératives objectives, performe mieux en termes de prédiction de ceux qui vont réussir ou non le cours. Cette contribution est developée dans cette thèse au chapitre 7.

(v) **Prédiction précoce des succès des étudiants basée uniquement sur la représentation d'interactions vidéo TMED :** Boniface Mbouzao, Michel Desmarais and Ian Shrier, *Video interaction based student performance prediction in MOOC.* sera sou-

mis à la conférence EDM 2021.

La représentation TMED est d'abord validée sur sa capacité à discriminer de manière efficace les séquences d'interactions vidéo des étudiants avant d'être utilisée dans la prédiction du succès ou non en termes de réussite ou d'échec. L'hypothèse fondamentale de cette recherche est qu'il existe une façon d'interagir avec les vidéos qui caractérisent les étudiants qui réussissent ou l'échouent un cours. Dans la mesure où la représentation TMED distingue bien les écoutes vidéo, elle peut donc être utilisée pour la prédiction du succès ou de l'échec en identifiant les écoutes des deux groupes de façon efficace. Les résultats obtenus contribuent à valider cette hypothèse. Les résultats obtenus dans le cadre de cette prédiction en effet sont comparés aux résultats obtenus par les prédictions en utilisant les mesures agglomératives de l'utilisation des vidéos à travers les interactions vidéo de la recherche précédente. Les résultats montrent que cette prédiction basée sur la représentation TMED des interactions vidéo a des performances supérieures aux deux méthodes de la littérature présentées dans la recherche précédente.

Cette dernière contribution de cette thèse est l'utilisation de la présentation TMED en vue de la tâche de prédiction du succès ou non des étudiants en termes de réussite ou d'échec. La représentation TMED a l'avantage de montrer de façon unique l'interaction vidéo d'un étudiant. C'est pour cette raison qu'elle a été utilisée dans la comparaison des séquences d'écoute dans la seconde contribution de la thèse. Ici, il s'agit d'utiliser la façon d'interagir des étudiants pour prédire leurs succès ou non. La validation de cette tâche se fait à travers la comparaison des résultats de prédiction à ceux obtenus par les deux méthodes précédentes de la troisième contribution.

Pour les prédictions, notre étude utilise trois classificateurs :la machine à vecteurs de support (SVM), l'arbre de décision (GBM) et le plus proche voisin (KNN). Pour valider la méthode proposée à prédire les succès ou non, nous testons d'abord la capacité de la représentation des interactions vidéo que nous proposons pour distinguer les interactions vidéo individuelles des étudiants et, ensuite, celle de vérifier la possibilité de prédire les succès des étudiants à la fin du cours en se basant seulement sur la première semaine d'interaction vidéo des étudiants. La contribution de cette recherche peut être résumée comme l'utilisation de la représentation de séquence d'interaction vidéo TMED en vue d'identifier les patterns d'interaction vidéo, d'une part, et, de l'autre, de prédire de façon précoce le succès ou non de l'étudiant dans un cours en ligne en termes de réussite ou d'échec. Cette contribution va être soumise sous peu dans le cadre d'une

conférence.

Les performances des prédictions de cette recherche ont été ensuite comparées aux résultats obtenus dans les recherches précédentes publiées dans des conférences. Cette comparaison montre que les résultats de prédictions obtenus par l'utilisation de la représentation TMED sont plus performantes que celles obtenues par ces deux autres méthodes basées sur des mesures agglomératives d'interaction vidéo des étudiants. Ce résultat montre que la représentation proposée TMED de l'interaction vidéo des étudiants peut être utilisée pour la prédiction précoce des succès ou non des étudiants.

Les cinq contributions énumérées ci-haut peuvent se résumer : deux types de représentations des interactions des étudiants en vue des tâches d'analyses particulières, deux façons de prédire les succès ou non des étudiants une basée sur une une représentation et l'autre sur des mesures agglomératives d'écoute vidéo et en dernier une application de la méthode de comparaison des écoutes vidéo dans le domaine d'analyse textuelle des transcripts vidéo. La grande distinction entre les deux types de représentation des interactions des étudiants proposés est que SIVS est sous forme de matrice qui se situe dans le cadre de l'espace vectoriel euclidien. Les distances entre les représentations d'interaction vidéo se calculent entre les matrices comme une norme de Frobenius. Avec la représentation d'interaction vidéo SIVS, on utilise pour la classification des écoutes, un centroïde définit comme le centre d'écoute de l'ensemble de toutes les écoutes de la vidéo. A partir du centroïde de chaque vidéo, l'on pouvait en fonction de la distance entre une écoute quelconque et le centroïde d'une vidéo dire si l'écoute est de cette vidéo ou non.
Les contributions présentées ci-haut sont détaillées dans cette thèse comme suit :

1. Nous continuerons après cette introduction, au chapitre 2, dans le cadre de nos questions de recherches avec les problématiques concernant le développement de chaque technique d'analyse et d'évaluation des interactions vidéo en soulignant leur portée et leur importance. Il s'agira ici de faire ressortir les questions qui sous-tendent chaque contribution de la thèse présentée dans cette introduction.

2. Au chapitre 3, nous ferons une revue de la littérature dans le domaine de l'apprentissage en ligne, en mettant l'accent sur le développement des techniques d'analyse de l'interaction des apprenants avec les plates-formes d'apprentissage en ligne. Cette revue va donner une place de choix aux techniques utilisées dans le cadre de l'interaction de l'apprenant avec les vidéos lors de l'apprentissage.

3. Nous présenterons au chapitre 4, les données du cours que nous avons utilisées comme cadre d'étude et de test des techniques d'évaluation des interactions entre les apprenants et les vidéos. Un cours en particulier sur la plateforme Edx, portant sur deux sessions complètes a été particulièrement utilisé dans le cadre de nos recherches. Il s'agit aussi de présenter les traitements apportés aux traces brutes récoltées au niveau du serveur.

4. Nous présenterons la première technique d'analyse des interactions des apprenants avec les vidéos au chapitre 5, à savoir la représentation d'interaction vidéo (SIVS). Dans ce chapitre, nous présenterons la première série de contributions de la thèse.

5. La seconde technique d'analyse des interactions présentée au chapitre 6 est une représentation des interactions vidéo des étudiants TMED sous forme matricielle en vue de la mesure de similarité entre des interactions vidéo. Comment dire qu'une façon d'interagir avec les vidéos est semblable à une autre ? C'est la seconde et la troisième série des contributions qui seront développées dans ce chapitre 6.

6. La troisième technique d'analyse au chapitre 7 est une prédiction précoce des succès des étudiants à partir des mesures agglomératives de l'utilisation des vidéos. Nous présentrons dans ce chapitre la quatrième série des contributions évoquée ci-dessus.

7. La dernière technique d'analyse présenté au chapitre 8 est une prédiction précoce des succès ou non des étudiants en termes de réussite et d'échec à partir uniquement de la représentation de séquences d'interactions vidéo TMED. La cinquième série des contributions de cette thèse est présentée dans ce chapitre 8.

8. Une conclusion, suivie des pistes des travaux futurs, clôtureront cette thèse au chapitre 9. Il s'agira de tirer des conclusions des diverses techniques d'analyse d'interaction vidéo des apprenants et des nouvelles perspectives qu'ils inaugurent pour des travaux futurs. Cette conclusion montrera finalement que dans cette thèse, il s'agit de deux propositions de représentations d'interaction vidéo des étudiants et d'une nouvelle mesure d'écoute vidéo pour différentes tâches d'analyse des écoutes vidéo dans l'apprentissage en ligne. Enfin, il s'agit à travers cette conclusion de passer en revue les réponses apportées dans cette thèse aux questions de recherches présentées dans le chapitre 2.

CHAPITRE 2 QUESTIONS DE RECHERCHE

2.1 Contexte général

Parmi les recherches précédentes qui utilisent les traces vidéo de l'apprentissage en ligne, certains visent l'amélioration de la production des vidéos (voir Guo, Kim et Rubin 2014 ; Kim, Guo, Seaton et al. 2014, Lai, Young et N.-F. Huang 2015, Kim, Guo, Cai et al. 2014, Brame 2016), leurs effets sur l'engagement des étudiants (voir Hew 2016, Sinha, Jermann et al. 2014, Sooryanarayan et Gupta 2015) et leur style de présentation (voir Santos-Espino, Afonso-Suárez et Guerra-Artal 2016, Reutemann 2016, Chorianopoulos et M. N. Giannakos 2013 ; Chorianopoulos 2018, Rahim et Shamsudin 2019). Il semble important également de voir comment la manière d'interagir avec les vidéos pourrait caractériser l'écoute des étudiants et constater si cette écoute a un effet sur certains aspects de l'apprentissage (voir Hansch et al. 2015, Dissanayake et al. 2018).

L'enregistrement des traces des étudiants des MOOC sur un serveur permet d'avoir accès à de grande quantité de données concernant leurs activités au cours de l'apprentissage, de pouvoir les analyser et en tirer des conclusions, notamment dans la manière d'étudier (voir Hmedna, El Mezouary, Baz et Mammass 2017, Maldonado et al. 2016, Y. Shi, Peng et H. Wang 2017). Il est possible surtout de déterminer des patterns de décrochage et de prédire si l'étudiant va arriver au bout de l'apprentissage (voir Ye et Biswas 2014, Sherif Halawa, Daniel Greene et John Mitchell 2014, Yuanzhe Chen et al. 2016, X. Zhang et H. Lin 2017, X. Lu et al. 2017) . On peut aussi évaluer l'engagement ou la motivation de l'étudiant et prédire son succès ou non à partir de son utilisation du système (voir Evans, R. B. Baker et Dee 2016, Vitiello et al. 2018).

Avec la nouvelle génération des MOOC, les vidéos jouent un rôle de plus en plus central dans l'expérience d'apprentissage (M. Giannakos et al. 2014, F. Zhang, D. Liu et C. Liu 2020). Les cours en ligne sont organisés comme des séquences de vidéos produites par les instructeurs et complétées par les autres ressources, comme les textes, les problèmes et les démonstrations interactives (Höfler, Zimmermann et Ebner 2017, J. Li 2017). Dans les MOOC, on constate que la plupart des étudiants passent une bonne partie de leur temps à regarder les vidéos. Il y en a même qui sont prêts à sauter les devoirs et les exercices mais qui vont regarder systématiquement toutes les vidéos (Sharma, Jermann et Dillenbourg 2015). Étant donné la place prépondérante qu'occupent les vidéos dans les MOOC, il serait

donc pertinent de classifier les façons d'écouter les vidéos. Car la manière dont les étudiants interagissent avec les vidéos et leurs réponses aux questions post-vidéos peuvent être des indicateurs de leur détermination et de leur motivation d'aller jusqu'au bout du processus d'apprentissage.

Il serait également utile d'identifier les étudiants qui ont besoin d'une aide supplémentaire pour aller jusqu'au bout de leur apprentissage par le biais d'une prédiction précoce de leur succès ou non (Breakwell et Cassidy 2013, Khalil et Ebner 2016).

Pour cela, il serait intéressant de déterminer le pattern d'utilisation des vidéos qui caractérise particulièrement des groupes d'étudiants dans un MOOC (ceux qui vont réussir ou échouer, ceux qui ont besoin d'aide supplémentaire pour réussir (Y. Wang et R. Baker 2015, S.-F. Tseng et al. 2016, Khalil et Ebner 2017)). Ce sont les deux dernières catégories (ceux qui sont à risque d'abandonner ou qui vont échouer) qui intéressent les enseignants et les développeurs des MOOC. L'identification d'un tel groupe d'étudiants par la prédiction précoce des succès permettrait aussi aux développeurs et aux instructeurs d'apporter une aide supplémentaire à un groupe plus restreint et d'avoir un impact sur le nombre d'étudiants qui resteraient jusqu'au bout du processus d'apprentissage. Les développeurs, dans la mesure du possible, pourraient même personnaliser l'aide à apporter aux groupes étudiants pour qu'ils poursuivent l'apprentissage.

Dans la littérature, plusieurs ont utilisé des techniques pour représenter les traces vidéo des étudiants en vue de diverses sortes d'analyses comme la représentation d'interaction étudiante dans un MOOC avec la technique de chaîne de Markov (X. Ma, Schonfeld et Khokhar 2009) de chaîne semi-Markov (Faucon, Kidzinski et Dillenbourg 2016,Geigle et Zhai 2017), découvrir les patterns d'interactions en utilisant les techniques de séquence d'interactions (Boroujeni et Dillenbourg 2018,Lorenzen, Hjuler et Alstrup 2019), la technique des réseaux neuronaux pour la prédiction des performances des étudiants (Qu et al. 2019) et plusieurs autres techniques ont été développés pour répondre à divers types d'analyses. La problématique ici est de pouvoir trouver la meilleure technique de représentation des traces d'interactions vidéo pour des tâches d'analyse vidéo particulière. Dans la même tâche d'analyse vidéo, la recherche d'une technique plus efficace que celles existantes redynamise la recherche. C'est pour répondre à cet objectif de trouver les meilleures techniques que dans le cadre de cette thèse deux types de représentations des traces d'interactions vidéo des étudiants sont proposés pour améliorer les résultats des tâches d'analyses. Il s'agit entre

autre de la recherche des sections vidéo réécoutées ou pas écoutées, des comparaisons entre les écoutes, les prédictions des succès à partir du style d'écoute etc.

2.2 Représentations et algorithmes d'analyse des séquences vidéo

Les cours en ligne ouverts à grande échelle s'appuient souvent sur la vidéo comme premier choix de contenu médiatique. Compte tenu de leur importance, il n'est pas surprenant que de nombreuses études portant sur la manière dont les étudiants utilisent les vidéos dans le cadre des MOOC aient vu le jour ces dernières années. Dans cette perspective, des façons de représenter les interactions vidéo ont été mises en place pour aider à l'analyse et à la classification des types d'écoutes et des séquences vidéo des étudiants. Les représentations les plus connues et les plus utilisées sont celles basées sur les mesures cumulatives d'écoute vidéo comme le temps passé à voir la vidéo, le temps passé lorsqu'on a mis une pause à la vidéo, le nombre de fois que l'on a fait des recherches en avant et en arrière dans la vidéo, le fait d'avoir suivi ou non la vidéo toute entière (N. Li, Kidziński et al. 2015, Van der Sluis, Ginn et Van der Zee 2016, Arora et al. 2017, Dissanayake et al. 2018) et celles basées sur les séquences vidéo (Brinton, Buccapatnam et al. 2016, Wong et al. 2019).

La limite de la représentation basée sur les mesures cumulatives d'écoute dans la comparaison et dans l'agglomération des écoutes semblable vient du fait qu'elle ne tient pas compte des sections des vidéos vraiment écoutées par les étudiants mais en général du temps mis pour l'écoute. Elle ne prend pas également en compte la succession du temps des événements d'écoute mais s'attelle simplement à compter le nombre de fois que certains événements apparaissent dans l'interaction de l'étudiant.

La limite de la représentation séquentielle des événements vidéo (former une séquence d'interactions en fonction de l'apparition de l'événement d'interaction sur une ligne de temps), même si elle respecte l'apparition des événements dans le temps, ne permet pas de rendre compte des parties de la vidéo réécoutée dans sa représentation. Elle ne peut non plus rendre compte, en fonction de la longueur d'une pause dans la vidéo, de la ressemblance entre deux écoutes. Une longue pause peut en effet faire complètement différer deux séquences vidéo pourtant semblables dans leur style d'écoute.

Une limite commune de ces deux représentations (mesures cumulatives d'écoute et séquence d'interaction vidéo) tient au fait qu'elles ne peuvent rendre compte des sections des vidéos

réécoutées à plusieurs reprises ou pas du tout par l'étudiant pour pouvoir comparer deux écoutes et dire si deux étudiants ont écouté exactement les mêmes sections de la vidéo. La conséquence de cette limite peut conduire parfois fois à considérer des écoutes comme semblables alors que si l'on regarde en détail, elles seraient complètement différentes : un étudiant qui a écouté plusieurs fois une même section de la vidéo versus un autre qui a écouté toute la vidéo par exemple.

Ainsi, une représentation qui pourrait mieux rendre compte de l'écoute de l'étudiant au niveau d'événement tenant en compte le temps passé dans chaque événement mais également les sections écoutées dans la vidéo serait nécessaire pour améliorer les regroupements des écoutes semblables.

2.2.1 Représentation d'écoute vidéo sensible à la durée de la vidéo SIVS.

Nous introduisons une représentation d'écoute vidéo que nous appelons SIVS (*"Student Interactions in Vector Space"*) pour l'analyse détaillée des habitudes de visionnement des vidéos par les étudiants en comparaison avec des approches communément utilisées (*"feature based representation"*) basées sur les mesures cumulatives d'écoute.

La représentation SIVS encode les séquences d'interaction vidéo dans un espace vectoriel qui définit les mesures de distance entre elles. Un pattern d'écoute vidéo est défini comme un centroïde de séquences. Nous utilisons l'encodage SIVS et nous étudions la façon dont les étudiants interagissent avec les vidéos en analysant les différences de patterns d'interaction entre les différentes vidéos. Dans le cadre de la validation de SIVS, l'analyse de l'influence de la vidéo sur les motifs est comprise comme une tâche de classification. Pour la classification des séquences par vidéo, nous avons utilisé les approches de la machine à vecteur de soutien (SVM), de l'arbre de décision (GBM) et de la méthode du plus proche voisin (KNN). Les résultats révèlent qu'il existe une différence significative dans les patterns d'interaction (centroïde de chaque vidéo) entre les vidéos.

Cette proposition contribue à combler les lacunes ci-dessus (spécification des patterns d'écoute) en introduisant une représentation permettant d'étudier des patterns à partir de séquences, et en l'appliquant, à la mise en lumière des différences entre les façons d'interagir avec des différentes vidéos. Pour tester cette représentation, nous recherchons si les vidéos induisent des patterns d'interaction entre les étudiants.

2.2.2 QR.1 : La représentation SIVS est-elle plus performante que la représentation cumulative pour discriminer des écoutes de durée semblables ?

La représentation SIVS sensible à la durée de la vidéo contribue à combler les lacunes de la représentation cumulative en introduisant une structure permettant d'étudier des patterns à partir de séquences, et en l'appliquant, à la mise en lumière des différences entre les façons d'interagir avec les différentes vidéos. Pour tester cette représentation, nous recherchons si les vidéos induisent des patterns d'interaction.

Pour tester l'existence des patterns d'interaction dans les vidéos, nous définissons un centroïde d'écoute pour chaque vidéo. Pour un ensemble de plusieurs vidéos de durée semblables à la seconde près, nous vérifions les distances en leurs centroïdes. Ensuite, nous classifions les diverses écoutes vidéo en fonction de leurs distances par rapport aux centroïdes. Nos résultats de classification utilisant la représentation SIVS sont comparés aux résultats obtenus par les classifications en utilisant la représentation basée sur les mesures cumulatives des écoutes vidéo. La comparaison montre que les classifications basées sur la représentation SIVS sont plus performantes que celle basée sur la représentation à partir des mesures cumulatives d'écoute vidéo.

2.2.3 Représentation d'écoute vidéo insensible à la durée de la vidéo TMED

Dans les MOOC contemporains, l'analyse de la manière dont les étudiants interagissent avec ces vidéos devient essentielle pour comprendre les styles d'apprentissage et prédire les succès des étudiants.
Dans la littérature, deux types de représentation sont principalement utilisés pour l'analyse des écoutes vidéo dans le cadre des comparaisons : la représentation séquentielle des interactions vidéo (Sinha, Jermann et al. 2014) et la représentation en chaîne de Markov (Faucon, Kidzinski et Dillenbourg 2016, Bishara et al. 2017). Ces deux types de représentations sont utilisés pour comparer aussi bien les écoutes vidéo et les classer que pour faire des prédictions des succès (Tang, Peterson et Pardos 2016, H. Chen et al. 2019, L.-Y. Li et Tsai 2017).

La limite de la représentation séquentielle des interactions vidéo des étudiants pour la com-

paraison des séquences est qu'elle n'arrive pas à détecter les mêmes styles d'interactions. Il suffit d'un décalage dû à une pause plus longue dans l'une des séquences pour découvrir une grande différence entre deux séquences semblables. La ressemblance étant basée sur le calcul de distance d'édition entre les séquences, il serait nécessaire de trouver une représentation qui pourrait tenir compte du style d'écoute en prenant en compte la succession de transition.

La limite de la représentation en chaîne Markov bien qu'elle comprenne dans sa structure la succession de transition, elle ne réussit pas à prendre en compte le temps mis dans chaque état. Deux séquences peuvent être déclarées semblables avec des durées très différentes dans les états à condition qu'elles aient les mêmes probabilités de transition d'un état à l'autre. Ici la nécessité de trouver une représentation qui fasse ressortir l'importance des temps mis dans chaque état serait indispensable.

Nous introduisons une représentation de séquences d'interaction vidéo nommée TMED (*Transition Matrix and the Edit distance*) que nous comparons aux représentations existantes (séquentielle et chaîne de Markov) et nous utilisons trois classificateurs (la machine à vecteurs de support (SVM), l'arbre de décision (GBM) et le plus proche voisin (KNN)) pour comparer les séquences d'interactions des étudiants et pouvoir reconnaître et comparer des patterns particuliers d'interaction vidéo. Les résultats montrent qu'il est possible de comparer en déterminant le niveau de similarité des séquences d'interaction des étudiants.

2.2.4 QR. 2 : La présentation TMED est-elle plus performante que les approches séquentielles et de chaîne de transition pour discriminer entre différents types d'écoutes ?

La représentation TMED est-elle plus performante que d'autres représentations dans la recherche de similarité entre les écoutes vidéo ?

Discriminer les divers types d'écoute dans un ensemble des séquences d'écoute est un attribut important d'une représentation d'écoute. Il nous faudrait donc tester si la représentation TMED, plus que les approches séquentielles et de chaîne de transition, peut mieux spécifier une écoute vidéo. Il s'agit donc de pouvoir regarder la capacité de chaque représentation à pouvoir distinguer les écoutes vidéo les unes des autres et à pouvoir reconnaitre les écoutes semblables.

Nous introduisons à cet effet une méthodologie de calcul de degré de similarité entre les interactions vidéo des étudiants. En utilisant la représentation TMED, nous testons la méthodologie visant à comparer les écoutes vidéo des étudiants provenant des interactions avec divers vidéos. Pour valider la méthodologie, des cas prototypes ont été choisis connaissant les résultats d'avance pour voir la capacité de la représentation TMED par rapport aux représentations séquentielles et de chaîne de Markov à pouvoir déterminer la similarité entre des séquences. Les résultats montrent que la représentation TMED est à mesure de mieux déterminer la similarité entre les écoutes vidéo que les deux autres.

Nous avons appliqué cette méthodologie pour déterminer le degré de similarité entre des textes. Nous avons pu extraire le vocabulaire commun à tous les textes pour construire un texte atypique du corpus des textes considérés. L'application dans un cadre particulier des transcrits des vidéos d'interviews des migrants en Europe a été d'une grande utilité pour produire un texte atypique pour des analyses sociologiques par des experts.

2.3 Caractère unique des traces d'interaction d'un étudiant et d'une vidéo

Dans la perspective de tester les diverses sortes de représentations des interactions vidéo à l'hypothèse que les traces d'écoutes sont uniques à chaque étudiant et à chaque vidéo, nous allons comparer la capacité de chaque représentation à rendre compte de cela. Les deux représentations que nous avons proposées (SIVS et TMED) vont être comparées dans leur capacité à rendre compte de l'unicité des traces d'un étudiant (SIVS et TMED) et d'une vidéo (TMED) en comparant leurs résultats à ceux obtenus avec les représentations de la littérature. La représentation SIVS va ainsi être comparée à la représentation basée sur les mesures cumulatives (car répondant aux mêmes types d'analyses que cette dernière). La représentation TMED quant à elle va être comparée aux représentations en chaîne de Markov et la représentation en séquence d'interaction vidéo.

2.3.1 QR. 3 : Un étudiant possède-t-il un style d'écoute qui lui est propre ? Une vidéo possède-t-elle aussi une signature d'écoute ?

Nous utilisons la représentation SIVS pour montrer notamment l'existence d'un pattern d'interaction autour des vidéos. Nous avons montré que cette représentation plus que celle basée sur les mesures cumulatives d'écoute vidéo est capable de reconnaitre à travers la représentation SIVS d'écoute, la vidéo avec laquelle l'étudiant est en train d'interagir.

Nous utilisons également la représentation TMED pour montrer que cette représentation est capable plus que deux autres représentations sauvent utilisées dans la littérature (la représentation en chaîne de Markov et la représentation en séquence d'interaction vidéo) pour identifier les étudiants à partir de la représentation de leurs interactions vidéo. Avec la représentation TMED, mieux que d'autres, on est capable à partir d'une représentation des interactions vidéo quelconque de déterminer de quel étudiant provient cette interaction.

Utilisant la même représentation TMED nous avons investigué si une vidéo en particulier impose aux étudiants une façon d'interagir avec elle. En utilisant les classificateurs mentionnés dans la section 2.2.3, nous avons testé la capacité des représentations des interactions vidéo à reconnaitre avec quelle vidéo les étudiants étaient en train d'interagir. Les résultats montrent que la représentation TMED est la mieux capable de pouvoir identifier la vidéo avec laquelle un étudiant interagit à partir de sa représentation d'interaction vidéo comparés aux résultats obtenus avec les représentations séquentielles et en chaîne de Markov.

2.4 Prédiction du succès à partir des traces d'écoute vidéo

Dans la littérature, la question de la mesure des interactions vidéo des étudiants pour comparer leur utilisation des ressources en ligne en général mais surtout l'utilisation des vidéos a commencé par le développement des outils visuels. En particulier, C. Shi et al. 2015 ont développé un outil appelé VisMOOC qui peut aider à voir de manière visuelle comment les étudiants utilisent les ressources du MOOC. Ils ont développé plusieurs vues possibles à savoir une vue d'ensemble des différences de clics entre les vidéos, la vue par contenu pour montrer les variations temporelles du nombre total de chaque type d'action de clic le long de la ligne de temps de la vidéo, la vue du tableau de bord pour afficher diverses informations statistiques telles que des informations démographiques et temporelles.

Brochenin et al. 2017, ont étendu ce concept de calcul de l'utilisation des ressources pour prédire les décrochages. Afin d'arriver à ce but, ils effectuent un aperçu de la manière dont les ressources éducatives sont utilisées et considèrent l'abandon comme un comportement atypique à identifier. Ils ont développé un prototype, appelé RUAF, qui peut être appliqué aux données et leurs résultats mettent en évidence des modèles montrant comment les apprenants utilisent les ressources. Dans le cadre spécifique de visualisation de l'utilisation quotidienne

des vidéos dans un MOOC, He, Dong et al. 2019 ont développé un système appelé VUC qui permet de visualiser l'utilisation de la ressource vidéo chaque jour par les étudiants. Par ailleurs, ils ont mis en place des mesures agglomératives d'utilisation des ressources vidéo pour la prédiction des succès des étudiants (He, Zheng et al. 2018).

2.4.1 Les mesures agglomératives d'écoutes vidéo

Avec des milliers d'étudiants qui s'inscrivent à des cours en ligne chaque année, il est très intéressant pour les développeurs et les instructeurs de MOOC, de prédire, dès les premières semaines d'un cours, quels sont les étudiants qui sont susceptibles de réussir ou d'échouer le cours. Par exemple, des indicateurs qui pourraient prédire le nombre d'étudiants en mesure de réussir ou échouer le cours. Une telle prédiction peut, par exemple, alimenter les tableaux de bord des instructeurs, les aider à adapter la structure et le matériel de leurs cours, ou déclencher une aide et adapter les interventions à des groupes d'étudiants spécifiques. Notre investigation se concentre sur trois mesures agglomératives de la façon dont les étudiants interagissent avec les vidéos du MOOC définies par He, Zheng et al. 2018 afin de prédire quels groupes d'étudiants qui réussiront ou échoueront le cours. Ces trois mesures agglomératives sont : le taux d'assiduité des étudiants (AR), le taux d'utilisation (UR) et le taux de visionnement (WR).

La limite de leur perspective est qu'elle est étroitement liée dans sa procédure de prédiction à des éléments graphiques pour diviser les groupes d'étudiants. En effet, la division entre les étudiants dont on prédit qu'ils réussiront et ceux qui vont échouer se fait de manière graphique et peut éventuellement introduire un aspect subjectif. D'où la nécessité de trouver une méthode plus objective de division de groupe d'utilisation de ressources vidéo pouvant prédire les deux classes d'étudiants. Les mesures agglomératives qu'ils ont définies étant objectives, comment les utiliser pour prédire le succès ou l'échec des étudiants ?

2.4.2 QR. 4 : Les mesures agglomératives permettent-elles de prédire avec précision les chances du succès des étudiants ?

Pour dépasser la limite de la dépendance vis-à-vis de l'aspect graphique, nous introduisons une nouvelle mesure que nous appelons l'index de visionnage (WI) que nous utilisons pour séparer de façon objective les étudiants par groupes d'utilisation vidéo. Ces mesures agglomératives sont prises après la première semaine et au milieu du cours. Les résultats montrent

que ces mesures peuvent être très efficaces pour prédire quels sont les étudiants qui réussiront ou échoueront le cours.

L'apport du WI dans la prédiction précoce du succès des étudiants est qu'elle enlève la dépendance vis-à-vis de la composante graphique assujettie à l'interprétation subjective du chercheur pour reposer uniquement sur des mesures objectives. Ainsi la prédiction du succès ou de l'échec d'un étudiant dépend uniquement des mesures agglomératives de son mode d'utilisation des ressources vidéo.

2.4.3 Utilisation de la représentation TMED pour la prédiction précoce du succès

Dans la problématique de prédiction précoce du succès à partir de la représentation TMED, il s'agit d'explorer les liens entre la représentation des interactions vidéo et le succès à la fin du cours. En d'autres termes, il est question de tester la manière dont la représentation TMED fait ressortir les styles d'interactions qui conduisent au succès ou à l'échec. Dans la littérature plusieurs études se sont penchées sur la prédiction du succès des étudiants dans le cours utilisant plusieurs types de représentation d'interactions (Ren, Rangwala et Johri 2016, Wan et al. 2017, T.-Y. Yang et al. 2017, Conijn, Van den Beemt et Cuijpers 2018, Brinton, Buccapatnam et al. 2016, Malhotra 2020).

La limite principale de toutes ces représentations pour la prédiction des succès des étudiants est que toutes ces représentations ont besoin d'un traitement particulier avant de pouvoir être utilisées comme facteur de prédiction du succès. De plus, ces représentations sont combinées à d'autres facteurs dans le processus de prédiction. D'où la nécessité de trouver une représentation qui pourrait en elle-même être le seul facteur de prédiction du succès de l'étudiant. Ceci pourrait réduire éventuellement le temps de calcul et réduire les données à traiter en vue de la prédiction.

2.4.4 QR. 5 : Comment la représentation TMED se compare-t-elle aux autres méthodes de prédiction de succès ?

Nous souhaitons déterminer, à partir d'une représentation particulière d'interaction vidéo (TMED introduit précédemment), s'il s'agit d'une interaction de réussite ou d'échec. Cette

représentation de l'interaction vidéo des étudiants est basée sur une matrice de transition transformée qui expose à la fois la succession des états comme dans une chaîne de Markov et en même temps, le temps passé dans chaque état. Nous avons déjà montré dans la section 2.3.1 sa sensibilité de la représentation en testant sa capacité de distinguer l'interaction vidéo des étudiants les uns par rapport aux autres. Si les étudiants ont une signature individuelle dans leurs interactions vidéo et que la représentation est capable de la reconnaître, nous supposons qu'elle est également assez puissante pour détecter si ces interactions sont propres à une personne qui va réussir ou échouer le cours. Par conséquent, la représentation que nous proposons sera validée par sa capacité à reconnaître une interaction particulière de l'étudiant avec la vidéo conduit au succès ou à l'échec. Ensuite, nous cherchons à savoir si la représentation proposée, combinée à un algorithme de classification standard, peut prédire les résultats de l'étudiant à la fin du cours, en termes de réussite ou d'échec. Nos résultats montrent que l'utilisation de la représentation TMED pour les prédictions précoces du succès des étudiants a une bonne performance de prédiction.

Nous avons comparé pour cela, la performance de prédiction du succès ou non des étudiants obtenus en utilisant la représentation TMED aux performances des deux méthodes de prédiction du succès des étudiants basées sur les mesures agglomératives d'écoute vidéo. Cette comparaison montre que l'utilisation de la représentation TMED pour prédire le succès des étudiants est plus performante que l'utilisation des mesures agglomératives d'écoute vidéo. Cette comparaison montre également l'importance de la représentation TMED par sa capacité de pouvoir être utilisé pour plusieurs tâches d'analyses d'interactions des étudiants dans un cours en ligne.

CHAPITRE 3 REVUE DE LA LITTÉRATURE

3.1 Introduction

L'analyse des traces des étudiants à travers les événements recueillis sur les serveurs a fait l'objet de plusieurs investigations. Ces diverses analyses ont conduit à une meilleure compréhension de la façon dont les étudiants interagissent avec les systèmes d'apprentissage en ligne. Ces analyses ont permis également non seulement d'améliorer l'expérience d'apprentissage mais aussi d'améliorer la façon d'évaluer les étudiants. Les divers domaines de recherches qui ont été identifiés par Yousef, Chatti et U. Schroeder 2014 résument bien dès 2014 les divers types de travaux menés sur les traces d'interaction des étudiants dans les systèmes d'apprentissage en ligne en général et des MOOC en particulier (voir pour la période 2007-2017, Poquet et al. 2018, l'analyse des prédictions pour la période 2008-2018 Moreno-Marcos et al. 2018). L'analyse des données des interactions avec les systèmes d'apprentissage en ligne a pour but d'améliorer l'implémentation des systèmes d'apprentissage en ligne et de rendre l'expérience apprentissage plus accessible et plus agréable aux étudiants. Nous passerons en revue les publications qui ont utilisé principalement les traces vidéo des étudiants pour analyser l'expérience des systèmes d'apprentissage en ligne.

Nous allons dans un premier temps présenter comment les recherches qui ont utilisé les traces vidéo touchent à plusieurs aspects de l'apprentissage comme : l'efficacité du système et des étudiants, les méthodes d'enseignement, le design du système. On qualifie ici d'efficacité la représentation et l'analyse des interactions vidéo et les prédictions des succès qui sont deux volets de nos recherches dans le cadre de cette thèse. Dans les deux volets de l'efficacité nous allons explorer plusieurs recherches qui ont été faites. Nos recherches visent donc à améliorer les techniques existantes dans le cadre des analyses des interactions vidéo des étudiants. Nous proposons deux représentations d'interactions vidéo qui améliorent les résultats des d'analyses existantes dans le domaine de l'identification des interactions, leur degré de similarité et leur utilisation pour les prédictions précoces des succès des étudiants. Il s'agira dans cette revue de littérature de passer en revue les principales publications qui ont trait à ces divers champs d'investigations qui touchent à l'analyse des traces vidéo des étudiants dans les systèmes d'apprentissage en ligne. Nous mettrons un accent particulier à l'analyse les interactions vidéo des étudiants et mettrons de côté les études qui utilisent d'autres types de traces d'interaction avec les plates-formes d'apprentissage en ligne.

Dans la seconde partie de cette revue de littérature nous allons présenter les recherches liées directement à chacune des problématiques de recherches que nous abordons dans cette thèse. Il s'agit de présenter l'état de l'art lié à chacune des cinq contributions de cette thèse qui répondent aux cinq questions de recherche du chapitre 2.

3.2 Analyse de l'apprentissage utilisant des interactions vidéo des étudiants

L'analyse des traces vidéo des étudiants à travers les données collectées par les serveurs qui hébergent les systèmes d'apprentissage en ligne a principalement pour but d'améliorer pour les étudiants l'expérience d'apprentissage et d'aider les instructeurs et les développeurs à mieux gérer les divers groupes d'étudiants pour pouvoir accommoder le plus grand nombre. A cet effet, nous pouvons trouver dans la littérature des grandes lignes de recherche qui se dégagent dans l'analyse des traces vidéo. Nous allons principalement nous intéresser aux recherches qui utilisent les traces vidéo des étudiants pour aider la communauté d'apprentissage en général.

3.2.1 L'évaluation de l'efficacité par l'analyse des traces vidéo

Beaucoup de recherches utilisant des traces vidéo ont comme objectif de pouvoir, à travers l'étude des interactions des étudiants avec les vidéos, trouver les diverses pistes d'améliorations du système d'apprentissage et de l'expérience d'apprentissage tout en favorisant une meilleure performance des étudiants (D. Zhang et al. 2006, D. Zhang et al. 2006, Liaw 2008, Delen, Liew et Willson 2014). Dans l'analyse des traces vidéo avec le but d'améliorer des aspects de l'apprentissage touchent principalement les résultats d'apprentissage, les formes d'enseignement, l'amélioration de la présentation des contenus vidéo, le changement de style vidéo.

Résultat d'apprentissage

Certains chercheurs ont analysé les traces vidéo des étudiants pour évaluer les apprentissages du cours. Cela concerne aussi bien l'évaluation des connaissances acquises, que la prédiction de succès et de décrochage ou simplement évaluer le niveau d'engagement des étudiants à travers les traces vidéo et l'influence du style vidéo dans l'apprentissage (prédiction de succès : Dekker, Pechenizkiy et Vleeshouwers 2009, niveau d'engagement :Molinari et al. 2016, prédiction de décrochage :Eagle et Barnes 2014a, Eagle et Barnes 2014b,S. Halawa, D. Greene et J. Mitchell 2014,J. Whitehill et al. 2015, influence du style vidéo : Ilioudi, M. N. Giannakos et Chorianopoulos 2013). En général on peut les

décrire comme des connaissances, des compétences (C.-c. Lin et Y.-f. Tseng 2012, Hsu et al. 2013) et les capacités que les apprenants doivent atteindre grâce au processus d'apprentissage (Merkt, Weigand et al. 2011 ; Merkt et Schwan 2014, Zahn et al. 2012, Ozan et Ozarslan 2016). Certaines études montrent par l'étude des traces d'ailleurs que l'enseignement à travers les vidéos n'a aucune différence statistiquement significative par rapport aux autres formes d'enseignements plus traditionnelles équivalentes (Donkor 2010, Comeaux 2005, U. D.-I. U. Schroeder p. d., Yousef, U. Schroeder et Wosnitza 2015). Certains par l'analyse des interactions vidéo étaient capables de percevoir les difficultés des étudiants à comprendre les vidéos ou simplement à identifier des décrochages au niveau de la vidéo (N. Li, Kidzinski et al. 2015 ; N. Li, Kidziński et al. 2015, Kim, Guo, Seaton et al. 2014). La spécification du profil étudiant s'est faite à travers les interactions vidéo (voir Belarbi et al. 2019, Man, Azhan et Hamzah 2019). L'identification des patterns d'écoutes et des spécificités d'écoutes à travers les interactions vidéo a été également analysée avec beaucoup d'attention (Sinha, N. Li et al. 2014, N. Li, Kidziński et al. 2015, Brinton, Buccapatnam et al. 2016, Shridharan et al. 2018).

Des chercheurs ont identifié le degré de satisfaction des étudiants à travers les traces interaction vidéo (D. Zhang et al. 2006, Liaw 2008, Kuo et al. 2014, Eom et Ashill 2016).

3.2.2 Les formes d'enseignement par l'analyse des traces vidéo

Certaines études d'analyse des traces vidéo des étudiants dans un cours en ligne avaient entre autres pour but d'améliorer la méthode d'enseignement. Plusieurs aspects de la méthodologie d'enseignement ont été étudiés. Nous aborderons quelques-uns d'entre eux qui ont fait l'objet des investigations sérieuses de la part des chercheurs à travers l'analyse des traces vidéo des étudiants.

La reproduction du micro-enseignement dans le cadre des cours en ligne à travers les vidéos en est une. Ce qu'on qualifie de micro-enseignement est une méthode d'enseignement caractérisée selon la taille et la durée de la classe (par exemple, de quatre à neuf apprenants dans une classe qui est tenue pendant cinq à dix minutes) reproduite dans le contexte d'apprentissage en ligne en utilisant les podcasts vidéo pour la rétroaction rapide avec chaque apprenant (Brantley-Dias et al. 2008, Zee 2005). C'est le constat que les vidéos de courtes durées (entre 4 à 5 minutes) donnent plus de flexibilité dans les micro-leçons ouvrant la porte au micro-enseignement (voir Brantley-Dias et al. 2008, Seidel, Blomberg et Renkl 2013, Fisher et Burrell 2011). Une telle approche est souvent très utilisée par les étudiants à

l'approche des évaluations.

L'analyse des traces vidéo des étudiants qui montrent que ces derniers ont tendance à aller écouter les parties des vidéos les plus importantes, la nécessité de rendre disponible un résumé de chaque vidéo se fait sentir. Le résumé vidéo est une stratégie d'enseignement qui extrait des parties importantes de la vidéo du cours pour les mettre à la disposition des étudiants (Fu et al. 2008, Fang et al. 2011). En particulier W.-H. Chang, J.-C. Yang et Y.-C. Wu 2011 ont conçu une façon de résumer les vidéos par mots clés dans une plate-forme d'apprentissage de synthèse appelée KVSUM qui fournit un nuage de mots-clés comme substitut textuel pour aider les apprenants à organiser l'information des vidéos et les améliorer pour suivre les vidéos et la réduction du temps d'apprentissage (voir aussi Choudhari et Bhalla 2015, Baralis et Cagliero 2015, Mohd Kamal et al. 2019).

En analysant les traces des interactions vidéo des étudiants, la nécessité de développer des stratégies pour orienter le cours de telle sorte que l'étudiant soit au centre du projet éducatif du cours se fait ressentir. En effet, beaucoup de systèmes d'apprentissage en ligne ont une stratégie centrée sur l'enseignant y compris plusieurs MOOC. Des études montrent même que 15% seulement des cours en ligne utilisent la stratégie centrée sur l'étudiant. Cette stratégie d'enseignement fait en sorte que le cours ne dépend pas de l'enseignant comme fournisseur du contenu. Elle donne à l'étudiant l'espace nécessaire pour être à son tour actif dans l'environnement d'apprentissage. Elle devient un environnement interactif où chaque apprenant peut avoir le soutien des autres participants pour prendre des décisions en faisant appel au sens et au jugement critique (Gainsburg 2009, Verst 2010, Baig 2012, Smyth 2011).

Certains chercheurs ont fait, à partir de l'analyse des interactions, des propositions de modification du design de la plate-forme du cours en ligne. Il s'agit de proposer un design plus facile en termes d'accessibilité à l'apprenant. Il est question dans certains cas de proposer une interface persuasive pour aider l'apprenant à s'engager dans une interaction intense avec les vidéos (voir Brangier et Michel C Desmarais 2013). Deux axes en particulier peuvent résumer les propositions dans le domaine du design.

Dans certaines situations, il a fallu ajouter une note, commentaire, explication et majoration de présentation à un document, une image ou une vidéo dans le but d'encourager les étudiants à s'engager davantage dans les interactions avec les vidéos. Ici on peut voir aussi l'ajout des transcrits des vidéos disponibles aux étudiants (Rich et Hannafin 2009, Marsh et N.

Mitchell 2014). L'annotation des vidéos fait également référence aux notes supplémentaires ajoutées à la vidéo sans modifier la ressource elle-même, ce qui facilite la recherche, la mise en évidence des sections vidéo, l'analyse, l'extraction et le retour d'information (Khurana et Chandak 2013, Chatti et al. 2016, Yousef, Chatti, Danoyan et al. 2015, Martin, Charlton et Connor 2016, Shama et Prakash 2018). Dans le cas des vidéos, l'annotation permet également l'indexation, la discussion, la réflexion et la conclusion du contenu (M. Wang et al. 2007, Schroeter, Hunter et Kosovic 2003, Risko et al. 2012, Dawson et al. 2012). Par exemple, Colasante 2011 a examiné l'intégration dans une vidéo l'outil d'annotation qu'ils appellent "MAT" dans l'apprentissage et l'évaluation des activités d'une classe de troisième année du cours "Éducation physique" à l'université RMIT. Cet outil a permis aux apprenants de sélectionner et annoter des parties d'une vidéo. Ces annotations sont ensuite utilisées par les étudiants et les enseignants pour discuter, recevoir des commentaires, réfléchir et évaluer leurs pratiques d'apprentissage et d'enseignement. Les résultats ont montré que le MAT était efficace pour recevoir les réactions des enseignants et des pairs. Mais, certains problèmes concernant la qualité de la contribution collaborative des pairs ont été notés.

A travers des analyses des interactions des étudiants, plusieurs études ont fait des suggestions pour pouvoir augmenter les interactions des étudiants avec le système d'apprentissage en ligne (Chunwijitra et al. 2012). Des outils bien connus comme celui utilisé pour synchroniser un flux vidéo avec la présentation par le biais de la synchronisation des clips vidéo (Chunwijitra et al. 2012), celui qui est capable de résumer le contenu par l'extraction des informations sommaires des vidéos de conférences et le fournir automatiquement aux apprenants (J. C. Yang et al. 2009, Ngo, Y.-F. Ma et H.-J. Zhang 2005, Viguier et al. 2015, Wouters, Tabbers et Paas 2007, Domagk, Schwartz et Plass 2010).

3.3 Les analyses des traces vidéo liées à nos investigations.

Nous allons dans cette seconde partie de notre revue de littérature présenter les publications ayant trait à chacune de nos questions de recherche. Des recherches ont été menées pour répondre à chacune de nos questions d'une manière ou d'une autre. Le but de nos investigations est d'apporter des réponses plus efficaces que celles existantes pour améliorer la représentation et l'analyse des traces vidéo des étudiants dans le cadre des cours en ligne. Les solutions actuelles à toutes ces questions de recherche sont ouvertes à des améliorations pour continuer la recherche des nouvelles techniques plus efficaces. Nous présenterons ces recherches en suivant les questions de recherche qui constituent la base de nos cinq contributions dans le

cadre de cette thèse (voir l'introduction).

3.4 Séquence d'activités vidéo des étudiants et représentations (QR.1 et QR.2)

Le regroupement des activités des étudiants sous forme de séquence sur une ligne de temps a été effectué par Michel Desmarais et François Lemieux 2013 ; François Lemieux, Michel C Desmarais et P.-N. Robillard 2014. Les activités sont définies comme les réponses aux exercices, la navigation dans les exercices, la navigation dans les notes, la pause sans activité dans les 5 dernières minutes, la réponse à un exercice, la visualisation du score et l'activité de connexion. Ils ont étudié la séquence des activités et visualisé les schémas d'étude de l'apprentissage des mathématiques au collège. Nous présenterons en détail cette approche de représentation séquentielle d'interaction vidéo dans cette section.

Le concept de séquence d'activité sur une ligne de temps en termes d'activités a été exporté vers une forme de représentation des interactions vidéo spécifiques. En utilisant les séquences d'événements vidéo des étudiants, un regroupement d'écoutes des étudiants a été fait pour les MOOC par N. Li, Kidziński et al. 2015 dans le but d'explorer le lien entre les interactions vidéo et la difficulté vidéo perçue dans la compréhension de ces vidéos. Ici, ils ont utilisé les événements vidéo uniquement par le biais de la plateforme pour générer chaque séquence d'interactions des étudiants.

Une autre approche de représentation des interactions vidéo des étudiants est celle basée sur les mesures cumulatives des interactions. Chaque événement vidéo de l'étudiant est exprimé par sa durée cumulative dans l'interaction de l'étudiant. Cette forme de représentation très utilisée dans la littérature sera également présentée dans cette section.

Une troisième approche de représentation des interactions vidéo des étudiants est celle sur les probabilités des transitions entre les événements à l'image d'une construction de chaîne de Markov. Une telle représentation est sensible au style d'interactions car cela donne une place importante aux transitions d'évènements vidéo.

Nous allons présenter dans un premier temps l'approche de représentation séquentielle des interactions vidéo. Ensuite, nous présenterons l'approche de représentation par mesures cumulatives des activités d'interaction vidéo des étudiants.

3.4.1 Représentation d'interaction vidéo basée sur les séquences d'événements

Les interactions vidéo peuvent être obtenues à partir d'événements captés et présents sous forme d'informations séquentielles des sessions d'interaction, et les ordonnent sous forme de ligne de temps en fonction de leur apparition.

Cette approche consiste à caractériser une interaction vidéo comme une séquence d'événements, ou d'activités Michel Desmarais et François Lemieux 2013 ; Lemieux, Desmarais et P. Robillard 2013 ; Sinha, Jermann et al. 2014. Pour Lemieux, Desmarais et P. Robillard 2013 ; Michel Desmarais et François Lemieux 2013, il s'agissait, dans le cadre de l'évaluation de l'apprentissage autonome, d'étudier les traces d'un guide d'étude pour apprentissage autonome des mathématiques et de pouvoir dégager la façon dont les apprenants autonomes utilisent le guide. Le guide d'étude en question est un répertoire de 1030 exercices de mathématiques accompagné de près de 150 pages imprimées des notes en ligne qui expliquent la théorie nécessaire pour les exercices. Il est organisé en dix thèmes de mathématiques pré-universitaires et les exercices sont classifiés selon 144 notions, avec, en moyenne, 8 exercices par notion. Les auteurs ont réussi à classifier et visualiser sept types de séquences d'activités menées par 119 utilisateurs du système, dont 53 ont tenté de résoudre les exercices. Ils ont résumé ces sept types d'activités comme suit : réponse à un exercice, furetage dans les pages d'exercices, furetage dans les pages de notes, aucun événement depuis 5 minutes ou plus, résolution d'exercice, furetage dans la section des résultats, page d'enregistrement. L'utilisation était laissée au libre choix des étudiants de première année d'université. L'utilisateur pouvait demander la réponse à chaque exercice et indiquer au système s'il a réussi ou non l'exercice. Le guide garde la trace des exercices ayant été déclarés réussis par l'utilisateur pour pouvoir juger de la progression de ce dernier. Les chercheurs pouvaient, à travers la séquence d'utilisation de chaque utilisateur, faire une analyse spectrale et obtenir les patterns d'utilisation en groupe du même type d'utilisation. L'avantage d'une telle méthode porte aussi sur sa représentation qui peut donner en un coup d'œil le mode d'utilisation de tous les étudiants (François Lemieux, Michel C Desmarais et P.-N. Robillard 2014, agrégation spectrale avec Von Luxburg 2007 et C. Li et Yoo 2006, visualisation des séquences d'état avec Gabadinho et al. 2011). La première caractérisation, que nous appelons *fonctionnalité*, conserve des caractéristiques numériques ou des facteurs booléens qui peuvent être utilisés pour prédire des variables d'intérêt telles que l'engagement, le style d'apprentissage, etc. Dans la dernière caractérisation, *sequence-based*, les interactions sont des séquences ordonnées dans le temps entre lesquelles nous pouvons mesurer les distances. Notre approche, qui se situe dans cette deuxième catégorie, sera exposée plus loin.

L'analyse des interactions basée sur les séquences, dans l'étude de Sinha, Jermann et al. 2014, est particulièrement pertinente et originale. Au lieu d'utiliser des événements bruts tels que *"pause"* et *"play"*, ils examinent également des paires et des n-tuples d'événements. Les auteurs développent un encodage basé sur n-gram des événements de visionnage vidéo et utilisent des sous-séquences d'événements en plus des événements individuels. Ils tentent d'utiliser les niveaux d'interaction n-gram et de prédire l'engagement, d'identifier les états d'excitation et la probabilité d'abandon. Leur approche atteint un bon niveau de précision de prédiction de l'engagement, du click suivant, du désengagement de la vidéo et le désengagement du cours avec un score Kappa qui varie entre 0,5 et 0,9 pour les différentes prédictions. Une approche similaire a également été appliquée à la résolution des problèmes liés aux données du journal de bord des étudiants par Sinha, N. Li et al. 2014.
D'autres cas d'analyse d'interaction basée sur des séquences ont été menés sur les données des journaux d'utilisation des étudiants de différentes applications d'apprentissage (Michel Desmarais et François Lemieux 2013 ; Klingler et al. 2016 ; Hao, Shu et Davier 2015). Ces approches utilisent une mesure de la distance (généralement la distance minimale d'édition, ou distance de Levenshtein) pour regrouper les sessions. Les séquences d'événements sont transformées en sessions d'activités représentées par des vecteurs. Chaque unité du vecteur représente une durée de temps et correspond à une activité. Le regroupement de ces vecteurs permet d'obtenir une vue synthétique des différents types de sessions.

L'approche que nous proposons dans cette thèse s'inspire des approches d'interaction basées sur les séquences et utilise une mesure de la distance entre les sessions de visionnage. Cependant, la similarité entre les sessions est basée sur la combinaison de la distance de Levenshtein et la distance d'édition. Elle définit un modèle type d'interaction d'une vidéo dans l'espace vectoriel qu'on qualifie de centroïde. De plus, une étiquette de classe est représentée par un point dans cet espace correspondant à la séquence prototype de l'étiquette de classe. Nous présentons plus loin les moyens par lesquels une séquence prototype d'interactions peut être définie et nous démontrons comment elle peut identifier le modèle de visionnement de chaque vidéo spécifique.

3.4.2 Représentations basées sur les transitions d'écoute vidéo

Dans le souci de vouloir tenir compte de la succession des événements vidéo des étudiants dans leurs interactions la représentation basée sur la chaîne de Markov ou semi-chaîne de Markov fut mise en place (Geigle et Zhai 2017, Geigle et Zhai 2017, Faucon, Kidzinski et

Dillenbourg 2016, H. Chen et al. 2019). Il s'agit de représenter les interactions vidéo des étudiants sous la forme d'une matrice de transition en exprimant chaque transition en termes de probabilité de chance de passer d'un état à l'autre.

La limite d'une telle approche, bien qu'elle soit bonne dans la comparaison des styles d'interactions en termes de transitions, elle reste très limitée dans sa capacité à retrouver le degré d'engagement et d'intensité d'interactions vidéo. Notamment, deux séquences d'interactions peuvent avoir une même matrice de transition sans avoir les mêmes temps d'activité d'interactions.

L'approche de représentation TMED que nous proposons dans cette thèse, veut garder l'avantage de la représentation en chaîne de Markov sur le style d'interactions tout en tenant en compte le temps passé dans chaque état. Cette approche de représentation gagne l'avantage du style apporté par sa structure en chaîne de Markov mais également de la durée dans les états qui constitue l'avantage des représentations séquentielles.

3.4.3 Représentations basées sur les mesures cumulatives d'écoute vidéo

Le niveau élevé d'abandons dans le système MOOC a conduit à de nombreuses études pour étudier comment les apprenants pourraient être classés entre ceux qui resteront et ceux qui s'arrêteront avant la fin en fonction de la séquence d'interactions y compris les interactions vidéo et comment intervenir pour aider les apprenants à continuer à utiliser le système (N. Li, Kidziński et al. 2015, Brinton, Buccapatnam et al. 2016, Kloft et al. 2014 ; D. Yang et al. 2013, Jacob Whitehill et al. 2015). Le but était d'évaluer l'engagement des étudiants et de prédire leurs performances (Ramesh et al. 2013, W. Jiang, Schenke et O'dowd 2014). Certains ont étudié les configurations structurelles dans les séquences d'activités d'apprentissage des étudiants dans le MOOC (Sinha, Jermann et al. 2014 ; Sinha, N. Li et al. 2014). Certaines de ces études ont pu trouver des processus d'apprentissage à partir des comportements des étudiants et en utilisant l'extraction des modèles pour classer les étudiants (Emond et Buffett 2015, Seaton, Nesterko et al. 2014).
Notre approche est comparée à l'approche de codage basée sur les mesures cumulatives d'écoute vidéo (temps mis pour visionner la vidéo, temps mis pour les pauses vidéo, nombre de fois de recherche dans la vidéo en avant et en arrière etc). Il s'agit ici d'une approche commune permettant de détecter des modèles uniques et de prédire des facteurs d'intérêt, tels que le niveau d'engagement des étudiants, l'abandon de la vidéo et l'achèvement ou l'abandon des cours (Kim, Guo, Seaton et al. 2014, Sinha, Jermann et al. 2014). Nous pouvons

résumer les cinq attributs vidéo qui peuvent intervenir dans l'encodage d'une séquence de visionnage des vidéos comme suit :

(i) **Durée de la vidéo :** L'encodage des interactions des étudiants peut être affecté par la longueur de la vidéo. Certaines études ont montré que les vidéos plus courtes sont beaucoup plus attrayantes (Guo, Kim et Rubin 2014). Dans la représentation SIVS proposée dans le cadre de cette thèse, la longueur de la vidéo est présente dans la structure de l'encodage à travers la taille de la matrice de tenseur dans l'encodage de séquences d'interaction des étudiants. De plus, dans cette étude, nous avons utilisé des vidéos de même durée afin de les comparer à la seconde près pour éviter toute discrimination entre les vidéos en fonction de la durée de chacune elles.

(ii) **La durée totale de lecture de la vidéo par l'étudiant :** Il s'agit de rendre compte dans l'encodage de la durée totale (durée cumulée) pendant laquelle l'étudiant a visionné une vidéo, y compris s'il écoute la même section de la vidéo plus d'une fois. En général, cette mesure cumulative est exprimée en pourcentage de la durée de temps de lecture de la vidéo par rapport à la durée du temps passé à interagir avec la vidéo par l'étudiant. La limite d'une telle approche est de ne pas pouvoir rendre compte dans l'encodage de la section de la vidéo vraiment écoutée par l'étudiant. L'approche proposée dans le cadre de cette thèse cherche à dépasser cette limite.

(iii) **La durée totale de pause de la vidéo par l'étudiant :** Selon des études, des pauses fréquentes ou plus ou moins longues peuvent révéler certains aspects d'une vidéo, comme son niveau de difficulté, ou le niveau de l'engagement (N. Li, Kidzinski et al. 2015). Rendre compte donc de la durée des pauses dans l'encodage des interactions peut aider dans ces types d'investigations. La limite des encodages actuels est l'impossibilité de voir la fréquence, le lieu d'occurrence et la durée à chaque occurrence en même temps dans l'encodage. Les encodages cumulent généralement le temps de pause dans toute la durée des interactions de l'étudiant avec une vidéo. C'est pour dépasser cette limite que l'encodage SIVS proposé dans cette thèse tient compte de l'occurrence des événements "pause" dans la vidéo, du lieu de l'occurrence dans la vidéo et ainsi que de la durée à chaque occurrence.

(iv) **Nombre de fois où l'étudiant a reculé la vidéo :** La recherche à rebours pour un étudiant montre l'importance de la lecture d'une partie de la vidéo ou l'importance de ces sections de la vidéo dans le processus d'apprentissage. L'encodage actuel nous informe sur le nombre de fois que l'étudiant a reculé la vidéo mais ne nous renseigne pas jusqu'où le recul s'est fait. Mieux, est-ce que l'étudiant réécoute les mêmes sections lors du recul ou des sections différentes ? La représentation SIVS répond à ces interrogations en montrant dans sa structure les lieux où ont eu lieu les reculs et jusqu'où l'étudiant a reculé la vidéo à chaque fois.

(v) **Nombre de fois où l'étudiant a avancé dans la lecture de la vidéo :** La recherche vers l'avant indique le niveau de désengagement de l'étudiant en sautant des sections de la vidéo. Pour N. Li, Kidzinski et al. 2015, le fait de regarder peu fréquemment ou de sauter de grandes parties de la vidéo suggère que la vidéo est perçue comme étant d'un niveau de difficulté plus élevé. Mais dans les encodages actuels nous n'avons pas de renseignements sur les sections vidéo sautées et donc non écoutées. Avec la représentation SIVS, nous avons dans la structure même de l'encodage les sections vidéo sautées par une avancée vers l'avant.

(vi) **Nombre de fois où l'étudiant a arrêté la vidéo :** En général, l'événement d'arrêt se produit à la fin de la vidéo, bien que l'étudiant puisse arrêter la vidéo à tout moment. Dans ce dernier cas, le curseur de lecture de la vidéo revient à son début et l'étudiant doit reprendre la lecture de la vidéo au début de celle-ci si un clic sur "play" se fait. Ainsi, le nombre d'arrêts peut indiquer combien de fois l'étudiant est revenu au début de la vidéo. La limite de ce trait dans un encodage est de ne pas pouvoir savoir à quel endroit de la vidéo l'étudiant a arrêté la lecture de celle-ci. Avec la représentation SIVS l'on est à mesure de dépasser cette limite et de pouvoir savoir à la seconde près quand est apparut la fin de la lecture de la vidéo et la reprise ou pas de la lecture de cette dernière à partir du début de la vidéo.

L'étude très citée de Guo et al. Guo, Kim et Rubin 2014 est un exemple de l'approche basée sur les mesures cumulatives, qui permet d'extraire des informations des sessions d'interaction vidéo. Leur étude est menée sur un très large échantillonnage de MOOC et de sessions d'étudiants. Les auteurs ont découvert que de multiples facteurs associés aux vidéos peuvent affecter l'engagement des étudiants, qui peut être mesuré par la durée de visionnage et la

volonté des étudiants de réaliser une activité d'évaluation. Une autre conclusion intéressante de Guo, Kim et Rubin 2014 est que les tutoriels sont souvent regardés plusieurs fois et pendant des périodes plus longues que les conférences. Cette constatation permet de supposer que certaines vidéos peuvent imposer un schéma de visionnage. Dans le même ordre d'idées, Bhat, Chinprutthiwong et Perry 2015 ont constaté que de légers changements dans la façon dont l'enseignant présente le contenu de la matière dans une vidéo peut influencer l'engagement en termes de proportion de vidéo regardée, ce qui est un autre exemple d'analyse basée sur les cumulatives d'écoute vidéo.

La limite de l'encodage SIVS proposée est son lien très étroit dans sa structure à la durée de la vidéo. On est alors contraint dans les analyses des traces vidéo à considérer des vidéos de mêmes durées à la seconde près ou de faire des analyses à l'intérieur de l'ensemble des traces d'interactions avec la même vidéo. Par exemple dans l'étude des interactions des étudiants avec une vidéo, la présence dans une vidéo d'une pause obligatoire imposerait à tous les apprenants une structure d'écoute qui rendrait difficile la distinction d'une séquence d'écoute d'une autre. Le défi ici est alors de trouver un moyen d'encoder les interactions qui pourrait discriminer (reconnaître) une séquence d'écoute d'un étudiant des autres indépendamment de la durée de la vidéo. C'est ce qui nous a motivés pour rechercher un autre type de représentation d'interactions vidéo des étudiants.

3.5 Codage d'interaction vidéo basé sur les modèles de visionnement (QR.1 et QR.2)

Des chercheurs comme Sinha, Jermann et al. 2014 ; Sinha, N. Li et al. 2014 ont étudié les modèles de visionnement vidéo au niveau du clic de l'étudiant afin de prédire le caractère unique du modèle par la prédiction du niveau d'engagement de l'étudiant, du clic suivant, de l'abandon de la vidéo et de l'achèvement ou de l'abandon du cours. D'autres chercheurs ont pu déterminer des processus d'apprentissage à partir des comportements des étudiants et en utilisant l'extraction de modèles pour classer les étudiants selon leurs compétences (Emond et Buffett 2015). L'étude du comportement des apprenants au cours du processus d'apprentissage en ce qui concerne l'engagement (Kizilcec, Piech et Schneider 2013), l'engagement social (Brinton, Chiang et al. 2014) ou la performance (S. Jiang et al. 2014) peut être utilisée pour prédire les abandons (Sherif Halawa, Daniel Greene et John Mitchell 2014) ou pour analyser les données démographiques (Guo et Reinecke 2014).

La construction de modèle d'interaction vidéo se fait généralement à partir des enregistrements d'écoutes vidéo des étudiants. Plusieurs approches ont été explorées, notamment par L. Chen, Zhou et Chiu 2013 ; L. Chen, Zhou et Chiu 2014 ; Yishuai Chen et al. 2014 qui ont étudié les systèmes de vidéo à la demande en ligne sur Internet afin de construire un modèle précis pour caractériser la répartition du temps de visionnage des utilisateurs par vidéo. Ils ont découvert que le temps de visionnage des vidéos peut être modélisé par une concentration de la distribution exponentielle, au début des vidéos, et une distribution tronquée de la loi de puissance, pour le reste de la vidéo, dès que l'utilisateur regarde la vidéo sans interruption. En effet, ils montrent à travers l'analyse de plus de 540 millions de sessions d'étudiants interagissant avec des vidéos, que les étudiants passent beaucoup de temps dans la navigation (visionner une partie de la vidéo après l'autre et seulement dans 20% de cas regarder complètement la vidéo). Pour eux, il s'agit-là d'une forme particulière de navigation (répétition partielle) dans le visionnage de la même vidéo. Ils ont proposé un nouveau modèle avec une alternative plus détaillée à la formulation du réseau de files d'attente fermées (voir D. Wu, Y. Liu et Ross 2009), qui peut aider à mesurer la popularité d'une vidéo. Leurs études mettent en évidence les paramètres et les distributions d'un tel modèle de comportement stochastique sur la base d'observations dans la pratique. En ce qui concerne les vidéos en tant que moyen majeur de diffusion de contenu MOOC, l'objectif de N. Li, Kidzinski et al. 2015 était justement d'analyser les différents types d'interactions vidéo des étudiants pour percevoir leurs difficultés en matière de vidéo. Ceci peut potentiellement aider les instructeurs à identifier les vidéos avec lesquelles certains étudiants pourraient avoir des difficultés (Kamahara et al. 2010) et leur donner une aide appropriée et prévenir l'abandon.

Certaines recherches se sont intéressées à la manière dont les étudiants dans leur ensemble interagissent avec une vidéo pour déterminer les points d'intérêt ou les parties importantes d'une vidéo. Il convient alors de souligner le point de convergence entre les diverses écoutes étudiantes. Dans ce cadre, Kim, Guo, Seaton et al. 2014 ont utilisé des traces vidéo edX de 862 vidéos et ont constaté que les segments de transition vidéo, ou d'autres points d'intérêt, correspondent à un nombre important d'interactions des étudiants avec une partie commune d'une vidéo, qu'ils appellent "pic d'interaction". Ces pics peuvent être visualisés par des outils spécifiques d'analyse du clic (Q. Chen et al. 2015). Kim, Guo, Seaton et al. 2014 ont également établi que les vidéos revisionnées affichent un taux d'abandon plus élevé ; ce qui suggère des modèles d'interaction plus typiques des étudiants à la recherche de segments spécifiques.

Certaines données démographiques et sociales peuvent justifier l'apparition de certains mo-

dèles d'interaction vidéo dans cette ligne, Guo, Kim et Rubin 2014 se sont intéressés à la différence démographique des étudiants dans un MOOC à travers la manière dont ces étudiants naviguent dans la matière du cours. Ils ont étudié la navigation de 140 546 étudiants de 4 cours du MOOC EdX provenant de 196 pays différents. Ils ont découvert que les étudiants qui cherchent une certification ignorent en moyenne 22% du contenu de la matière en revenant souvent aux contenus plus anciens (en vue d'avoir juste ce qu'il faut pour obtenir la certification). Ensuite, ils se sont intéressés au niveau d'utilisation des vidéos dans l'apprentissage à grande échelle en ligne. Ils montrent, par une étude empirique, comment la décision de production de vidéos affecte l'interaction avec les vidéos éducationnelles en ligne. Leur approche se limite uniquement au temps passé à regarder la vidéo, en ignorant les temps de pause, leurs fréquences et la succession des états dans les interactions vidéo.

3.6 Mesure de la similarité des séquences des étudiants (QR.3)

De nombreuses recherches sont consacrées à l'étude des interactions vidéo des étudiants, y compris les paramètres d'écoute vidéo spécifiques des comportements de visionnage de la vidéo.

Ces dernières années, Atapattu et Falkner 2017 ont étudié comment le comportement des étudiants, en matière d'engagement vidéo, dans le cadre du MOOC, influe sur leur réussite ou leur échec. Les questions qui ont soutenu leurs recherches sont celles de savoir si les particularités du discours sont-elles corrélées aux séquences d'interaction des vidéos MOOC ? Si oui, quelles sont leurs caractéristiques spécifiques ? En analysant 1,5 million de vidéos, ils ont scruté le raisonnement qui sous-tend la variation temporelle de l'interaction vidéo. L'analyse du discours consiste à extraire des phrases des transcriptions pour mesurer le discours et les caractéristiques linguistiques. Ils ont utilisé un outil particulier Coh-Metrix 3.0 (McNamara et al. 2014) pour identifier les caractéristiques du discours (nombre de mots descriptifs, nombre de syllabes, facilité d'utilisation du texte, narrativité, simplicité syntaxique, mot au le caractère concret, la cohésion référentielle, cohésion profonde, connectivité, diversité lexicale, pronom, fréquence des mots de contenu, familiarité, caractère concret, taux de prise de parole etc). Mais, ils se sont davantage concentrés sur la corrélation entre les caractéristiques du discours et les modèles d'interactions vidéo de MOOC pour trouver une corrélation entre la façon d'interagir (séquence vidéo d'interaction) avec la richesse du discours tenu dans la vidéo.

Certains chercheurs comparent les interactions vidéo des étudiants pour voir si un étudiant garde le style d'interaction vidéo similaire d'un cours à un autre. Par exemple, Dissanayake et al. 2018, en utilisant deux cours en ligne, ils ont pu déterminer si un étudiant a le même style d'interaction avec la vidéo. Les résultats de leurs recherches montrent qu'un étudiant peut changer de style d'apprentissage d'un cours à l'autre. Pour l'identification des styles d'apprentissage, ils ont regroupé les interactions des étudiants en quatre groupes : scénario de visualisation en une passe, scénario en deux passes, scénario répétitif et scénario de zapping. Chaque groupe a une distribution des mesures cumulatives spécifiques. Les voici : nombre de pauses (NP), nombre de recherches en arrière (NB), nombre de recherches en avant (NF), durée médiane des pauses (MP), longueur de la vidéo lue (RL), proportion de contenu vidéo sauté (SR), vitesse moyenne de la vidéo (AS), changement de vitesse effectif de la vidéo (SC), durée totale de lecture (TP). En utilisant ces mesures cumulatives pour regrouper les sessions d'interaction avec les vidéos, ils ont démontré que les mêmes sessions d'étudiants pouvaient être regroupées en différents groupes d'un cours à l'autre. Ils ont conclu qu'un étudiant pouvait changer son style d'apprentissage d'un cours à l'autre. Une meilleure démonstration du fait qu'un étudiant a le même style d'apprentissage dans un cours pourrait être de voir à quel niveau il est possible de reconnaître une session d'interaction spécifique d'un étudiant. C'est le but dans cette thèse : découvrir si un étudiant, dans le même cours, a le même style d'apprentissage en prédisant la séquence d'interaction de l'étudiant (par la chaîne de Markov de la séquence), parmi d'autres séquences, en utilisant différents classificateurs. Notre étude, dans le cadre de cette thèse, montre ainsi qu'un étudiant garde le même style d'interaction qui lui est unique dans un même cours (nous pouvons identifier les interactions d'un même étudiant). Cela se justifie par le fait que les étudiants ont le même intérêt, engagement et objectif au sein d'un cours alors dans un autre cours l'objectif, l'engagement et l'intérêt ne peuvent pas être les mêmes d'où cela influence sur leur style d'interaction avec les vidéos du cours considéré.

Certains chercheurs par ailleurs essaient de prendre en compte des parcours similaires en comparant la façon dont les étudiants étudient. N. Patel, Sellman et Lomas 2017 isolent, pour les étudier, des parcours fréquents d'apprentissage à partir d'un vaste ensemble de données éducatives. Ils se sont servis d'une méthodologie de *"pattern mining"* en utilisant un grand graphique de séquences de parcours et en isolant les principaux parcours. Pour obtenir ce graphique, ils ont utilisé la méthode de regroupement avec les distances de Levenshtein et la méthode "Ward". Le principal défi de l'utilisation des méthodes de regroupement est de définir le nombre approprié des groupes. Il n'existe pas encore de meilleure méthode pour trouver le nombre de groupes significatifs dans une donnée. La limite de cette méthode est

liée, cependant, à la limite de l'utilisation des techniques de mise en groupes. Un pattern distribué dans plusieurs groupes peut ne pas être significatif dans chaque groupe mais, si il est détecté dans son ensemble, il peut être significatif. Il est donc nécessaire de disposer d'une méthodologie qui puisse donner tous les patterns, avec le nombre de séquences dans chaque groupe, pour mieux aider à spécifier l'importance de chaque pattern. Leur méthodologie ne permet pas de savoir à quel point les différentes séquences sont proches (une mesure de similarité). Il est nécessaire de disposer d'une méthodologie qui pourrait donner une meilleure mesure de la similarité entre les séquences d'interaction pour ce type de recherche.

Un autre objectif pour la recherche de la similarité entre les séquences d'interaction des étudiants dans les MOOC est de pouvoir regrouper les interactions semblables. Par exemple, Klingler et al. 2016 ont regroupé les séquences d'interactions des étudiants dans le MOOC en fonction de la similarité (ou distance) de leurs interactions. Pour le calcul des similarités, ils utilisent la plus longue sous-séquence commune (LCS voir annexe G la section G) et la distance de Levenshtein (voir annexe G dans la section G). Au lieu de calculer les distances directement sur les séquences d'actions, ils ont appliqué le calcul à leurs valeurs agrégées des chaînes de Markov. Au lieu d'utiliser la distance euclidienne entre les probabilités de transition de deux chaînes de Markov (voir Köck et Paramythis 2011), ils proposent d'utiliser des mesures spécifiquement conçues pour comparer les distributions de probabilité. Ils proposent de calculer les probabilités de transition attendues en utilisant la distribution stationnaire sur les actions et comparer les fréquences de transition prévues. Pour cela, ils utilisent deux métriques communes : la divergence de Jensen-Shannon et la distance de Hellinger (voir Pardo 2005) pour calculer les distances entre les fréquences de transition attendues des chaînes de Markov. La principale contribution de leur recherche est de permettre de disposer des groupes des séquences cohérents. Ils ont proposé un pipeline de regroupement évolutif qui peut aider à détecter l'évolution de l'apprentissage des apprenants au fil du temps ainsi que l'évolution pertinente des clusters dans le temps. Avec cette méthode, ils ont pu détecter les comportements importants des étudiants au fil du temps et les propriétés de l'environnement d'apprentissage. Même s'ils ont utilisé la chaîne de Markov dans leur pipeline, ils ne captent pas tous les types des comportements avec le nombre des séquences utilisant chaque pattern.

Partant de l'hypothèse que l'approche d'une évaluation peut avoir une influence sur la façon d'interagir, certains chercheurs vont comparer les interactions des étudiants autour des périodes d'évaluation pour les comprendre. Boroujeni et Dillenbourg 2018 ont extrait des modèles d'étude latents, des séquences d'interactions des étudiants. Ils se sont intéressés aux patterns d'interactions des étudiants pendant les périodes d'évaluation dans le MOOC.

Ils ont utilisé deux approches fondées sur des hypothèses. Ils ont prédéfini des modèles en fonction de la première activité de l'étudiant dans la séquence. Ils ont ensuite regroupé les étudiants en fonction de leur première activité pour obtenir la chaîne de Markov de chaque groupe d'étudiants. Pour cela, ils ont utilisé pour le regroupement la représentation de la chaîne de Markov de leurs interactions. Ils ont pu identifier en période d'évaluation onze (11) types d'interactions (11 groupes d'interactions semblables) répartis entre des approches fixes d'interactions (3 groupes) et les approches changeantes (7 groupes). Dans la séquence des données d'activité de certains étudiants, le premier événement ne peut pas définir adéquatement la séquence des activités de l'étudiant. Dans ces cas, la méthode qu'ils ont proposée ne pouvait pas convenir pour comparer deux séquences vidéo d'interactions connaissant la limite de la chaîne de Markov pour les interactions (ne prend pas en compte du temps passé dans les états).

L'utilisation de la représentation en chaîne de Markov pour regrouper les écoutes semblables dans les vidéos a également vu le jour. C'est ainsi que Mongy, Djeraba et Simovici 2007 ; Mongy, Bouali et Djeraba 2007 ont regroupé les comportements de visionnage de vidéos en chaînes de Markov de premier ordre, dans les MOOC, afin d'extraire les types de comportements observés. Ils ont demandé à un spécialiste du domaine d'analyser leurs "clusters" et ils les ont nommés : "visionnage rapide de la vidéo", "visionnage d'une séquence vidéo spécifique" et "visionnage complet de la vidéo". Ils ont regroupé les comportements des utilisateurs en appliquant l'algorithme de k-means pour produire des modèles de visionnement bien distincts qui couvrent la diversité des comportements observés.

La méthodologie de similarité proposée dans le cadre de cette thèse, en exprimant les similarités en termes de degré de similarité, peut donner des regroupements des interactions les plus semblables. Les séquences ayant les degrés de ressemblance les plus élevés peuvent être classifiées dans le même groupe. Un test de la méthodologie a été réalisé dans le cadre des regroupements étiquetés en considérant comme étiquètes les vidéos et voir comment les interactions de la même vidéo se retrouvent dans le même groupe en prenant en compte l'hypothèse que chaque vidéo impose dans une certaine mesure une manière particulière d'interagir avec elle.

3.7 Prédiction des succès d'étudiants basée sur les mesures agglomératives d'interaction vidéo (QR.4)

De nombreuses recherches sont consacrées à l'étude des interactions vidéo des étudiants, y compris les analyses spécifiques des comportements de visionnage de la vidéo. Par exemple, M. N. Giannakos, Chorianopoulos et Chrisochoides 2015 ont établi une relation entre les interactions vidéo, le visionnage répété, et le niveau de la connaissance requise pour un segment vidéo spécifique. L.-Y. Li et Tsai 2017, ont également étudié le lien entre les modèles de comportement des étudiants et les succès d'apprentissage. Leur étude comprenait des éléments comme : des vidéos de cours, des diapositives de cours, des devoirs partagés et des messages de forum ouvert. Ils ont conclu que les cours vidéo et les diapositives de cours étaient les plus utilisés par les étudiants. He, Zheng et al. 2018 ont mesuré l'utilisation de la vidéo par les étudiants dans un MOOC en relation avec le succès académique. Ils ont proposé des indicateurs basés sur les mesures agglomératives d'interaction vidéo des étudiants pour quantifier l'utilisation des ressources vidéo, par exemple, le taux d'assiduité, le taux d'utilisation et le taux de visionnement. Notre étude utilise également les mesures du taux d'utilisation et du taux d'assiduité en vue de la prédiction des succès.

De leur côté, Hughes et Dobbins 2015, montrent comment l'utilisation de techniques d'analyse de données dans les MOOC peut aider à prédire le risque d'abandon scolaire des étudiants avant qu'il ne se produise. Il est alors possible d'identifier suffisamment tôt les étudiants qui risquent d'abandonner leurs études pour leur apporter une aide adéquate avant que cela ne se produise (Kamahara et al. 2010). Les succès des étudiants peuvent être influencés par certaines caractéristiques des vidéos elles-mêmes (Guo, Kim et Rubin 2014) qui les incitent à visionner une vidéo plus souvent que les autres (Breslow et al. 2013 ; Seaton, Bergner et al. 2014), ou à passer plus de temps à regarder se dérouler la vidéo que sur d'autres ressources en ligne (Ozan et Ozarslan 2016).

Au-delà des interactions vidéo, la prédiction du succès et de l'échec a été faite par R. S. Baker et al. 2015 sur la base de l'activité en ligne des étudiants qui utilisent l'environnement d'apprentissage Soomo. En utilisant de nombreux modèles de prédiction, ils montrent que, avec le modèle de régression logistique de leur proposition de modèle combiné, ils peuvent identifier jusqu'à 59,5 % des étudiants qui auront de mauvais résultats dans le cours : ce qui est mieux que des résultats aléatoires. Récemment, Sheshadri et al. 2019 ont analysé les journaux de bord des étudiants de trois cours mixtes pour prédire leurs succès (voir également O. H. Lu et al. 2018). Les résultats de leur recherche montrent qu'il est possible de

prédire les succès des étudiants en fonction de leur niveau d'utilisation du système en ligne. La limite fondamentale de ces études comme celles de C.-H. Yu, J. Wu et A.-C. Liu 2019, Goulden et al. 2019, W. Wang, H. Yu et Miao 2017 et T.-Y. Yang et al. 2017 est qu'elles utilisent pour leur prédiction de la performance des étudiants beaucoup plus que les mesures agglomératives d'écoute vidéo. En effet, ils utilisent toutes les interactions de l'étudiant avec la plateforme et non pas uniquement les interactions vidéo. Et dans certaines situations si une de ces mesures vient à manquer dans le processus de collecte des données, la prédiction n'est plus possible. L'objectif également dans cette thèse est de pouvoir faire des prédictions à partir des données simples à récolter et avec moins des mesures possibles à manipuler.

Des études ont été réalisées sur l'impact de certains facteurs d'apprentissage sur les succès des étudiants dans les MOOC (Almeda et al. 2018, Deshpande et Chukhlomin 2017). Quelques-unes se sont concentrées sur l'impact des facteurs externes (Rai et Chunrao 2016, Namestovski et al. 2018). L'impact géographique et culturel sur les succès, par exemple, a également été étudié par Z. Liu et al. 2016. Même les facteurs externes qui facilitent un plus grand engagement de l'étudiant dans les MOOC ont été examinés par Li Q 2016 de même que Ferguson et Clow 2015. La relation entre les facteurs d'apprentissage, comme les forums de discussion et les résultats d'apprentissage, joue un rôle important dans le succès de l'étudiant selon Bergner, Kerr et Pritchard 2015. Le comportement des étudiants dans les MOOC, en particulier dans les forums de discussion, a mis en lumière le fait qu'il peut affecter l'apprentissage acquis par les étudiants, en particulier par leur comportement cognitif (X. Wang et al. 2015). Une étude a permis de prédire les succès des étudiants dans les exercices paramétrés, en connaissant leur historique de succès et d'échec dans les pratiques précédentes (Sahebi, Y. Huang et Brusilovsky 2014).

Pour étudier le succès des étudiants, certains chercheurs procèdent par regroupement des interactions semblables. Par exemple, N. Li, Kidziński et al. 2015 ont analysé l'interaction des étudiants en vidéo (comme la pause, la recherche en avant et en arrière et le changement de vitesse) pour caractériser les comportements vidéo en modèles, en employant une méthodologie de regroupement. Les auteurs se sont concentrés sur les relations entre l'interaction vidéo et la perception de la difficulté de la vidéo, les comportements de revisite de la vidéo et le succès des étudiants. Les résultats de leur analyse fournissent des indications pour améliorer l'expérience d'apprentissage du MOOC. Plus précisément, en nommant des comportements qui montrent des difficultés à comprendre les vidéos, ils ont pu identifier les vidéos qui devraient être améliorées pour la prochaine version de MOOC. Notre étude dans cette thèse tout en prédisant les succès des étudiants, s'intéresse également à voir si un

étudiant, en particulier, conserve le même style d'interactions vidéo pendant un cours en ligne.

Des chercheurs ont étudié la corrélation entre le style d'interaction et le succès des étudiants. Des chercheurs comme Brinton, Buccapatnam et al. 2016 ont utilisé deux cadres de présentation pour expliquer les flux de clics d'observation vidéo où l'un est basé sur la séquence des événements de l'étudiant et l'autre sur la séquence des positions visitées. Ils ont constaté que certains comportements sont significativement en corrélation avec les changements de probabilité, qu'un étudiant réussisse dès la première tentative ou qu'il ne réponde pas à un test vidéo. Ils ont construit une prédiction du succès dans un modèle de quiz, avec un cadre basé sur les positions, à partir des positions visitées dans une vidéo. Dans leur étude, ils ont constaté que le comportement de l'étudiant qui regarde la vidéo peut être utilisé pour améliorer la prédiction des succès de l'étudiant sur une base de la vidéo de l'étudiant en train de regarder la vidéo. Sur la base des conclusions de leur étude, la présente recherche cherche à voir s'il est possible de construire un modèle de prédiction de la séquence d'interaction de l'étudiant, basé sur l'interaction vidéo, pour identifier le style d'apprentissage de l'étudiant dans la vidéo. Ce modèle peut ensuite être étendu à un modèle de prédiction du succès des étudiants, basé sur le style d'interaction des étudiants dans la vidéo. Kahan, Soffer et Nachmias 2017 ont amélioré l'idée de trouver le type de comportement des participants dans un MOOC par leur interaction avec le système. Ils ont trouvé sept types de participants en fonction de leur niveau d'interaction, allant des "goûteurs" aux "engagés", et du pourcentage d'étudiants qui se trouvent dans chaque type.

La complexité de la vidéo a une influence sur le succès des étudiants. Van der Sluis, Ginn et Van der Zee 2016 ont étudié l'influence de la complexité de la vidéo sur le temps de connexion des étudiants (temps passé à regarder la vidéo) et le taux de "logement" (nombre de vidéos regardées). Ils ont formalisé une définition de la complexité de l'information en vidéo. La complexité de la vidéo et le logement des étudiants montrent une relation polynomiale dans laquelle une complexité vidéo faible et élevée augmente le logement des étudiants. Sans tenir compte de la complexité de la vidéo, la présente étude vise à déterminer si un étudiant a une façon spécifique d'interagir avec les vidéos.

Le taux de réussite dans un cours en ligne est en lien avec l'interaction avec les vidéos. Par exemple Hu et al. 2019 ont mesuré l'efficacité des cours en ligne en analysant le comportement des étudiants lorsqu'ils interagissent avec des vidéos. Ils ont utilisé une analyse statistique du comportement de chaque étudiant lorsqu'il interagit avec des vidéos, puis ils

ont développé un algorithme de traitement des données basé sur la plateforme Spark pour calculer le comportement de visionnage des vidéos à chaque instant. Ils ont utilisé la méthode de l'entropie pour pondérer d'autres événements d'interaction vidéo qui n'étaient pas "lus en vidéo". Grâce à cette méthode basée sur Spark, ils ont pu analyser rapidement les caractéristiques du comportement de visionnage des vidéos. Ils ont répondu au même problème que Hmedna, El Mezouary et Baz 2017 avaient précédemment abordé en utilisant des réseaux neuronaux pour l'identification et le suivi des styles d'apprentissage des apprenants dans les MOOC. La présente étude reprend la même idée ; mais, au lieu d'utiliser l'analyse statistique et l'entropie des événements vidéo, elle utilise les représentations de la chaîne de Markov de chaque séquence d'interaction avec la vidéo.

Les antécédents des étudiants peuvent, d'une manière ou d'une autre, affecter les succès des étudiants dans le MOOC (DeBoer et al. 2013). Certains ont pu prédire les succès des étudiants à l'examen en analysant simplement les données de leur réseau social (Fire et al. 2012). D'autres ont classé les apprenants en fonction de leur interaction avec les cours vidéo et de l'évaluation (Kizilcec, Piech et Schneider 2013). Ils ont constaté qu'un étudiant peut être engagé sans soumettre d'évaluation. Certains ont souligné l'importance des vidéos sans tenir compte de leur impact sur les succès des étudiants. La fonction d'apprentissage, utilisée pour prédire les succès des étudiants dans le MOOC, montre les meilleures pratiques en termes de processus d'apprentissage (S. Halawa, D. Greene et J. Mitchell 2014). Les mesures de l'engagement, selon le style de vidéo, dans les MOOCs, ont été examinées par Guo, Kim et Rubin 2014 et montrent que les vidéos de type "talking head" et "khan" sont plus engageantes que les cours préenregistrés en classe. Pourtant, il n'existe pas de classification complète des styles vidéo en termes d'engagement, auquel les étudiants se réfèrent. De nombreuses études comme celle de Korkut et al. 2015 ; Diwanji et al. 2014 ont défini des styles de vidéo pouvant être utilisé dans les cours en ligne. Ces différents styles de vidéo sont utilisés dans les cours en ligne, comme Coursera, Udacity, Edx, Khan Academy, TED et Video Lecture. Dans la littérature, il existe sept styles vidéo différents : le style tête parlante (Ilioudi, M. N. Giannakos et Chorianopoulos 2013 ; Ozan et Ozarslan 2016), le style présentation (Ozan et Ozarslan 2016), le style image dans l'image (Chorianopoulos et M. N. Giannakos 2013), le style voix au dessus d'une présentation (Griffin, D. Mitchell et S. J. Thompson 2009), le style Khan (Ilioudi, M. N. Giannakos et Chorianopoulos 2013), le style interview et le style capture d'écran de l'instructeur. La présentation de chaque style vidéo en détail se trouve à l'annexe E. La limite de ses études est la non prise en compte de l'effet du présentateur dans la motivation d'écouter ou non une vidéo. Dans le cadre de cette thèse, au lieu de nous intéresser à la motivation par style de vidéo, nous proposons une

méthodologie pour pouvoir reconnaitre une vidéo à partir des interactions des étudiants avec la vidéo, quel que soit le style de cette dernière.

3.8 La représentation d'interaction vidéo en vue de la prédiction des succès des étudiants (QR. 5)

Les premiers travaux sur l'utilisation d'algorithmes d'apprentissage automatique pour prédire le décrochage des étudiants ont été réalisés par Dekker, Pechenizkiy et Vleeshouwers 2009. Ils recueillent des données académiques du premier semestre du programme de génie électrique pour prédire si l'étudiant terminera ou abandonnera le programme. Le même concept de prédiction des premiers résultats des étudiants a été appliqué à un niveau de cours, dans le cours en ligne, par Barber et Sharkey 2012 qui ont construit un modèle pour l'Université de Phoenix afin d'identifier les étudiants qui risquent d'échouer le cours auquel ils sont actuellement inscrits. Zafra et Ventura 2012 ont utilisé ce concept de prédiction des succès des étudiants dans les environnements éducatifs basés sur le web en développant une méthode (apprentissage à instances multiples) qui pourrait être très utile pour l'identification précoce des étudiants à risque, en particulier dans les très grandes classes en ligne, et qui permet à l'instructeur de fournir aux étudiants à risque des informations sur les activités les plus pertinentes, pour aider les étudiants à avoir de meilleures chances de réussir un cours.

Dans le même ordre d'idée, Wolff et al. 2013 ont analysé le comportement des étudiants en matière de clics pour prédire les étudiants qui risquent d'échouer à un module de l'Open University ("OU" est l'un des plus grands établissements d'enseignement à distance au monde). Ils ont mené leurs recherches en appliquant la méthode d'analyse de données GUHA (*General Unary Hypothesis Automation*) pour obtenir des résultats prometteurs afin de dégager une hypothèse précise sur les étudiants qui échouent. L'exploration de l'interaction des étudiants sur le web pour prédire leur note finale a été réalisée par Romero et al. 2013 qui ont comparé les succès de différentes techniques d'exploration de données (telles que les méthodes statistiques, les règles des arbres de décision, la règle floue et le réseau neuronal) pour la classification des étudiants. S. Jiang et al. 2014 ont concentré leurs prédictions de la note finale des étudiants sur la première semaine d'interaction, en considérant cette période comme celle de fort décrochage (Hill 2013). Ils ont utilisé une combinaison du succès des étudiants dans le devoir de la première semaine et de l'interaction sociale au sein du MOOC en vue de prédire leur succès final dans le cours. En utilisant la régression logistique comme classificateur, ils ont pu prédire avec une grande précision la probabilité que les étudiants obtiennent des certificats pour l'achèvement du MOOC, ainsi que la mention obtenue (c'est-à-dire distinc-

tion et normal). En adoptant également les données de la première semaine d'interactions des étudiants, Kloft et al. 2014 ont pu prédire l'abandon de manière significativement meilleure que les méthodes de base.

Certains chercheurs pour prédire le succès des étudiants, recourent à des modèles bien établis de prédiction. Par exemple, Fei et Yeung 2015 ont utilisé des modèles temporels en considérant la prédiction d'abandon comme un problème de classification des séquences. En particulier, ils ont découvert qu'un modèle de réseau neuronal récurrent (RNN), avec des cellules à mémoire de longue et de courte durée (LSTM), surpasse de loin les méthodes de base et les autres méthodes.

En combinant les activités en ligne des étudiants et en utilisant l'environnement d'apprentissage Soomo et les notes des devoirs pour prédire l'échec du cours, R. S. Baker et al. 2015 ont découvert que les étudiants qui ont des chances de réussir sont ceux qui accèdent tôt aux ressources et qui continuent d'y accéder tout au long de la première semaine du cours : ils obtiennent aussi de bons résultats lors des activités formatives. En utilisant ces indicateurs précoces, un modèle de régression logistique a atteint un maximum de 0,58 comme F_1.

Allant plus loin dans la prédiction des succès des étudiants, R. Raga et J. Raga 2019 ont utilisé des cours mixtes, pour lesquels des données en ligne de niveau universitaire et académique sont disponibles. En utilisant les réseaux neuronaux profonds (*Deep RNN*), ils ont obtenu des précisions élevées dans la prédiction des résultats finaux, allant jusqu'à 91,07 % avec un score ROC_AUC de 0,88.

Les chercheurs comme Sinha et Cassell 2015 ont quant à eux utilisé le visionnage de conférences vidéo, combiné à l'accès aux travaux de cours et à la publication sur un forum de discussion tiré d'un cours en ligne complet dans MOOC. Un modèle CRF (Conditional Random Fields) a atteint une précision de 0,581, un rappel de 0,660 et un F_1 pondéré de 0,560. Notre étude est unique en ce sens que nous nous concentrons uniquement sur l'interaction vidéo pour les prédictions.

L'utilisation des interactions des étudiants pour prédire leur succès a été réalisé par Brooks, C. Thompson et Teasley 2015 qui ont développé une méthode pour convertir les données des journaux de bord éducatifs en caractéristiques adaptés à la construction de systèmes de prédiction de réussite des étudiants. Contrairement à la modélisation cognitive ou d'analyse

de contenu, ces modèles sont construits à partir des interactions entre les apprenants et les ressources, une approche qui n'exige pas de contribution de la part d'experts en matière d'enseignement ou de domaine et peut être appliquée à tous les cours ou environnements d'apprentissage.

Dans le cadre de cette thèse, nous utilisons la représentation des interactions des étudiants pour reconnaitre les interactions qui peuvent conduire au succès ou à l'échec. Les prédictions en utilisant les interactions vidéo avec la représentation TMED sont plus performantes que les autres méthodes de prédictions basées entre autres sur des mesures agglomératives.

CHAPITRE 4 ENSEMBLE DE DONNÉES ET TRAITEMENT

4.1 Introduction

Ce chapitre fait mention des données que nous avons utilisé pour nos recherches. Nous allons présenter dans un premier temps les statistiques générales des d'inscriptions dans le cours considéré (Body 101x) sur la plateforme edX. Ensuite, nous parlerons de l'organisation interne du cours y compris les travaux et sujets abordés. Nous terminerons par les statistiques de l'utilisation des ressources vidéo par les étudiants et comment nous avons utilisé ces données dans le cadre de nos recherches. Dans le cadre de cette thèse, nous appellerons événement dans le cadre d'une trace d'étudiant, la capture du début d'une activité par un étudiant. Par exemple lorsqu'un étudiant clique sur "play" dans une vidéo, un événement est créé avec le temps de l'apparition (date, heure, minute, seconde) de l'événement et est enregistré sur le serveur. Jusqu'au prochain changement d'état de l'étudiant, le serveur n'enregistre rien. Nous appellerons activité d'un étudiant dans le cadre de cette thèse lorsque nous prenons en considération dans les analyses la durée entre deux événements. Par exemple un événement "play" apparait au temps 00 :00 :00 et un événement "pause" au temps 00 :00 :10. On dira alors que l'activité "play" de l'étudiant a duré dix (10) secondes.

4.2 Statistiques générales du cours Body 101x

Dans le cadre de cette thèse, nous avons utilisé les traces d'un cours en ligne de l'Université McGill à Montréal sur la plateforme edx qui comptait un nombre important d'étudiants inscrits (Tableau 4.1). Le cours Body101x a été donné par le professeur Ian Shrier à l'automne 2015 et à l'hiver 2016. Nous avons eu accès aux traces de ces deux sessions. Les étudiants sans interactions vidéo ont été exclus de cette étude. Le cours comporte 138 vidéos (plus une vidéo diffusée en direct). Ce cours a une durée de treize (13) semaines à chaque session.

Pour la session de l'automne 2015, 30 640 étudiants ont interagi avec les vidéos. Parmi eux, 30 519 étudiants étaient inscrits en tant que "honours" et 1 084 ont réussi le cours avec une note finale supérieure ou égale à 50 % dont 970 sont parmi ceux qui ont interagi avec toutes les vidéos de la première semaine du cours. Nous avons identifié 10 424 étudiants qui ont interagi avec toutes les vidéos de la première semaine de cours sans exception. Nous avons utilisé les traces de ces étudiants qui ont interagi avec toutes les vidéos de la première semaine dans le cadre de la prédiction précoce des succès ou échec des étudiants. Notons que 90% des

étudiants ayant réussi le cours ont interagi avec toutes les vidéos de la première semaine. La répartition des notes finales des étudiants se retrouve dans la table 4.2.

Pour la session d'hiver 2016, des 13 135 étudiants qui ont interagi avec les vidéos, 11 555 étudiants étaient inscrits en tant que "honours" et 339 ont réussi le cours avec une note finale supérieure ou égale à 50 %. Les données de cette session ont été utilisé en fin de comparaison. C'est à cause du plus petit nombre d'étudiants enrôlés à cette session que nous n'avons pas reporte les résultats des analyses de cette session dans cette thèse. Toutes les analyses faites avec la session d'automne ont été refaites avec celle d'hiver par souci de vérification de nos analyses avec beaucoup plus des traces de la session d'automne. Comme il n'y a pas eu de divergence nous avons simplement présenté nos résultats avec les traces de la session d'automne.

Tableau 4.1 Informations générales sur le cours.

Cours	Body101x	Body101x
Session	Hiver 2015	Automne 2016
Institution	Université McGill	Université McGill
Etudiants inscrits	30 640	13 135
Etudiants notés	30 519	11 553
Etudiants ayant réussi	1084	339
Nombre des vidéos	138 + 1 direct	138 + 1 direct

Tableau 4.2 Répartition des notes finales des étudiants.

Notes finales sur 100	**Nombre des étudiants 2015**	**Nombre des étudiants 2016**
0 to 10	28 219	10 853
11 to 20	491	85
21 to 30	391	156
31 to 40	209	85
41 to 50	137	35
51 to 60	123	47
61 to 70	196	62
71 to 80	222	76
81 to 90	377	100
91 to 100	154	54

La durée totale de toutes les vidéos pour le cours est d'environ seize (16) heures, plus 64 minutes pour la session qui a été livrée "en direct". La durée moyenne des vidéos est de sept (7) minutes par vidéo (voir le tableau 4.1pour plus d'informations).

Sur les quatre (4) giga-octet de données d'activités étudiantes enregistrées en 2015, nous avons extrait 2 733 169 événements étudiants différents. Le temps de l'occurrence de chaque événement est aussi enregistré au millième de seconde. Ainsi chaque événement est associé à une vidéo et un étudiant dans les traces et nous permet ainsi de connaître pour chaque étudiant la succession des événements dans le temps et de calculer le temps des activités de lecture, de pause, de recherche dans la vidéo de chaque étudiant pour chaque vidéo. Les événements ont été collecté à partir de toutes les interactions vidéo des étudiants, telles que le clic de *"play"*, *"pause"*, *"seek"* (en avant et en arrière pour revisiter des sections d'une vidéo), et *stop*. Ce sont les événements enregistrés dans les traces vidéo des étudiants. La durée d'une activité d'un étudiant est calculée en premier lieu en soustrayant le temps de l'événement à celui du précédent. Pour préparer nos données, nous avons créé une nouvelle base de données qui comprenait : les étudiants par leur identification, la séquence de toutes les activités de chaque étudiant y compris le temps passé dans chaque activité et l'identification de vidéo s'il s'agit d'une interaction avec une vidéo. Tous ces événements sont compilés en fonction de leur occurrence dans une ligne de temps.

L'étudiant a également la capacité de pouvoir choisir le rythme de rapidité d'écoute vidéo. L'étudiant peut choisir l'option de rendre plus lente la vidéo à deux niveaux : 0.5x et 0.75x ou écouter normalement la vidéo au rythme de son enregistrement : 1.0x ou écouter plus rapidement avec trois niveaux de rapidité : 1.25x, 1.50x et 2.0x comme l'illustre la figure 4.1. Dans nos recherches, nous avons étudié l'influence de changement de la rapidité de la vidéo. Nous avons remarqué que très peu d'étudiants (moins de 0.1%) ont changé au moins une seule fois le rythme des vidéos. En comparant leurs interactions avec la vidéo lorsque le rythme est changé et lorsqu'ils écoutent avec le rythme par défaut, il n'y a aucun changement significatif. Ce qui nous a conduit à ne pas prendre cet aspect de la plateforme en considération dans nos analyses. L'hypothèse est que le peu des étudiants qui ont changé le rythme d'une vidéo l'ont fait surtout pour tester cette fonctionnalité plutôt que par choix de commodité.

La caractéristique intéressante de la plateforme de ce cours est que les apprenants ont une autre alternative que l'utilisation de vidéos. Par exemple, l'étudiant peut choisir de ne pas

Figure 4.1 Plate-forme avec vidéo montrant les divers niveaux d'écoute de la vidéo.

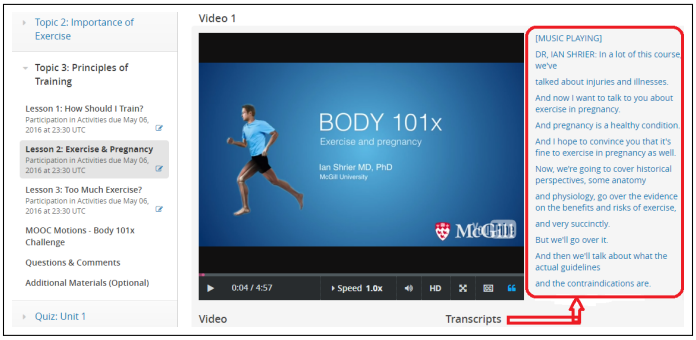


Figure 4.2 Plate-forme avec vidéo montrant la transcription de la vidéo.

interagir avec la vidéo et de se contenter de lire la transcription de la vidéo qui est disponible à côté de chaque vidéo, comme le montre la figure 4.2.

L'utilisation des transcrits par les étudiants dans leurs interactions est signalée dans les traces par un événement. Très peu d'étudiants ont utilisé le transcrit comme moyen unique de suivi de la vidéo (moins de 0.01%). La plateforme donne par contre la possibilité de cliquer sur une phrase dans le transcrit pour avoir la vidéo au lieu où cette phrase est prononcée (une forme de recherche ciblée dans la vidéo). Lorsque la vidéo est écoutée le transcrit montre exactement en une autre couleur la phrase est qui est en train d'être prononcée par la personne qui parle dans la vidéo. Ceci permet à un étudiant qui aurait des difficultés avec l'accent de la personne qui parle d'avoir un support textuel. Des études comme celles de Grgurović et Hegelheimer 2007 montrent que très peu d'étudiants utilisent le transcrit en général comme mode soutient dans la compréhension des vidéos comparé aux sous-titres. Comme

souligné précédemment, dans nos données, nous avons remarqué que très peu d'étudiants ont utilisé le transcrit dans leurs interactions. Nous avons fait une étude pour voir si l'utilisation des transcrits par un étudiant peut changer son style d'interaction (temps d'écoute, nombre de recherches, temps de pause) par rapport à une vidéo où le même étudiant n'a pas utilisé le transcrit. Nous avons trouvé que le fait d'utiliser le transcrit n'a pas un effet observable sur le style d'écoute de l'étudiant. Pour cette raison, nous n'avons pas pris en compte dans nos analyses le fait d'avoir ou non utilisé le transcrit.

4.3 Organisation interne du cours

Dans cette section nous présenterons la description et les objectifs du cours selon les instructeurs. Ceux-ci présentent l'organisation du cours autour des objectifs qu'ils se sont donnés.

L'exercice physique est promu comme une composante fondamentale d'un mode de vie sain. La bonne question est de se demander pourquoi ? L'exercice physique est plus qu'un "art de transformer les gros repas en étirements musculaires en soulevant des objets lourds qui n'ont pas besoin d'être déplacés, ou en courant quand personne ne vous poursuit" (Citation anonyme). Notre corps a évolué pour se déplacer sur plusieurs millénaires.

En effet, les personnes obèses physiquement actives vivent plus longtemps que les personnes minces inactives. Vous voulez voir les preuves des avantages et des risques de l'étirement ? Quelle est la meilleure façon de traiter une blessure à la cheville, au genou et/ou à l'épaule ? Comment une blessure affecte-t-elle l'humeur et quelles en sont les conséquences ?

Que vous soyez un athlète de compétition, un musicien ou un danseur en herbe, que vous jouiez pour le plaisir ou que vous souhaitiez simplement mener un style de vie actif, ce cours vous divertira et vous mettra au défi. Les sujets abordés comprendront les principes de base et avancés du mouvement corporel et les questions biologiques, psychologiques et sociales liées à l'activité/au sport/aux blessures/à la réadaptation. Le cours comprendra du contenu provenant d'experts internationaux de premier plan dans de nombreux domaines liés à la science de l'exercice. Le Syllabus du cours décrit les trois grandes unités en forme d'objectifs comme suit :

1. Examiner les avantages de l'activité physique
   (a) Discuter des avantages de l'activité physique
   (b) Identifier les défis liés à la promotion de l'activité physique
   (c) Décrire les principes généraux sur la façon de former

2. Discuter de la manière de prévenir les blessures
   (a) Expliquer quand et pourquoi l'étirement est efficace et quand il ne l'est pas
   (b) Décrire les erreurs de l'entraînement qui causent des blessures
   (c) Discuter de la similitude entre les athlètes, les musiciens, les danseurs et les artistes de cirque

3. Déterminer ce qu'il faut faire en cas de blessure
   (a) Décrire ce que vous pouvez faire par vous-même en cas de blessure
   (b) L'importance de la réhabilitation et de toutes ses composantes
   (c) Identifier les principes sous-jacents pour décider du moment approprié pour reprendre une activité après une blessure

4.4 L'utilisation des traces vidéo

La base des données transformées que nous avons obtenues provient de la mise en ligne de temps des activités de chaque étudiant et pour chaque vidéo comme nous l'avons décrit précédemment. Nous avons, pour chaque étudiant constitué une séquence d'interaction de chaque session d'interaction vidéo. Une session d'interaction vidéo d'un étudiant est l'interaction avec une vidéo sans arrêt de plus de 5 minutes. Lorsqu'un étudiant quitte la page d'une vidéo de plus de 5 minutes, une nouvelle session est initiée. D'une manière générale, nous avons observé dans les données très peu de retour sur la même vidéo après avoir quitté la page de la vidéo. Par contre, pour les vidéos qui ont fait l'objet de quiz ou de question post vidéo, nous avons constaté un nombre plus élevé de recherches à l'intérieur de la vidéo, et des parties réécoutée de la vidéo.

Pour certaines données comme la note finale de chaque étudiant nous les avons reçues des données prétraitées, que nous avons intégré dans notre base de données. Nous avons traité donc les données brutes enregistrées sur le serveur, comme il nous était possible de mettre sur une ligne de temps les événements de chaque étudiant pour chaque vidéo. D'où une représentation sous forme d'une séquence d'interactions en prenant en considération le temps

de chaque activité de l'étudiant.

Le cours et les vidéos sont organisés autour de trois grands axes classés en termes d'unités de cours. Les unités du cours sont composées de modules qui sont les divers axes d'explorations de chaque unité décrite plus haut à l'intérieur de chaque unité. Nous avons ainsi neuf (9) modules répartis en trois (3) modules dans chaque unité.

Les modules sont composés des leçons que l'on peut faire correspondre à des chapitres du cours. Chaque module a plusieurs leçons comme le montrent les détails dans les tableaux des trois unités (Tableaux des annexes A, C, D).

Le contenu de ce cours en ligne est organisé pour une période de treize (13) semaines avec une durée moyenne de 4 semaines par unité. La répartition de la matière en termes de semaine est observable dans les tableaux des annexes A, C, D. En somme, l'organisation du cours est comme suit : Trois unités composées des modules qui eux-mêmes sont subdivisées en leçons qui sont subdivisées de plusieurs vidéos chacune et des exercices post vidéo.

Dans les tableaux des annexes A, C, D, chaque cellule représente une vidéo et dans chaque cellule le temps en minutes suivi des secondes de la durée de chaque vidéo (sous forme hh :mm :ss). Chaque leçon est composée de plusieurs vidéos. Les vidéos sont des courtes séquences présentant chaque aspect de la leçon, utilisant divers styles vidéo comme :

1. **Le style tête parlante :** Ilioudi, M. N. Giannakos et Chorianopoulos 2013 ; Ozan et Ozarslan 2016
2. **Le style présentation :** Ozan et Ozarslan 2016
3. **Le style image dans l'image :** Chorianopoulos et M. N. Giannakos 2013, Hansch et al. 2015
4. **Le style de la voix au dessus d'une présentation :** Griffin, D. Mitchell et S. J. Thompson 2009
5. **Le style Khan :** Ilioudi, M. N. Giannakos et Chorianopoulos 2013
6. **Le style interview :** Kizilcec, Papadopoulos et Sritanyaratana 2014
7. **Le style capture d'écran de l'instructeur :** Santos-Espino, Afonso-Suárez et Guerra-Artal 2016

Dans ce cours quatre styles vidéo ont été utilisés : le style présentation, le style tête parlante, le style interview, le style Khan et le style image dans l'image. La table 4.3 montre le nombre des vidéos par style vidéo contenu dans les 139 vidéos du cours. Il faut noter qu'une même vidéo peut être classée dans deux styles différents du moment où il y a dans la même vidéo des segments de vidéo de ces deux styles différents.

Tableau 4.3 Le nombre des vidéos par style vidéo du cours Body 101x.

Style vidéo	**Nombre des Vidéos**
Tête parlante	88
Présentation	25
Image dans l'image	52
Voix au dessus d'une présentation	0
Khan	2
Interview	21

Les tableaux des annexes A, C, D indiquent également les évaluations des divers travaux et quiz durant le cours et le nombre des points que l'étudiant peut obtenir pour chaque activité d'évaluation. Le total des points possibles est de 301. Pour obtenir une note sur 100, les points accumulés par chaque étudiant à la fin du cours seront divisés par 3 pour avoir la note finale dans le cours. Dans le cas de cette plateforme, la note finale de chaque étudiant à la fin du cours est calculée automatiquement par le système.

La figure 4.3 montre la structure des données brutes qui sont enregistrées sur le serveur que nous avons traité pour nos analyses. On peut noter que les données sont pour tous les étudiants en même temps. Dans cette figure, nous avons classé les données par étudiant. L'identifiant des étudiants pour réunir les événements de chaque étudiant avant de poursuivre le traitement des données brutes.

```
> Ovweview_3_2015[1:25,-c(14:17)]
             id  event_type              path user_id        code curentTime                       module_id new-time old-time new_speed old_speed         time_event_emitted               course_id
182569  25356504  load_video           /event     296 BGT_3cEcV64         \\N 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-26 23:22:50.933608 McGillX/Body101x/1T2015
182589  25356524  play_video           /event     296 BGT_3cEcV64          0  2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-26 23:22:57.237322 McGillX/Body101x/1T2015
182505  25356440 pause_video           /event     296 BGT_3cEcV64    62.5289 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-26 23:24:00.211174 McGillX/Body101x/1T2015
182472  25356407  play_video           /event     296 BGT_3cEcV64    62.5289 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-26 23:24:10.229123 McGillX/Body101x/1T2015
182574  25356509 pause_video           /event     296 BGT_3cEcV64    439.806 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-26 23:30:27.692353 McGillX/Body101x/1T2015
182576  25356511  stop_video           /event     296 BGT_3cEcV64    439.806 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-26 23:30:27.687487 McGillX/Body101x/1T2015
38999   25202947  load_video           /event     523 BGT_3cEcV64         \\N 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 22:58:29.615225 McGillX/Body101x/1T2015
38990   25202937  play_video           /event     523 BGT_3cEcV64          0  2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 22:58:31.587173 McGillX/Body101x/1T2015
39335   25203298  load_video           /event     523 BGT_3cEcV64         \\N 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 23:05:50.833405 McGillX/Body101x/1T2015
39360   25203323  play_video           /event     523 BGT_3cEcV64    435.792 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 23:05:56.347958 McGillX/Body101x/1T2015
39376   25203339 pause_video           /event     523 BGT_3cEcV64     439.61 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 23:06:00.709476 McGillX/Body101x/1T2015
39378   25203341  stop_video           /event     523 BGT_3cEcV64     439.61 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 23:06:00.731104 McGillX/Body101x/1T2015
302442  25486227  load_video           /event     578 BGT_3cEcV64         \\N 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-28 22:42:38.013591 McGillX/Body101x/1T2015
302470  25486255  play_video           /event     578 BGT_3cEcV64          0  2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-28 22:46:10.389749 McGillX/Body101x/1T2015
302662  25486451  stop_video           /event     578 BGT_3cEcV64    439.666 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-28 22:53:30.293322 McGillX/Body101x/1T2015
302663  25486453 pause_video           /event     578 BGT_3cEcV64    439.666 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-28 22:53:30.281381 McGillX/Body101x/1T2015
818961  26057427  load_video           /event     663 BGT_3cEcV64         \\N 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-03-09 21:54:55.005172 McGillX/Body101x/1T2015
1495962 26798824  load_video           /event     815       html5         \\N 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-03-22 14:57:28.002703 McGillX/Body101x/1T2015
1495965 26798830  play_video           /event     815       html5          0  2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-03-22 14:57:32.243349 McGillX/Body101x/1T2015
1496004 26798894 pause_video           /event     815       html5    439.573 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-03-22 15:02:26.524900 McGillX/Body101x/1T2015
1496006 26798896  stop_video           /event     815       html5    439.573 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-03-22 15:02:26.528739 McGillX/Body101x/1T2015
1546065 26853384  seek_video           /event     815       html5         \\N 2a87925a74fa474fb674c7fcaf53a594       414       440        \\N        \\N 2015-03-23 00:31:45.119869 McGillX/Body101x/1T2015
50677   25215700  load_video /segmentio/event     823      mobile         \\N 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 17:49:32.873000 McGillX/Body101x/1T2015
7883    25167215 pause_video /segmentio/event     823      mobile        0.1 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 17:49:41.874000 McGillX/Body101x/1T2015
7884    25167216  play_video /segmentio/event     823      mobile          0 2a87925a74fa474fb674c7fcaf53a594       \\N       \\N        \\N        \\N 2015-02-25 17:49:41.874000 McGillX/Body101x/1T2015
```

Figure 4.3 Structure des traces brutes extrait du serveur.

CHAPITRE 5 ENCODAGE SIVS DES INTERACTIONS VIDÉO

5.1 Introduction

Dans ce chapitre nous présentons le travail qui correspond à la première contribution mentionnée dans l'introduction de cette thèse (*"A Methodology for Student Video Interaction Patterns Analysis and Classification"*, In Proceedings of the International Conference on Educational Data Mining (EDM) (12th, Montreal, Canada, July 2-5, 2019). Dans les versions contemporaines des MOOC, les vidéos jouent un rôle important dans la diffusion du contenu des cours. Les vidéos représentent souvent le principal canal d'enseignement dans les MOOC, et éventuellement pour les environnements d'apprentissage à distance en général (voir par exemple pour la vidéo Breslow et al. 2013 ; Seaton, Bergner et al. 2014 ; Chauhan et Goel 2015). Les événements d'interaction vidéo tels que *play*, *pause*, *seek*, et *stop*, caractérisent la façon dont un étudiant a écouté une vidéo.

Bien qu'il existe un nombre important de recherches visant à déterminer les facteurs liés aux vidéos qui ont un impact sur l'engagement des étudiants (Bonafini 2017) et l'efficacité de l'apprentissage (Stöhr et al. 2019), nous trouvons encore relativement peu d'études sur la façon dont les étudiants interagissent avec les vidéos elles-mêmes, au-delà de simples agrégations d'événements de visionnage de vidéos (Hu et al. 2019 ; Wong et al. 2019). La codification des interactions vidéo des étudiants dans un système d'apprentissage en ligne est un défi non résolu pour les tâches d'analyses et principalement de classifications. Dans ce chapitre, notre proposition contribue à combler cette lacune méthodologique. Nous proposons une nouvelle représentation (codification) que nous appelons SIVS (*Sequence of Interaction in Vector Space*) et introduisons une méthodologie d'analyse qui permet d'étudier les divers patterns d'interaction avec les vidéos (séquences d'interaction avec les vidéos). Pour la validation de cette méthodologie, nous partons de l'hypothèse que les vidéos induisent des patterns d'interaction lorsque les étudiants interagissent avec elles. En d'autres termes, une vidéo spécifique a sa propre "signature" en termes de modèle de visionnage. Nous allons donc tester la force de la représentation d'interaction SIVS à rendre compte de cette hypothèse en la comparant à la force de la représentation cumulative d'activités vidéo des étudiants.

5.2 Méthodologie pour coder et classifier les interactions vidéo

Dans cette section, nous présentons la méthode utilisée pour coder et classer les interactions vidéo qui permet de comparer les séquences codées en termes de distance. Nous utiliserons le calcul de distance de Frobenius pour mesurer la distance entre des écoutes et pouvoir les classifier. Dans cette étude, nous utiliserons dans un premier temps des classes vidéo (données synthétiques) et les styles vidéo (données réelles) comme étiquettes cibles d'intérêt pour démontrer la méthodologie de classification. Il est question de déterminer si l'hypothèse selon laquelle une vidéo impose un style d'écoute reconnaissable aux étudiants peut se vérifier et à quelle hauteur dans l'encodage des interactions vidéo.

5.2.1 Encodage vidéo basé sur les mesures cumulatives d'écoute vidéo.

Une grande part des recherches portant sur l'analyse de vidéo repose sur la représentation d'écoute vidéo. On appelle mesures cumulatives d'écoute vidéo, un ensemble des paramètres d'écoute vidéo d'un étudiant qui spécifie sa façon d'interagir avec la vidéo à l'instar du temps total de visionnement de la vidéo, du temps total de pause, de nombre de recherches dans la vidéo (Brinton, Buccapatnam et al. 2016), le nombre de fois que l'étudiant est intervenu dans le forum (Klüsener et Fortenbacher 2015, Shaffer 2015), si l'étudiant a soumis ou pas un devoir (Boyer et Veeramachaneni 2015) et bien d'autres combinaisons de paramètres (Xing et Du 2019). Ces différents paramètres constituent ainsi la spécificité de la session d'écoute de l'étudiant. Des classifications et comparaisons unissent ces différents paramètres pour comparer ou classifier les sessions d'interaction vidéo des étudiants.

La limite d'une telle représentation est qu'elle ne tient pas en compte la succession réelle des activités vidéo dans l'interaction des étudiants. Dans la classification et la comparaison, une telle représentation peut déclarer semblables deux types d'interactions très différents pourvu que le temps soit le même (notamment le total du temps passé à jouer la vidéo et le temps de pause de la vidéo soient semblables pour ceux qui ont le même nombre de fois de recherches dans la vidéo). La nécessité de trouver une autre représentation qui puisse tenir compte de la succession des événements à l'intérieur des interactions vidéo se fait sentir surtout lorsqu'on recherche à classifier les styles d'écoutes.

Notre approche basée sur les centroïdes est comparée à l'approche de codage basée sur les mesures cumulatives. Il s'agit d'une approche fort rependue permettant de détecter des modèles uniques et de prédire des facteurs d'intérêt tels que le niveau d'engagement des étudiants,

l'abandon de la vidéo et l'achèvement ou l'abandon des cours (Sinha, Jermann et al. 2014).

Pour cette étude, nous avons reproduit cette méthode d'encodage basée sur les mesures cumulatives d'écoute vidéo des étudiants utilisée dans la littérature (Sinha, Jermann et al. 2014 ; Guo, Kim et Rubin 2014 ; N. Li, Kidzinski et al. 2015) pour la comparer à celle que nous proposons en vue de trouver une meilleure méthode d'encodage d'écoute vidéo des étudiants.

À partir de nos données brutes sur les clics des étudiants, nous élaborons un codage détaillé de ces clics dans les 5 catégories suivantes :

1. Lecture de la vidéo (Pl),
2. Pause de la vidéo (Pa),
3. Recherche en en avant dans la vidéo (Sf),
4. Recherche en arrière dans la vidéo (Sb),
5. Arrêt de la vidéo (St).

Cet encodage permet d'obtenir pour chaque étudiant un vecteur d'interaction avec une vidéo. Les éléments du vecteur représentent :

1. L'identifiant unique de l'étudiant,
2. Le pourcentage de temps passé à l'écoute de la vidéo (Pl) par rapport au temps total d'interaction avec la vidéo,
3. Le pourcentage de temps passé pendant que l'étudiant a fait une pause dans la vidéo (Pa) par rapport au temps total d'interaction avec la vidéo,
4. Le nombre de fois où l'étudiant a fait une recherche en rentrant la vidéo en arrière (Sb) par rapport à son point de départ,
5. Le nombre de fois où l'étudiant a fait une recherche en avançant la vidéo (Sf) par rapport à son point de départ,
6. Le nombre de fois où l'étudiant a arrêté la vidéo (St).

La Structure des données d'interaction vidéo étudiante basée sur les mesures cumulatives d'écoute utilisée dans notre base des données est sous forme vectorielle comme suit :

Vecteur d'encodage cumulative = [Identifiant étudiant, Pourcentage Pl, Pourcentage Pa, Nombre Sf, Nombre Sb, Nombre St]

La séquence d'activité de chaque étudiant est représentée par son vecteur d'encodage cumulative. Les analyses sur cette forme d'encodage utilisent ces types des vecteurs.

5.2.2 Passage de séquence d'activité à l'encodage SIVS

Nous proposons à partir des séquences d'activités un encodage des interactions étudiant avec la vidéo dans un espace vectoriel que nous nommerons SIVS (*Sequence of Interaction in Vector Space*). Dans l'encodage SIVS nous prenons chaque séquence d'activités des étudiants et nous étendons chaque cellule de l'activité étudiant sous forme d'un vecteur de longueur cinq (5). Les cinq éléments du vecteur représentent les diverses valeurs que peut prendre une cellule de la séquence d'activité comme "play", "pause", "seek backward", "seek forward", "stop". Le vecteur représentant une cellule de la séquence d'activité aura la valeur un (1) à la position correspondante à la cellule de sa valeur (voir équation 5.1). La représentation SIVS va donc simuler la navigation dans la vidéo de l'étudiant en alignant les activités l'étudiant en respectant la disposition dans la vidéo des activités de l'étudiant. Par exemple si un étudiant fait une recherche en arrière de cinq (5) transitions (correspond à une unité de temps par secondes définit en fonction du rythme de l'enregistrement des traces par le système), SIVS va ramener l'activité prochaine de l'étudiant à la position dans la vidéo qui correspond à un retour en arrière de cinq (5) transitions par rapport à sa position de départ tout en respectant le temps de référence choisi pour la représentation en fonction du nombre de transitions maximale par seconde. La figure 5.1 montre la procédure de codification de la séquence d'activités à la codification SIVS.

Afin d'obtenir la représentation SIVS, nous élargissons la séquence originale d'événements individuels en utilisant une représentation vectorielle des événements. Un événement est défini comme un vecteur de cinq (5) types d'événements défini dans la représentation SIVS.

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} = \begin{bmatrix} "play" \\ "pause" \\ "seek\ backward" \\ "seek\ forward" \\ "stop" \end{bmatrix} \tag{5.1}$$

La représentation de chaque élément de la matrice SIVS de l'activité vidéo se fait donc à travers un vecteur de longueur cinq (5) où la valeur une (1) indique la position de l'élément de l'activité dans l'encodage. Par exemple, une séquence d'activité, $\mathbf{s} = \{(\text{play})_1, (\text{play})_2, (\text{pause})_3, (\text{stop})_4\}$ sur un ensemble de cinq types d'événements qui comprend également des événements *seek forward* et *seek backward* (recherche avant/arrière), est représentée comme :

$$S = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4] = \begin{array}{c} \\ play \\ pause \\ seek\ backward \\ seek\ forward \\ stop \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{5.2}$$

Pour trouver la représentation d'une classe vidéo dans le cadre des données synthétiques d'interaction ou la représentation d'un style vidéo dans le cas des données réelles, nous définissons le centroïde de l'ensemble des écoutes vidéo. Par exemple, en supposant que nous avons un ensemble $\mathcal{S}$ de n séquences d'activité, où chaque séquence $\mathbf{s}$ est de longueur m ($|\mathcal{S}| = n$ et $|\mathbf{s}| = m$), nous définissons le centroïde de cet ensemble comme une matrice de m vecteurs définit comme suit : $\overline{\mathbf{s}} = [\overline{\mathbf{e}}_1, \overline{\mathbf{e}}_2, ..., \overline{\mathbf{e}}_m]$ où $\overline{\mathbf{e}}_i$ est défini comme

$$\overline{\mathbf{e}}_i = \begin{bmatrix} \sum_{j=1}^{n} e_{1,j} \\ \vdots \\ \sum_{j=1}^{n} e_{5,j} \end{bmatrix} /n \tag{5.3}$$

Dans chaque $e_{i,j,k}$ dans la définition d'une cellule d'un centroïde en général, i représente la position du chiffre dans la représentation SIVS du vecteur de la séquence donc varie de 1 à 5, j représente la position de la cellule dans la succession des transitions de la séquence d'activité donc varie entre 1 à n (pour un ensemble de n séquences) et k représente la position de la cellule de la séquence d'activité donc varie entre 1 à m (pour des séquences de m transitions). Le centroïde de cet ensemble est donc défini comme :

$$\overline{\mathbf{S}}(\mathcal{S}) = \begin{bmatrix} \sum_{j=1}^{n} e_{1,j,1} \ldots \sum_{j=1}^{n} e_{1,j,m} \\ \sum_{j=1}^{n} e_{2,j,1} \ldots \sum_{j=1}^{n} e_{2,j,m} \\ \sum_{j=1}^{n} e_{3,j,1} \ldots \sum_{j=1}^{n} e_{3,j,m} \\ \sum_{j=1}^{n} e_{4,j,1} \ldots \sum_{j=1}^{n} e_{4,j,m} \\ \sum_{j=1}^{n} e_{5,j,1} \ldots \sum_{j=1}^{n} e_{5,j,m} \end{bmatrix} /n \tag{5.4}$$

Un exemple plus complexe d'une séquence d'interaction entre un étudiant i et une vidéo de longueur n transitions (ou l'unité de temps retenu par le chercheur) peut être représenté comme dans la figure 5.1. Dans cet exemple, l'étudiant utilise des actions *seek* pour naviguer, et donc certaines parties de la séquence contiennent plus d'un seul événement *play* à un moment donné.

Le principe général consiste à définir le style de la vidéo par le *centroïde de la vidéo*, et par laquelle on peut calculer la distance par rapport à la séquence d'interaction vidéo d'un étudiant. Pour les distances entre les centroïdes puis entre la représentation SIVS de chaque séquence d'activité et un centroïde ou entre les représentations SIVS des séquences d'activités nous utilisons la distance de Frobenius comme expliqué en détail dans la section 5.2.3. Cela nous permettra de calculer la distance entre une séquence d'interaction de l'étudiant et chacun des différents centroïdes des vidéos et pouvoir ainsi déterminer à quelle vidéo appartient une séquence d'activité en particulier en fonction de sa distance (plus proche) du centroïde d'une vidéo.

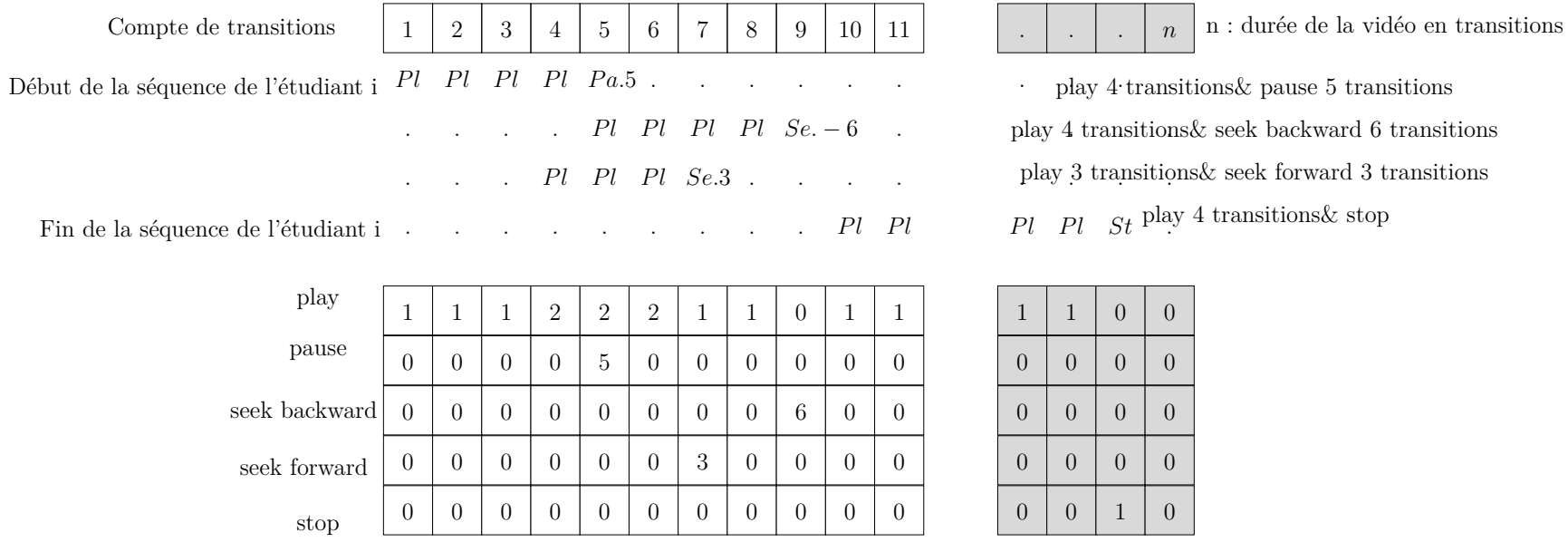
5.2.3 Utilisation des centroïdes pour définir les classes

Le centroïde d'un ensemble de séquences est défini comme une séquence prototypique et représente une classe des séquences. Dans l'expérience que nous allons effectuer sur les vidéos, cet ensemble sera les séquences d'interactions d'une vidéo. Une cellule de cette séquence de classe (centroïde) contient les moyennes des 5 types d'événements (définis dans l'équation 5.1). Cela permet de définir la distance entre une séquence individuelle, S, et un centroïde de vidéo $\overline{\mathbf{S}}$ (voir équation 5.4), comme la norme euclidienne (Frobenius) de la différence entre deux matrices : $\|\mathrm{S}-\overline{\mathbf{S}}\|$, puisque S et $\overline{\mathbf{S}}$ sont des matrices des mêmes dimensions.

Pour déterminer dans quelle mesure la séquence prototype de chaque vidéo est représentative et discriminante, nous allons utiliser les classificateurs comme le support de vecteur machine (*Support Vector Machine*), l'arbre décision (*Gradian Boosted Machine*) et l'approche du plus proche voisin (*K Nearest-Neighbour*) pour classer les séquences dans quatre vidéos de même longueur, comme le montre la figure 5.1.

Chaque séquence d'interaction des étudiants est codifiée selon SIVS. Il faut noter qu'en fonction de l'interaction, il peut avoir des passages au même endroit de la vidéo un, plusieurs fois ou jamais. La valeur dans la ligne correspondant à "Play", indique le nombre de fois que ce point de la vidéo a été écouté. Lorsqu'une colonne est constituée uniquement de zéro (0)

Figure 5.1 Représentation de l'interaction entre l'étudiant i et une vidéo d'une durée de n transitions.



ceci indique que l'étudiant n'a jamais écouté cette partie de la vidéo. En résumé, la valeur de chaque colonne à la ligne correspondant à "Play" constitue le nombre de fois que ce segment de la vidéo a été écouté.

Une telle représentation a pour avantage de montrer dans sa structure même les segments de la vidéo écoutée une ou plusieurs fois par un étudiant à la transition près (quelquefois la transition correspond à la seconde ou un passage est possible par le nombre de transitions par seconde connu). Cette information contenue dans la représentation de l'interaction permet de classifier des interactions semblables allant jusqu'à tenir compte des segments des vidéos écoutés par les étudiants et même les sections écoutées plusieurs fois ou pas du tout. On peut apercevoir tous les vas-et-viens dans la vidéo, les écoutes, les réécoutes et les sauts réalisés par un étudiant lors de ses interactions avec une vidéo à la transition près (voir figure 5.1).

5.2.4 Classification des interactions vidéo

L'évaluation de la puissance de la représentation SIVS est réalisée une fois qu'un centroïde vidéo est défini par $\overline{\mathbf{S}}$. La classification consiste à déterminer pour des données synthétiques à quelle classe (ensemble des données synthétiques) représentée par un centroïde, une représentation d'interactions vidéo appartient. Pour des données réelles, il s'agit à partir des calculs de distances avec les centroïdes des vidéos de classifier une représentation vidéo en particulier dans un style d'interactions de vidéo défini par le centroïde de chaque vidéo. En effet, le centroïde de chaque vidéo défini le style de la vidéo. Il s'agit donc de classifier les interactions vidéo dans une des catégories de style vidéo que constitue chaque centroïde. La classification d'une séquence d'interaction vidéo donnée S, consiste ainsi à trouver le centroïde le plus proche sur la base de la norme euclidienne : $\| S - \overline{\mathbf{S}} \|$. De plus, nous pouvons calculer un centroïde pour une vidéo à partir de l'ensemble de ses séquences d'écoute, et déterminer son style le plus proche en calculant la distance de ce centroïde avec les centroïdes des autres vidéos (pour la norme entre deux matrices voir les détails au chapitre 6, section 6.3.2).

5.3 Expérimentations

Nous appliquons la méthodologie décrite dans la section 5.2.3 pour vérifier l'hypothèse selon laquelle les vidéos induisent différents styles d'interactions. En d'autres termes, les vidéos imposent une façon d'interagir avec elles. Pour cette validation de l'hypothèse, nous allons procéder à une tâche de classification : étant donné un ensemble de séquences d'interactions vidéo, déterminer la vidéo correspondante à partir de la distance au centroïde de l'ensemble.

Il s'agit essentiellement de pouvoir tester la discrimination de la représentation SIVS proposée des interactions étudiantes avec les vidéos par rapport à la codification existante basée sur les mesures cumulatives (section 5.2.1).

La représentation SIVS étant sensible à la durée de la vidéo dans sa représentation, nos tests de validation vont prendre en compte cette limitation de la représentation pour qu'ils soient valides. Ainsi, le choix d'utiliser un ensemble de vidéos de même durée vise à éviter un biais dû à la durée : étant donné que la durée des vidéos influence la taille de la matrice SIVS, la classification des séquences dans telle ou telle classe pourrait se faire juste en fonction de la taille de la matrice. En ce sens, la durée de la vidéo peut introduire une distinction d'une vidéo à l'autre (pour les styles vidéo, voir le chapitre 4 section 4.4 et l'annexe E). La sélection des vidéos de même durée permet d'éviter ce biais potentiel (la taille des matrices de codification des interactions sera identique). Le choix également d'un style vidéo par ensemble des vidéos à identifier est pour éviter le biais que peut introduire un style vidéo en particulier dans la manière d'interagir avec elle. C'est pour éviter d'être en train d'identifier des styles vidéo au lieu d'identifier la vidéo avec laquelle un étudiant interagit.

5.3.1 Expérimentation 1 : données synthétiques

Pour tester la robustesse de la codification proposée par rapport à la méthode de codification basée sur les mesures cumulatives d'écoute, dans un premier temps, nous avons généré quatre ensembles de 100 séquences synthétiques de même durée en introduisant un pattern d'interaction différente (une signature) pour chaque vidéo. En effet, le pattern d'interaction introduit dans les séquences vidéo joue sur la succession des événements comme suit : 100 séquences d'activités pour chaque ensemble d'interactions. Chaque ensemble est caractérisé par son pourcentage de "play" dans ses transitions comme suit :

1. Le premier ensemble a 50% des transitions de chaque séquence qui sont des transitions "play",
2. Le second ensemble se reconnait avec 55% des transitions "play" dans chaque séquence,
3. Le troisième ensemble se reconnait avec 60% des transitions "play" dans chaque séquence,
4. Le dernier ensemble se reconnait avec 65% des transitions "play" dans chaque séquence

Il est question donc de créer des ensembles des séquences d'interactions ayant des signatures pour vérifier la capacité de chaque représentation à reconnaitre ces signatures.

En déterminant le centroïde de chaque ensemble, il est possible de vérifier à quelle hauteur par codification nous pouvons reconnaitre les classes d'interactions des différentes séquences. Nous traduisons en codage de mesures cumulatives chaque séquence d'interaction pour voir également la puissance de cette codification dans la vérification de l'hypothèse. Il s'agit alors de voir comment la codification proposée SIVS et la codification existante seront capables d'identifier à quel ensemble de données une séquence de l'ensemble test appartient en particulier. Nous avons dans un premier temps entraîné les classificateurs SVM, GBM et KNN (pour plus des détails sur les classificateurs utilisés voir l'annexe F). Ces classificateurs sont entraînés sur 90% des données et testés sur les 10% des données restantes.

Le choix de la signature d'interaction vidéo caractérisé par le pourcentage des transitions "play" n'est qu'un choix au hasard parmi tant d'autres choix possibles de signature d'écoute vidéo. En effet, nous avons essayé de faire varier le pourcentage de transitions "play" dans la détermination de la signature et nous avons eu des résultats similaires à savoir que la représentation SIVS performe toujours mieux que la représentation cumulative dans la reconnaissance de ces diverses signatures. C'est ainsi qu'un exemple en particulier a été choisi au hasard avec des pourcentages de transitions "play" arbitraires. Ce test peut être reproduit en changeant les pourcentages et les transitions tout en gardant le principe de créer une signature reconnaissable pour chaque ensemble d'interactions vidéo pour trouver que la représentation SIVS performe mieux que la représentation cumulative dans la reconnaissance des signatures dans les interactions vidéo.

5.3.2 Expérimentation 2 : données réelles

Dans une deuxième partie de nos expérimentations, nous utilisons les données réelles pour évaluer la codification SIVS. Bien que l'ensemble du cours que nous utilisons contienne 139 vidéos (Pour plus sur les vidéos voir le chapitre 4 et annexes A, C, D), nous en retenons 4 séries de 4 vidéos ayant dans chaque série les mêmes durées pour l'expérimentation. A chaque ensemble de 4 vidéos, les vidéos sont de même durée à la seconde près (entre 8 et 10 minutes). Et l'autre critère consiste à choisir pour chaque ensemble de 4 vidéos du même style. Les 4 ensembles des vidéos sont des styles différents correspondant à un style vidéo par ensemble de 4 vidéos. Nous avons ainsi couvert les principaux styles vidéo présents dans le cours (style présentation, style tête parlante, style interview et style image dans l'image) comme l'indique la table 5.1. Nous allons ainsi prédire la vidéo des séquences d'activités provenant des mêmes styles vidéo et de même durée. Pour plus de détails au sujet des styles vidéo, voir l'annexe E

et pour les divers styles vidéo présents dans nos données ainsi que le nombre des étudiants en moyenne qui ont interagi avec chaque style par vidéo, voir le résumé dans la table 5.1.

Tableau 5.1 Fréquence des styles vidéo (entre parenthèses, les fréquences retenues dans l'expérience) et nombre moyen d'étudiants qui ont interagi avec les vidéos.

Style Vidéo	**Nombre des Vidéos**	**Nombre moyen d'étudiants par Vidéo**
Tête parlante	79 (25)	1,953
Présentation	26 (13)	1,705
Image dans l'image	15 (8)	1,616
Interview	19 (8)	2,167

Afin de comparer l'approche SIVS avec l'approche basée sur les mesures cumulatives d'écoute (approche cumulative du temps passé dans chaque état de transitions en terme de pourcentage par rapport au temps total mis dans l'interaction avec la vidéo), nous reproduisons la tâche de classification en utilisant l'approche basée sur les mesures cumulatives (section 5.2.3). On considère dans cette recherche les mesures cumulatives d'écoute vidéo le vecteur contenant la proportion de chaque activité pour chaque vidéo, les proportions en temps cumulatif passent dans les événements comme *pause*, *play* et le nombre des *seek forward*, *seek backward*, *stop* qui sont calculées comme décrite précédemment. La classification d'une séquence d'interaction donnée dans l'une des 4 vidéos se fait à l'aide des approches du support de vecteur machine (SVM), l'arbre de décision (GBM) et l'approche du plus proche voisin (KNN). Cette classification est faite aussi bien pour l'encodage basé sur les mesures cumulatives d'écoute vidéo que pour l'encodage proposé SIVS.

5.4 Résultats

Ayant dans chaque analyse 4 classes (données synthétiques) ou quatre vidéos (données réelles), le résultat attendu de précision doit être au moins supérieur à 0.25 (précision nulle) qui représente la valeur de la précision pour la classification au hasard. Nous allons donc analyser les résultats de nos classifications en regardant les résultats avec les données synthétiques d'un côté et avec les données réelles de l'autre.

Classificateur :	SVM		GBM		KNN	
Approche :	SIVS	Cumul.	SIVS	Cumul.	SIVS	Cumul.
Accuracy	**1.00**	0.63	**1.00**	0.38	**1.00**	1.00
F_1	0.44	0.27	0.45	0.89	0.34	0.10
Kappa	**1.00**	0.00	**1.00**	−0.33	**1.00**	1.00

Tableau 5.2 Précision des séries de validation croisée de l'identification de la vidéo décuplée de quatre ensembles d'interactions des données synthétiques de même longueur en utilisant l'approche basée sur les mesures cumulatives d'écoute vidéo et la représentation SIVS.

SVM = Support de Vecteur Machine, GBM = L'arbre de décision, KNN = Plus Proche voisin, Cumul. = Approche cumulative du temps passé dans chaque état de transitions en terme de pourcentage par rapport au temps total mis dans l'interaction avec la vidéo.

5.4.1 Résultats des données synthétiques

Les résultats en utilisant les données synthétiques montrent qu'aussi bien avec la représentation basée sur les mesures cumulatives que la représentation SIVS, la précision est au-dessus celui du hasard qualifié de précision nulle (0.25). Ce premier constat permet de confirmer l'hypothèse du départ, à savoir les vidéos imposent une façon d'interagir avec elle. Dans les résultats obtenus avec chaque classificateur, la représentation SIVS est capable de mieux discriminer les classes des séquences d'interactions que la représentation basée sur les mesures cumulatives. Dans cette partie de l'expérimentation, les résultats montrent que lors de la classification des données, la représentation SIVS a une performance nettement supérieure à la méthode basée sur les mesures cumulatives. Dans beaucoup des cas, la représentation SIVS proposée avoisine une discrimination de 100% comme le montre le tableau 5.2.

5.4.2 Résultats des données réelles

Les résultats de la classification des données réelles sont présentés dans le tableau 5.3, montrent qu'aussi bien pour la représentation basée sur les mesures cumulatives que pour la représentation SIVS, les précisions sont au-dessus de la précision nulle. Ceci prouve l'hypothèse du départ selon laquelle les vidéos imposent une certaine façon d'interagir avec elles. Les résultats montrent que l'approche basée sur SIVS est nettement meilleure dans la distinction des interactions vidéo des étudiants que l'approche basée sur les mesures cumulatives. Ces résultats suggèrent que la méthode d'analyse des traces vidéo basée sur la représentation SIVS est plus précise dans la spécification d'interactions vidéo que la méthode basée sur les mesures cumulatives, du moins lorsqu'elle est appliquée à la tâche de reconnaissance des styles d'interactions.

Classificateur :	SVM		GBM		KNN	
Approche :	SIVS	Cumul.	SIVS	Cumul.	SIVS	Cumul.
Balanced Accuracy	**0.61**	0.51	**0.62**	0.60	**0.57**	0.50
F_1	**0.44**	0.27	**0.45**	0.42	**0.34**	0.10
Kappa	**0.23**	0.02	**0.23**	0.18	**0.13**	−0.02

Tableau 5.3 Précision des séries de validation croisée l'identification de la vidéo décuplée de quatre vidéos de même longueur en utilisant l'approche basée sur les mesures cumulatives d'écoute vidéo et la méthode proposée basée sur SIVS. Ceux qui sont rapportés ici sont les précisions moyennes des quatre ensembles différents des vidéos de même longueur dans chaque ensemble.

SVM = Support de Vecteur Machine, GBM = L'arbre de décision, KNN = Plus Proche voisin, Cumul. = Approche cumulative des temps passés dans chaque état de transition en termes de pourcentage par rapport au temps total mis dans l'interaction avec la vidéo.

5.5 Conclusions

Partant de l'hypothèse que les vidéos imposent une signature d'interactions avec elles, nous avons testé le pouvoir de la représentation basée sur les mesures cumulatives d'interaction vidéo pour corroborer à cette hypothèse. Nous avons comparé les résultats obtenus par la représentation basée sur les mesures cumulatives à ceux obtenus par la représentation SIVS que nous avons proposé. Les résultats montrent que la représentation SIVS que nous proposons à l'avantage de fournir un codage plus précis d'un ensemble d'interactions individuelles des étudiants autour d'un centroïde pour caractériser les styles d'interactions vidéo. La représentation SIVS permet de définir des styles d'interactions vidéo dans notre cas, qui contrastent avec les autres représentations qui ont du mal à définir autour d'une vidéo un style qui puisse regrouper toutes les interactions des étudiants avec la vidéo.

L'avantage d'une telle représentation en plus de mieux reconnaitre les styles d'interactions vidéo comparés à la représentation basée sur les mesures cumulatives, est le fait qu'à la transition près, SIVS rend compte de la navigation à l'intérieur de la vidéo d'un étudiant. Avec SIVS, l'on est en mesure de connaître les segments réécoutés de la vidéo, les segments non écoutés ou des pauses répétitifs en des lieux spécifiques de la vidéo etc.

La limite de cette représentation se trouve dans le fait qu'elle est sensible à la durée de

la vidéo dans sa structure même. Ainsi l'on ne pourrait pas, sans adaptation, analyser des écoutes venant des vidéos de durée différentes. On est limité, par sa forme originale, à des analyses des écoutes de la même vidéo ou des écoutes des vidéos de même durée. En somme, avec cette représentation la taille de la matrice de la représentation de l'écoute dépend de la durée de la vidéo. Une adaptation est nécessaire pour analyser des écoutes provenant des vidéos de durée différentes.

Pour dépasser cette limite de la représentation SIVS, il est nécessaire de pouvoir trouver une représentation insensible à la durée des vidéos pour analyser des écoutes vidéo de différente durée se fait sentir. Une analyse qui pourrait être faite par une représentation insensible à la durée des vidéos et comparée à d'autres représentations semblables dans les tâches d'analyse d'interactions provenant des vidéos de différentes durées serait une solution à la difficulté posée. Dans cette perspective, nous avons proposé une autre représentation insensible à la durée de la vidéo dans le chapitre 6 qui va être comparée aux autres représentations insensibles à la durée de la vidéo, actuellement, dans diverses tâches d'analyse, notamment dans la comparaison des styles d'écoutes vidéo.

CHAPITRE 6 ENCODAGE TMED POUR SIMILARITÉ D'ÉCOUTE VIDÉO

6.1 Introduction

Ce chapitre présente notre publication(*"Methodology to measure of similarity in student video sequence of interactions"*, In Proceedings of the International Conference on Educational Data Mining (EDM), 13th, July 10-13, 2020.) et notre travail en cours de publication (*"Interactions « discours de migrants, flux migratoires, espaces géographiques »"*, SFSIC 2020, 27-29 janvier 2021, Echirolles (France).), correspondant à la seconde et à la troisième contribution qu'on retrouve dans l'introduction de cette thèse. Elle porte sur la comparaison entre les façons d'interagir des étudiants avec les vidéos dans les contextes d'apprentissage en ligne. La mesure la plus utilisée dans la littérature concernant les interactions des étudiants est le temps total d'écoute de la vidéo qui a été utilisé comme mesure de l'engagement dans certains cas (Guo, Kim et Rubin 2014). Mais la disponibilité d'interactions détaillées avec une vidéo permet des mesures plus sophistiquées, et la comparaison entre les interactions vidéo plus intéressante. Les deux représentations courantes utilisées pour trouver la similarité entre les interactions vidéo sont la chaîne de Markov et les mesures de distance d'édition.

La principale limite de l'utilisation de la chaîne de Markov pour comparer des séquences d'interactions vidéo est que les probabilités de transition d'état ne prennent pas en compte le temps entre les états. De nombreuses séquences peuvent avoir la même matrice des probabilités des transitions mais représentant des interactions et des durées différentes.

En revanche, l'approche de la distance entre les séquences d'interactions vidéo prend en compte la durée des activités si les séquences d'événements sont placées sur une échelle de temps et sont représentées sous forme de segments d'activité, comme dans Michel Desmarais et François Lemieux 2013. Cependant, un décalage important (translation), comme une pause, dans des séquences d'activité similaires créera de grandes distances de montage qui feront de l'ombre à la similitude.

Une représentation qui peut prendre en compte simultanément la succession des transitions et le temps mis dans chaque état pourrait aider à l'analyse de la similarité des interactions vidéo. Elle pourrait aider à l'analyse des MOOC et des systèmes d'enseignement en ligne

dans des environnements à forte intensité vidéo, et pourrait aider à extraire des modèles significatifs d'interactions vidéo qui représentent une tâche difficile pour les chercheurs (voir par exemple : N. Patel, Sellman et Lomas 2017 ; Klingler et al. 2016 ; Boroujeni et Dillenbourg 2018).

6.2 Contexte

Nous passons d'abord en revue les bases des deux familles de représentations et expliquons plus en détail leurs limites, avant de décrire et d'évaluer la représentation proposée.

6.2.1 D'événements à une séquence d'activités

Les données sur l'interaction des étudiants avec les vidéos reposent sur la notion d'événements associés à des horodatages, tels que "play" à 0 :00 :00 et "pause" à 0 :00 :10. L'étudiant peut être considéré comme étant dans un état d'écoute d'une vidéo entre la seconde 0 et la seconde 10 et en état de pause par la suite.

Par exemple, supposons que nous ayons des interactions de deux étudiants comme suit :

— Étudiant 1 : "Play" (4 secondes) puis "Pause" (4 secondes) et ensuite "Play" (4 secondes),

— Étudiant 2 : "Pause" (2 secondes) puis "Play" (8 secondes) et ensuite "Pause" (2 secondes).

Chaque étudiant a passé 12 secondes en interaction totale avec la vidéo. Nous pouvons transformer ces deux modèles d'interaction en une séquence d'états d'activité d'intervalles de 1 seconde comme suit :

$$\begin{aligned} &\texttt{Séquence 1} : \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pa} - \texttt{Pa} - \texttt{Pa} - \texttt{Pa} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} \\ &\texttt{Séquence 2} : \texttt{Pa} - \texttt{Pa} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pl} - \texttt{Pa} - \texttt{Pa} - \texttt{Pa} \end{aligned} \tag{6.1}$$

(Pl=Lecture de la vidéo et Pa=Pause de la vidéo, séquence 1 pour l'étudiant 1 et séquence 2 pour l'étudiant 2)

Ces deux séquences d'activités peuvent également être représentées sous la forme d'une chaîne de Markov, comme nous le verrons ci-dessous.

6.2.2 La représentation sous la forme de la chaîne de Markov

Une chaîne de Markov est spécifiée par un ensemble d'états $S = \{s_1, s_2, s_3, ..., s_t\}$ et le processus commence dans un des états s_i, puis passe d'un état à l'autre s_j avec une probabilité de $p_{i,j}$. La caractéristique d'une chaîne de Markov est que la probabilité d'un état repose uniquement sur le dernier état, s_i. $p_{i,j}$ est alors appelé probabilité de transition. Le tableau spécifiant toutes les probabilités de transition entre les états est appelé *matrice des probabilités de transition*, ou *matrice de transition* (voir Grinstead et Snell GRINSTEAD et SNELL 2006, chapitre 11 pour une introduction).

Une séquence d'interaction d'un étudiant peut être représentée par une matrice de transition d'état de Markov, où les cellules contiennent les fréquences des transitions dans la séquence. Elle est normalisée de telle sorte que les sommes des lignes soient 1 et représentent donc les probabilités de transition. Une mesure de la distance entre les séquences peut être calculée à partir des deux matrices de Markov. Elle est définie comme la norme de Frobenius de la différence entre les matrices dans les cellules.

La limite de l'utilisation de la chaîne de Markov pour comparer des séquences d'interactions vidéo réside dans le fait que les probabilités de transition peuvent être les mêmes pour des séquences très différentes. Par exemple, les deux séquences de 6.1 provenant des deux scénarios d'interaction des étudiants (séquence 1 et séquence 2) ont la même matrice de Markov décrivant la probabilité de transition dans un espace à deux états :

$$\mathbf{M}_{\text{seq1}} = \mathbf{M}_{\text{seq2}} = \begin{array}{c} \\ play \\ pause \end{array} \begin{array}{c} \begin{array}{cc} play & pause \end{array} \\ \begin{pmatrix} 6/7 & 1/7 \\ 1/4 & 3/4 \end{pmatrix} \end{array}$$

Si nous prenions les chaînes de Markov basées uniquement sur les événements au lieu de les transformer en une séquence d'activités, le résultat serait le suivant :

$$\begin{aligned} &\texttt{Séquence1.1} : \texttt{Pl} - \texttt{Pa} - \texttt{Pl} \\ &\texttt{Séquence2.1} : \texttt{Pa} - \texttt{Pl} - \texttt{Pa} \end{aligned} \tag{6.2}$$

et nous obtiendrions deux chaînes de Markov identiques suivantes :

$$\mathbf{M}_{\text{seq1.1}} = \mathbf{M}_{\text{seq2.1}} = \begin{array}{c} \\ Play \\ Pause \end{array} \begin{array}{c} \begin{array}{cc} Play & Pause \end{array} \\ \begin{pmatrix} 0/1 & 1/1 \\ 1/1 & 0/1 \end{pmatrix} \end{array} \tag{6.3}$$

La similitude se maintiendrait entre $\mathbf{M}_{\text{seq1.1}}$ et $\mathbf{M}_{\text{seq2.1}}$ (forme usuelle de la chaîne de Markov) si des transitions supplémentaires entre "Pl" et "Pa" devaient se produire, car les cellules de "Pl" et "Pa" seraient peuplées de ratios similaires dans les deux matrices. Cette similitude masque donc la disparité importante lorsqu'on observe le temps des événements.

Par conséquent, chaque méthode a sa force, mais ne parvient pas à détecter les similitudes et les dissemblances évidentes à l'œil humain.

6.2.3 Distance de séquence d'édition : OM Distance

La distance de séquence d'édition (OM Distance : *"Optimal Matching Distance"*) sera utilisée dans le calcul de distance entre deux séquences en vue de déterminer leur degré de ressemblance. Si la distance entre deux séquences est de zéro (0) alors les deux séquences sont identiques alors que si la distance entre les deux séquences est égale au maximum de la longueur des deux séquences, alors les deux séquences sont complètement dissemblables et la dissemblance entre les deux séquences sera d'un (1).

La distance de séquence d'édition (distance ED) repose sur des mesures des distances entre les mots, où la similarité de l'alphabet entre les mots est la base du calcul de la similarité.

La distance d'édition (distance ED), génère des distances qui représentent le coût minimal en termes d'insertions, de suppressions et de substitutions pour transformer une séquence d'activité en une autre (voir l'annexe G la section G). Le coût de chaque suppression, insertion ou substitution est de 1 par défaut. Cet algorithme a été proposé à l'origine par Levenshtein 1966 et est le plus courant lors du calcul des distances entre les mots comme le remarque Navarro 2001. Pour les séquences d'écoute vidéo, le principe est le même mais l'alphabet est représenté par l'activité (les divers états d'interaction de l'étudiant avec la vidéo en tenant compte de la durée dans chaque activité traduit comme étant des transitions de l'état à lui-même comme l'illustre les exemples des séquences 1 et 2 précédentes). Par exemple, la distance de séquence d'édition pour les séquences 1 et 2 ci-dessus donne une distance de 9

sur un maximum de 12.

Une propriété notable de la distance de séquence d'édition (distance ED) entre des séquences de longueurs différentes est qu'elle ne peut pas être inférieure à la différence de durée des séquences d'activité. Cela peut poser un problème lorsque l'on compare l'interaction d'écoute vidéo des étudiants sur des vidéos. Il suffit qu'un étudiant maintienne la même interaction d'écoute, comme une pause fréquente, pour que la distance entre des séquences d'activités vidéos des étudiants puisse être plus grande.

Nous allons utiliser le fait que la distance d'édition minimale entre deux séquences est de zéro (lorsque deux séquences sont identiques) pour une similitude totale. La distance d'édition maximale entre deux séquences est égale à la longueur de la plus longue séquence entre les deux séquences (lorsqu'on a besoin de faire une substitution, une suppression ou une insertion à chaque cellule de la séquence pour transformer l'une en l'autre séquence) pour la dissimilarité totale. Nous allons utiliser ces deux propriétés extrêmes pour mesurer le degré de ressemblance entre deux séquences en utilisant la distance d'édition.

6.2.4 Performance de classification multi-classe

Dans le cadre de cette thèse nous rapportons des performances de classification de plusieurs classes. En général nous rapportons la moyenne des performances de toutes les classes. Pour ce faire, nous utilisons la matrice de confusion pour déterminer la performance de chaque classe. Nous prendrons l'exemple ici de la matrice de confusion pour trois classes qui peut être généralisée à n classes (avec $n \geq 2$). Voici la matrice de confusion pour trois classes (Figure 6.1) :

Il s'agit de trouver pour chaque classe les valeurs de vrai positif (VP), vrai négatif (VN), faux positif (FP), faux négatif(FN) et pouvoir déterminer les performances du classificateur pour chaque classe. Les principales mesures de performance que nous utilisons sont les suivantes pour chaque classe :

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \tag{6.4}$$

$$BalanceAccuracy = \frac{TVP + TVN}{2} \tag{6.5}$$

		Prédite		
		Classe A	Classe B	Classe C
Réel	Classe A	7	8	9
	Classe B	1	2	3
	Classe C	3	2	1

Figure 6.1 Exemple de matrice de confusion d'une classification en vue de déterminer les performances du classificateur dans le cas de 3 classes.

$$F_1 = \frac{2 \cdot Précision \cdot Rappel}{Précision + Rappel} = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \tag{6.6}$$

Où $TVP = \frac{VP}{VP+VN}$=Taux de Faux Positif, $TVN = \frac{VN}{VN+FP}$ = Taux de Vrai Négatif,
$Rappel = \frac{VP}{VP+FP}$ et $Précision = \frac{VP}{VP+FN}$.

A partir de notre matrice précédente (figure 6.1) à trois classes, nous déduirons les valeurs de performance suivante pour chaque classe :

1. **VP** : Le vrai positif de chaque est le nombre de vrai bien classifié. C'est la valeur de la diagonale de la matrice de confusion correspondant à chaque classe. Dans notre exemple de la figure 6.1 :
   classe A : VP = 7,
   classe B : VP = 2,
   classe C : VP= 1

2. **VN** : Le vrai négatif pour chaque classe correspond à la somme des valeurs excluant les valeurs de la colonne et ligne de la diagonale correspondant à la classe dans la matrice de confusion. Dans l'exemple de la figure 6.1 :

classe A : VN = (2+3+2+1)= 8,
classe B : VN = (7+9+3+1)=20,
classe C : VN= (7+8+1+2)=18

3. **FP** : Le faux positif correspond à la somme des valeurs sur la même ligne que la diagonale de la matrice de confusion correspondante a la classe excluant la valeur de la diagonale. Pour l'exemple de la figure 6.1 on obtient :
   classe A : FP = (8+9)= 17,
   classe B : FP = (1+3)=4,
   classe C : FP= (3+2)=5

4. **FN** : Le faux négatif correspond pour chaque classe à la somme des valeurs de la même colonne que la colonne de la diagonale de la matrice de confusion correspondante à la classe en question excluant la valeur de la diagonale. Dans notre exemple de la figure 6.1 nous aurons :

   classe A : FN = (1+3)= 4,
   classe B : FN = (8+2)=10,
   classe C : FN= (9+3)=12

Pour notre exemple de la figure 6.1 nous aurons donc comme résultat des performances pour chaque classe ce que montre la table 6.1.

Classe	Précision	Rappel	F_1-score
classe A	0.29	0.64	0.40
classe B	0.33	0.17	0.22
classe C	0.17	0.08	0.11

Tableau 6.1 Résultats de performance de classification par classe.

Dans les résultats que nous présentons, nous donnons les résultats de la moyenne de toutes les performances de toutes les classes. Par exemple, le F_1 que nous reportons dans nos tables, pour notre exemple de la figure 6.1 serait :

F_1 = F_1(classe A) +F_1(classe B)+F_1(classe B)/3 = (0.40 + 0.22 + 0.11)/3=0.24

Nous faisons la moyenne ainsi de toutes les mesures de performance. Cette méthode est connue dans la littérature comme *macro* F_1.

Lorsque nous avons le même nombre des données pour chaque classe (ce qui est le cas pour beaucoup de nos expérimentations), ces valeurs sont qualifiées de valeurs pondérées. Si le nombre des données sont différentes, alors la valeur pondérée par exemple du F_1 de l'exemple de la figure 6.1 est :

Valeur pondérée de F_1 = F_1(classe A)·11+F_1(classe B)·12+F_1(classe B)·13/(11 + 12 + 13) = ((0.40 · 11) + (0.22 · 12) + (0.11 · 13))/(11 + 12 + 13)

Nous avions onze (11) données dans la classe A, 12 données dans la classe B et 13 données dans la classe C dans notre exemple comme le montre la matrice de confusion. Plusieurs auteurs ont utilisé cette méthodologie d'une manière ou d'une autre pour la performance de leur classification dans le cadre de leurs recherches : A. C. Patel et Markey 2005 ; Diri et Albayrak 2008 ; Sokolova et Lapalme 2009 ; Patro et Patra 2015.

6.3 Représentation proposée, TMED

La représentation proposée, appelée TMED (*"Transition Matrix with Edit Distance features"*), est une combinaison de deux techniques : la matrice de transition qui prend en compte la durée de chaque activité de l'étudiant traduite en termes de transitions et la succession des transitions. La combinaison de ces deux éléments dans le cadre de la comparaison des écoutes vidéo donnerait une mesure de similitude plus acceptable entre chaque paire de séquences d'interactions des étudiants bénéficiant des avantages de ces deux techniques.

6.3.1 Construction de la matrice de transition

La matrice de transition vidéo d'un étudiant s pour une vidéo i est exprimée de telle sorte que chaque cellule de la matrice est la somme du nombre des transitions d'un état de la ligne à l'état de la colonne en tenant compte de la durée de chaque activité comme il sera expliqué en détail plus loin. Contrairement à la chaîne de Markov qui compte les transitions sans tenir compte de la durée des activités (voir la forme traditionnelle de la chaîne de Markov dans l'équation 6.3), la matrice des transitions dans le cadre de ce que nous proposons prend en

compte la durée de chaque activité comme expliqué ci-dessous. Nous exprimons une matrice de transition comme suit :

$$\mathbf{M}_{s,i} = \begin{array}{c} \\ load \\ play \\ pause \\ seek \\ stop \end{array} \begin{array}{c} \begin{array}{ccccc} load & play & pause & seek & stop \end{array} \\ \begin{pmatrix} a_{1.1} & a_{1.2} & a_{1.3} & a_{1.4} & a_{1.5} \\ a_{2.1} & a_{2.2} & a_{2.3} & a_{2.4} & a_{2.5} \\ a_{3.1} & a_{3.2} & a_{3.3} & a_{3.4} & a_{3.5} \\ a_{4.1} & a_{4.2} & a_{4.3} & a_{4.4} & a_{4.5} \\ a_{5.1} & a_{5.2} & a_{5.3} & a_{5.4} & a_{5.5} \end{pmatrix} \end{array}$$

Où $a_{j,k}$ est le nombre de transitions de l'événement j à l'événement k dans la séquence d'activité de type séquence 1 et 2 précédents (prenant en compte la durée de chaque activité) exprimés en nombre de transitions par unité de temps comme expliqué plus loin. Le choix de l'unité de temps discuté à la section 6.3.3.

Pour la construction de cette matrice de transition, nous quittons des interactions vidéo d'un étudiant. Au départ toutes les valeurs de transition sont à zéro. Un nouvel événement se produit lorsqu'un étudiant arrête une activité en commençant une autre. Par exemple, un étudiant met une vidéo en pause pendant qu'il la regarde. Le clic sur "pause" crée un nouvel événement qui passe de l'état "play" à l'état "pause" (changement de l'état "play" à l'état "pause"). Ce type de changement augmente d'une unité dans la matrice la cellule représentée par $a_{i,j}$ (i et j représentant deux états différents). Dans le cas où aucun événement ne se produit, cela signifie que l'étudiant est dans le même état : lecture de la vidéo ("Play") ou pause de la vidéo ("Pause") ou autre état, l'augmentation de l'élément de la matrice $a_{i,i}$ se fait en suivant le nombre maximum de transitions possibles lors du temps passé dans cet état comptant la transition d'un état à l'autre, ce qui correspond à l'idée de l'équation 6.7 dans la section suivante. Ensuite, chaque $a_{i,i}$ de la matrice représente à chaque fois le nombre de transitions possibles qui pourraient se produire lorsqu'un étudiant reste dans un état i sans passer à un autre état. Chaque fois qu'un étudiant atteint un état i, le compte de $a_{i,i}$ commence à augmenter sa valeur de l'unité de transitions par unité de temps passé dans cet état jusqu'à ce que l'étudiant clique pour passer à un autre état j et, dans ce cas, l'élément de matrice $a_{i,j}$ est augmenté de un et le compte dans l'état j de la matrice ($a_{j,j}$) commence à augmenter en fonction du temps passé dans j et ainsi de suite.

Pour éviter de perdre des transitions lors de la construction de la matrice de transition, nous allons trouver dans les traces enregistrées sur le serveur le plus petit intervalle de temps

que nous pouvons avoir entre deux événements. Cela s'exprime en nombre de transitions à la seconde ($N_{t/s}$). En utilisant cette référence, l'on est sûr de pouvoir représenter toutes les transitions des étudiants.

Le nombre total des transitions dans un état j lorsqu'un étudiant interagit avec une vidéo i et reste dans cet état une durée de $L_{s1,i,j}$ en seconde sachant que le système est capable de collecter $N_{t/s}$ transition par seconde est exprimé comme :

$$T_{s1,i} = L_{s1,i} * N_{t/s} \tag{6.7}$$

Où $T_{s1,i}$ est le nombre de transitions de l'étudiant $s1$ dans son interaction avec la vidéo i dans l'état j,
$L_{s1,i,j}$ est la durée de temps passé par l'étudiant $s1$ dans l'état j en seconde dans son interaction avec la vidéo i,
$N_{t/s}$ est le nombre maximum de transitions par seconde collecté par le système.

Lorsqu'un étudiant est dans un état j, on augmente la cellule $a_{j,j}$ de $T_{s1,i,j}$ transitions lorsque l'étudiant passe à un autre état. Dans notre cas, le nombre maximum de transitions par seconde collecté par notre système est de trois (3) transitions par seconde. Pour trouver ce nombre, il faut considérer toutes les données collectées et trouver le nombre maximum d'événements vidéo qui se produisent dans la même seconde. La normalisation est la prise en compte du temps passé dans chaque état par un étudiant y compris les périodes de transition d'un état à l'autre.

Le nombre total des transitions d'un étudiant $s1$ est exprimé comme suit :

$$T_{s1,i} = \sum_{i=1}^{5} \sum_{j=1}^{5} a_{i,j} \tag{6.8}$$

Avec $a_{i,j}$ l'élément de la ligne i et de la colonne j de la matrice de transition de l'étudiant $s1$.

6.3.2 Distance entre deux matrices : distance matricielle

Dans nos recherches, nous avons conservé cinq états propres aux interactions vidéo des étudiants : *"load"* (la vidéo est chargée dans l'interface prête à être joué), *"play"*(l'étudiant regarde la vidéo), *"pause"* (l'étudiant a mis en pause la vidéo), *"seek"* (l'étudiant déplace le curseur de lecture de la vidéo pour chercher), *"stop"* (l'étudiant arrête la vidéo). La matrice de transitions va donc être une matrice cinq par cinq dont chaque colonne et chaque ligne représente l'un des cinq états.

La distance entre deux matrices d'interaction vidéo est exprimée comme une distance de Frobenius qui est calculée comme suit :

$$d(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i}) = \|\mathbf{M}_{s1,i} - \mathbf{M}_{s2,i}\|_F = \sqrt{\sum_{j=1}^{5}(\sum_{k=1}^{5}(a_{j,k} - b_{j,k})^2)} \tag{6.9}$$

Où

$$\mathbf{M}_{s1,i} = \begin{pmatrix} a_{1.1} & a_{1.2} & a_{1.3} & a_{1.4} & a_{1.5} \\ a_{2.1} & a_{2.2} & a_{2.3} & a_{2.4} & a_{2.5} \\ a_{3.1} & a_{3.2} & a_{3.3} & a_{3.4} & a_{3.5} \\ a_{4.1} & a_{4.2} & a_{4.3} & a_{4.4} & a_{4.5} \\ a_{5.1} & a_{5.2} & a_{5.3} & a_{5.4} & a_{5.5} \end{pmatrix}$$

$$\mathbf{M}_{s2,i} = \begin{pmatrix} b_{1.1} & b_{1.2} & b_{1.3} & b_{1.4} & b_{1.5} \\ b_{2.1} & b_{2.2} & b_{2.3} & b_{2.4} & b_{2.5} \\ b_{3.1} & b_{3.2} & b_{3.3} & b_{3.4} & b_{3.5} \\ b_{4.1} & b_{4.2} & b_{4.3} & b_{4.4} & b_{4.5} \\ b_{5.1} & b_{5.2} & b_{5.3} & b_{5.4} & b_{5.5} \end{pmatrix}$$

Ici i représentant la vidéo considérée puis $s1$ et $s2$ les deux étudiants qui interagissent avec la même vidéo i (ou des vidéos differentes).

6.3.3 Similarité TMED (S_{mat}) entre deux matrices

Dans le cadre de la mesure de degré de similarité entre deux matrices d'interactions vidéo, nous allons utiliser la distance matricielle entre elles. L'objectif de cette mesure étant de comparer les degrés de similarité obtenus par diverses représentations d'interactions vidéo

pour trouver celles qui expriment mieux la similarité entre deux écoutes.

La distance maximale entre deux matrices de Markov ou des matrices TMED pour une vidéo est atteinte lorsque la première matrice n'a aucune transition en commun avec la seconde. Dans un tel cas, la dissimilarité entre ces deux matrices est maximale d'une valeur de 1 (la dissimilarité est exprimée entre 0 et 1). La dissimilarité entre deux matrices TMED ou deux matrices de Markov est égale à zéro (0) lorsque les deux matrices sont identiques (la distance entre les deux matrices est nulle). La dissimilarité entre deux matrices TMED ou deux matrices de Markov est égale à 1 lorsque les deux matrices n'ont aucune transition en commun. Donc, lorsqu'une transition dans une des matrices est différente de zéro, la même transition dans l'autre matrice est nécessairement nulle et vice-versa. Les valeurs correspondantes peuvent être par contre nulles pour les deux matrices. Nous nommons $dist_{max}$ la distance entre les matrices n'ayant aucune transition commune et est exprimée comme suit :

$$dist_{max}(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i}) = \sqrt{\sum_{i=1}^{5}(\sum_{j=1}^{5} a_{i,j}^2) + \sum_{i=1}^{5}(\sum_{j=1}^{5} b_{i,j}^2)} \tag{6.10}$$

Avec $a_{i,j}$ les cellules de la première matrice et $b_{i,j}$ les cellules de la seconde. Ainsi la valeur de la dissimilarité est de un (1) dans cette situation. Toutes les distances entre les matrices vont être divisées par la distance maximale ($dist_{max}$) pour obtenir le degré de dissimilarité entre elles.

La dissimilarité entre deux séquences d'interaction est exprimée en pourcentage en utilisant les matrices de transition des deux séquences. On définit donc la dissimilarité entre les deux matrices de l'équation 6.9 comme suit :

$$Dis(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i}) = \frac{d(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i})}{\sqrt{\sum_{i=1}^{5}(\sum_{j=1}^{5} a_{i,j}^2) + \sum_{i=1}^{5}(\sum_{j=1}^{5} b_{i,j}^2)}} \tag{6.11}$$

Où $Dis(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i})$ est le niveau de dissimilitude entre $\mathbf{M}_{s1,i}$ et $\mathbf{M}_{s2,i}$ deux matrices d'interactions de deux étudiants avec la vidéo i et $d(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i})$ est la distance entre elles. Les valeurs $a_{i,j}$ et $b_{i,j}$ sont les cellules respectivement des matrices de transitions $\mathbf{M}_{s1,i}$ et $\mathbf{M}_{s2,i}$. Si $Dis(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i})$ est égal à zéro (0), alors les deux séquences sont complètement similaires (la distance nulle entre les deux matrices de transition) et, lorsqu'il est égal à 1 (les deux matrices de transitions n'ont aucune transition en commun : deux cellules de la même po-

sition dans les deux matrices ne peuvent pas être différentes de zéro en même temps), elles sont complètement dissemblables. Entre 0 et 1 indique le pourcentage de dissimilarité entre les deux séquences de transitions.

La similarité entre deux séquences de vidéo d'interaction basées sur des matrices de transition avec une vidéo i est alors exprimée comme suit :

$$S_{mat}(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i}) = 1 - Dis(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i}) \tag{6.12}$$

Ici, $S_{mat}(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i})$ est le niveau de similarité entre la séquence d'interaction de l'étudiant $s1$ et de l'étudiant $s2$ de la vidéo i en utilisant la matrice d'interactions et $Dis(\mathbf{M}_{s1,i}, \mathbf{M}_{s2,i})$ est la dissimilitude elles.

6.3.4 Similarité OM Distance (S_{om}) entre deux séquences

Pour chaque paire de séquences étendues prenant en compte la durée de chaque activité du style 6.1, nous calculons la distance ED et à partir de là, nous calculons le niveau de similarité entre elles : la similarité OM Distance (S_{om}).

Le niveau de similarité entre deux séquences est calculé en utilisant la distance ED comme suit :

$$S_{om}(seq_{s1,i}, seq_{s2,i}) = 1 - \frac{dist_{om}(seq_{s1,i}, seq_{s2,i})}{max(T_{s1,i}, T_{s2,i})} \tag{6.13}$$

Où $S_{om}(seq_{s1,i}, seq_{s2,i})$ est le niveau de similarité entre la séquence de l'étudiant $s1$ et la séquence de l'étudiant $s2$ de la vidéo i et $dist_{om}(seq_{s1,i}, seq_{s2,i})$ est la distance ED en comptant les insertions, les substitutions et les suppressions pour rendre semblable dans l'algorithme OM (voir *Optimal Matching Distance* de l'annexe G à la section G) entre les deux séquences et $T_{s1,i}$ et $T_{s2,i}$ sont les nombres de transitions de la séquence de chaque étudiant donné dans l'équation 6.8. Enfin, $max(T_{s1,i}, T_{s2,i})$ est le maximum entre le nombre des transitions des deux séquences d'interactions des étudiants.

6.3.5 Similarité (S_{mark}) entre deux matrices de Markov

Le degré de similarité entre deux matrices de Markov est calculé en fonction de la distance entre deux matrices comme définie dans l'équation 6.9. Il faut bien noter ici qu'une matrice de Markov est différente d'une matrice TMED (la somme de chaque ligne d'une chaîne de Markov est 1 alors que la somme totale de toutes les valeurs d'une matrice TMED est de 1). Dans le calcul du degré de similarité des chaînes de Markov, nous prenons aussi en compte la distance maximale possible entre les deux chaînes. La similarité entre deux matrices de Markov va donc être définie comme :

$$S_{mark}(\mathbf{M'}_{s1,i}, \mathbf{M'}_{s2,i}) = 1 - \frac{d(\mathbf{M'}_{s1,i}, \mathbf{M'}_{s2,i})}{\sqrt{\sum_{i=1}^{5}(\sum_{j=1}^{5} {a'_{i,j}}^2) + \sum_{i=1}^{5}(\sum_{j=1}^{5} {b'_{i,j}}^2)}} \tag{6.14}$$

Où $S_{mark}(\mathbf{M'}_{s1,i}, \mathbf{M'}_{s2,i})$ est le niveau de similitude entre les chaînes de Markov $\mathbf{M'}_{s1,i}$ et $\mathbf{M'}_{s2,i}$ deux matrices d'interactions de deux étudiants avec la vidéo i et $d(\mathbf{M'}_{s1,i}, \mathbf{M'}_{s2,i})$ est la distance entre eux. Les valeurs $a'_{i,j}$ et $b'_{i,j}$ sont les cellules respectivement des matrices de transitions $\mathbf{M'}_{s1,i}$ et $\mathbf{M'}_{s2,i}$.

6.4 Étapes de la mesure de la similarité matricielle proposée

Pour récapituler la méthodologie de mesure de similarité que nous proposons en introduisant une représentation d'interaction vidéo, nous voulons dans cette section résumer les étapes pour y arriver. Nous allons ensuite comparer la capacité de la représentation TMED pour trouver le niveau de similarité entre les séquences comparée aux représentations existantes, à savoir la technique de la chaîne de Markov telle qu'utilisée par Klingler et al. 2016 et la représentation de séquence d'interaction vidéo étendue sensible à la durée des activités basée sur la distance édition (du style des séquences (6.1) dans la section 6.2.1).

Voici en somme les étapes dans la détermination de la similarité que nous proposons entre deux séquences d'écoute vidéo.

(i) **Convertir le temps passé dans les états en période de transition :** il s'agit pour chaque interaction étudiant avec la vidéo de transformer en termes de transitions les temps passés dans chaque activité d'interaction en utilisant l'équation 6.7 pour déterminer l'intervalle appropriée.

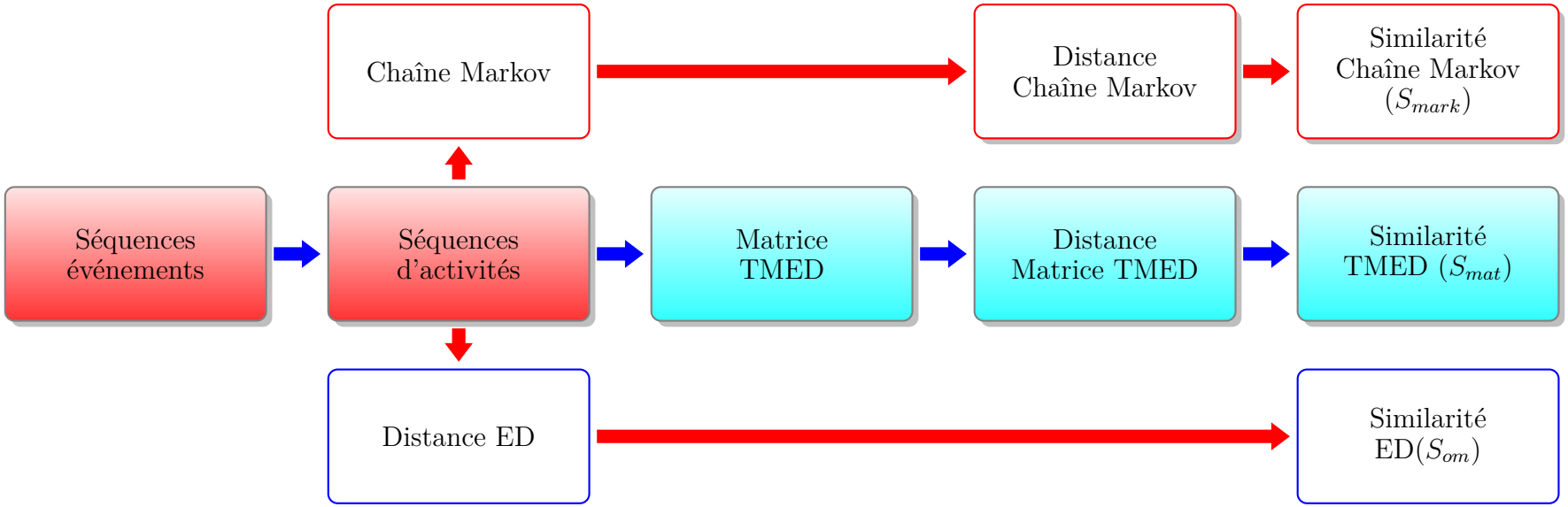
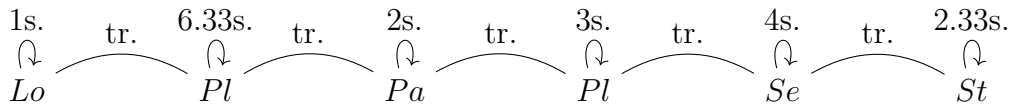
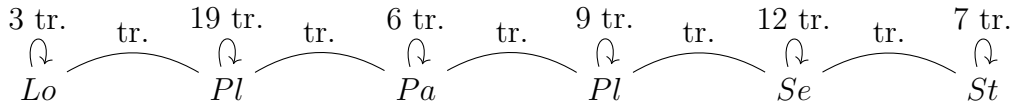


Figure 6.2 Flux de la méthode proposée pour calculer les similarités entre les séquences vidéo des étudiants. Nous obtenons les trois similarités (ED, Markov et TMED) en suivant le flux des opérations.

Prenons un exemple de séquence d'interaction vidéo d'étudiant suivant :



En prenant en compte le fait que dans notre cas $N_{t/s} = 3\ transitions/seconde$, la séquence d'interaction devient :



(ii) **Déduire la matrice de transition :** il s'agit à partir de la séquence d'interactions converti en termes de transitions de construire une matrice représentant toutes les valeurs de transitions d'un état à l'autre y compris les transitions d'un état à lui-même. Pour l'exemple précédent nous obtenons la matrice de transition suivante.

$$\mathbf{TM} = \begin{array}{c} \\ load \\ play \\ Pause \\ Seek \\ Stop \end{array} \begin{array}{c} \begin{array}{ccccc} load & Play & Pause & Seek & Stop \end{array} \\ \begin{pmatrix} 3 & 1 & 0 & 0 & 0 \\ 0 & 28 & 1 & 1 & 0 \\ 0 & 1 & 6 & 0 & 0 \\ 0 & 0 & 0 & 12 & 1 \\ 0 & 0 & 0 & 0 & 7 \end{pmatrix} \end{array}$$

(iii) **Déduire la matrice TMED représentant les interactions et la matrice de Markov de cette séquence :** Il s'agit de déduire la matrice de probabilité de transition à partir de la matrice de transition (chaîne de Markov) et la matrice de TMED en fonction des toutes les transitions. La matrice TMED est obtenue en divisant chaque élément de la matrice de transition par le nombre total des transitions. Dans l'exemple

précédent, la matrice TMED est exprimée comme suit :

$$\mathbf{TMED} = \begin{array}{c} \\ load \\ play \\ Pause \\ Seek \\ Stop \end{array} \begin{array}{c} \begin{array}{ccccc} load & Play & Pause & Seek & Stop \end{array} \\ \begin{pmatrix} 3/61 & 1/61 & 0 & 0 & 0 \\ 0 & 28/61 & 1/61 & 1/61 & 0 \\ 0 & 1/61 & 6/61 & 0 & 0 \\ 0 & 0 & 0 & 12/61 & 1/61 \\ 0 & 0 & 0 & 0 & 7/61 \end{pmatrix} \end{array}$$

La matrice de chaîne de Markov correspondante est :

$$\mathbf{M_{mark}} = \begin{array}{c} \\ load \\ play \\ Pause \\ Seek \\ Stop \end{array} \begin{array}{c} \begin{array}{ccccc} load & Play & Pause & Seek & Stop \end{array} \\ \begin{pmatrix} 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 28/30 & 1/30 & 1/30 & 0 \\ 0 & 1/7 & 6/7 & 0 & 0 \\ 0 & 0 & 0 & 12/13 & 1/13 \\ 0 & 0 & 0 & 0 & 7/7 \end{pmatrix} \end{array}$$

La particularité du TMED par rapport à la chaîne de Markov est le fait qu'elle prend en compte dans son expression toute la séquence de l'interaction vidéo de l'étudiant en même temps. Ce qui fait que dans TMED nous avons cette égalité :

$$\sum_{i=1}^{5} \sum_{j=1}^{5} b_{i,j} = 1 \tag{6.15}$$

Où $b_{i,j}$ est la cellule de la matrice TMED à la ligne i et la colonne j.

Pendant que la chaîne de Markov est exprimée par rapport à la probabilité de sortie de chaque état, TMED quant a elle est exprimée par rapport au nombre de transitions de toute la séquence d'interaction de l'étudiant avec la vidéo. Donc la somme des éléments de chaque ligne de la matrice d'une chaîne de Markov est égale à un (1) :

$$\sum_{j=1}^{5} b_{i.j} = 1, (i \in \{1, 2, 3, 4, 5\}) \tag{6.16}$$

Où $b_{i,j}$ est la cellule de la matrice de la chaîne de Markov à la ligne i et la colonne j.

(iv) **Calcul de la distance entre deux matrices TMED et deux matrices de Markov :** il s'agit d'utiliser à cette étape le calcul de la section 6.3.2 pour exprimer la distance entre les matrices TMED et la distance entre deux chaînes de Markov (voir l'équation 6.9).

(v) **Calcul de la similarité TMED et calcul de la similarité de la chaîne de Markov :** il s'agit d'utiliser ici les calculs de la section 6.3.3 pour déterminer le degré de similarité entre deux séquences d'écoutes à partir de leurs matrices TMED (voir l'équation 6.12). Puis de calculer le degré de similarité entre deux matrices de Markov en utilisant l'équation 6.14.

(vi) **Calcul de la similarité d'édition (ED) entre deux séquences d'activités :** il s'agit à partir de la distance d'édition entre deux séquences d'activités, de calculer le degré de similarité entre elles. Pour cela, nous utilisons l'équation 6.13 définie précédemment dans la section 6.3.4.

L'enchainement schématique de la procédure de détermination du degré de similarité entre deux séquences d'interactions se retrouve à la figure 6.2. Il s'agit de quitter de deux séquences sous leur forme étendue qui tient compte de la durée des activités pour aboutir à leur degré de ressemblance pour chaque type de représentation.

6.5 Validation de la méthodologie de similarité proposée

Pour valider la méthodologie de similarité proposée, nous allons procéder dans un premier temps au calcul de la similarité entre des séquences dont nous savons à l'avance le résultat de la similarité (cas prototypiques). Nous comparerons les résultats de la similarité entre la représentation proposée avec la similarité basée sur les représentations en chaînes de Markov et de la séquence d'édition pour voir laquelle de ces trois façons de représenter donne un degré de similarité semblable à celle attendue.

Dans la seconde partie de la validation nous allons prendre les données réelles pour voir

laquelle de ces trois représentations d'interactions vidéo (chaîne de Markov, séquence d'interaction et TMED) est mieux capable de discriminer d'un côté les séquences des étudiants et de l'autre les vidéos. Il s'agit en effet de voir la capacité de chaque représentation à corroborer les hypothèses selon lesquelles un étudiant a une "signature" propre d'interactions avec les vidéos et de l'autre qui a été vérifiée au chapitre 5, avec la représentation SIVS selon laquelle une vidéo impose une "signature" d'interaction avec elle. Nous comparerons les performances en utilisant ces trois types de représentations.

6.5.1 Cas prototypiques

Nous testons d'abord l'approche sur des cas prototypiques où les modèles sont évidents à l'œil nu. Pour ce faire, nous prenons deux cas principaux : des séquences de transitions de même longueur et des séquences de transitions de longueur différente. Pour des séquences d'interactions de même longueur, nous avons considéré une séquence cyclique de transitions de même longueur, comme l'illustre la figure 6.3. Le cycle des transitions est le suivant :

$$\texttt{Lo} - \texttt{Pl} - \texttt{Pa} - \texttt{Pl} - \texttt{Pa} - \texttt{Se} - \texttt{Pl} - \texttt{St} \tag{6.17}$$

Le cycle de transition peut commencer n'importe où et se terminer par St pour n'importe quelle séquence (figure 6.3). Le niveau de similarité attendue devrait être proche de 100 %, au moins plus que 80%, pour un même cycle d'activités. Le résultat basé sur la distance ED ne peut pas trouver ce niveau de similarité comme le montre la figure 6.4(a) par rapport à la méthode de Markov de la figure 6.4(b) où les résultats sont de l'ordre de ce qui est attendu avoisinant ainsi les 100% comme prévu et la représentation TMED proposée dans la figure 6.4 (c) trouve une similarité semblable à la chaîne de Markov avoisinant les 100 % comme prévu. Ici la représentation TMED proposée bénéficie de la propriété des chaînes de Markov qui tient compte de la succession d'activité dans sa structure.

Pour ces séquences cycliques, la méthode proposée et les méthodes de similarité basées sur la chaîne de Markov sont plus performantes que la méthode basée sur la méthode de distance ED (en raison de la disposition des états non concordants) pour trouver la similarité entre deux séquences d'interactions cycliques identiques.

La deuxième validation de la représentation proposée consiste à comparer ses capacités à celles de la méthode de Markov et la distance ED dans des séquences de longueur différente

avec des similarités connues. À cette fin, nous avons considéré quatre séquences de mêmes niveaux de transition, comme le montre la figure 6.5. Dans ce cas, le pourcentage de transition entre les états est le même, mais le temps passé dans chaque état est différent d'une séquence à l'autre. Le niveau de similarité attendu dépend ici de la longueur de chaque séquence, car la succession des états est la même pour les quatre séquences (figure 6.5). Ainsi, le degré de similarité de deux positions sur la même diagonale sera la même en partant des plus faibles jusqu'au 100% de la diagonale centrale de la matrice. En effet, selon la figure 6.5 la longueur de la séquence 1 est la moitié de la séquence 2 et longueur de séquence 2 est la moitié de la séquence 3 et la séquence 3 la moitié de la longueur 4. Donc le degré de similarité attendu entre les séquences 1 et 2, entre les séquences 2 et 3 puis entre les séquences 3 et 4 devrait être le même. D'où les diagonales de la matrice de similarité à chaque niveau devraient montrer le même degré de similarité. Nous devrions également avoir comme résultat une augmentation progressive du niveau de similarité de la séquence la plus courte à la plus longue.

Le résultat de la méthode basée sur la séquence d'édition dans la figure 6.6(a) a du mal à retrouver la similarité progressive attendue à cause de la variation de longueur des séquences, les opérations d'insertions et de substitutions sont inégalement réparties d'une séquence à l'autre. Contrairement à ce qui est attendu, les degrés de similarité sur la même diagonale sont différents. Le résultat de la méthode basée sur la chaîne de Markov, comme on peut le voir dans la figure 6.6(b), n'a pas pu non plus trouver les différents niveaux de similarité car le pourcentage de transition entre les états est conservé avec des longueurs de séquences différentes. La méthode proposée donne de meilleurs résultats, comme le montre la figure 6.6(c), car elle est basée sur le nombre de transitions (en prenant en compte la durée de temps dans chaque état) plutôt que sur la probabilité de transition comme l'est la chaîne de Markov. Comme on le prévoyait, on voit bien que les degrés de similarité d'une position à la même diagonale sont égaux.

6.5.2 Validation de la sensibilité de la représentation TMED proposée

Pour la première partie de la validation, nous allons identifier les étudiants auxquels le TMED appartient. Le but de l'expérience est d'évaluer la sensibilité de la représentation à différents styles d'interaction. En effet, nous voulons savoir jusqu'à quelle hauteur l'hypothèse selon laquelle un étudiant a sa propre signature d'interaction peut se vérifier avec la représentation TMED en comparant sa capacité à celle des deux autres représentations (séquence et chaîne de Markov). Si l'approche est en mesure d'identifier mieux que les deux autres à quel étudiant

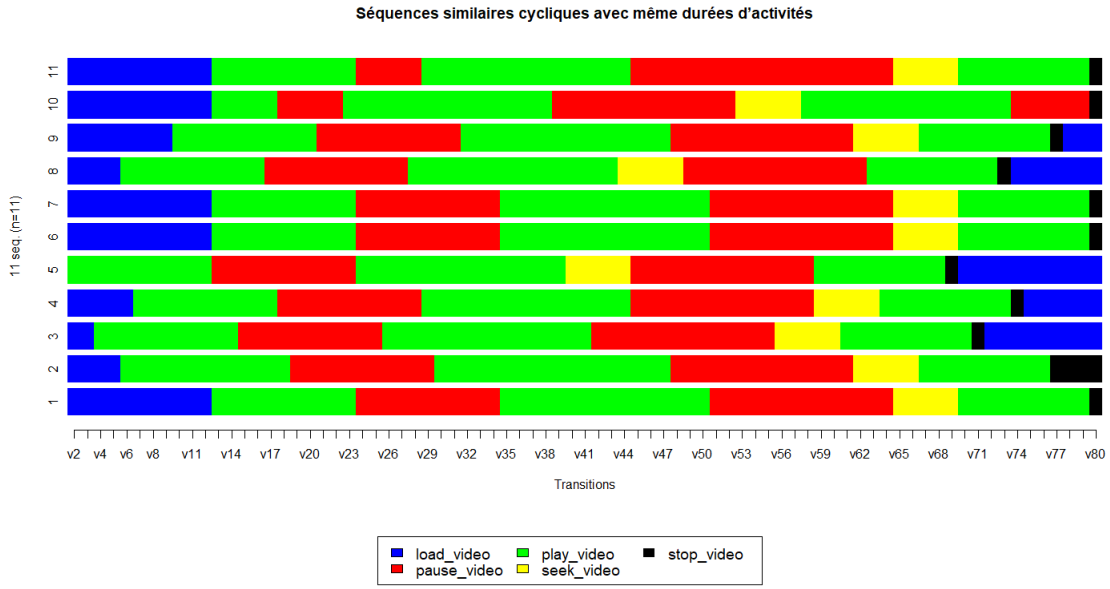
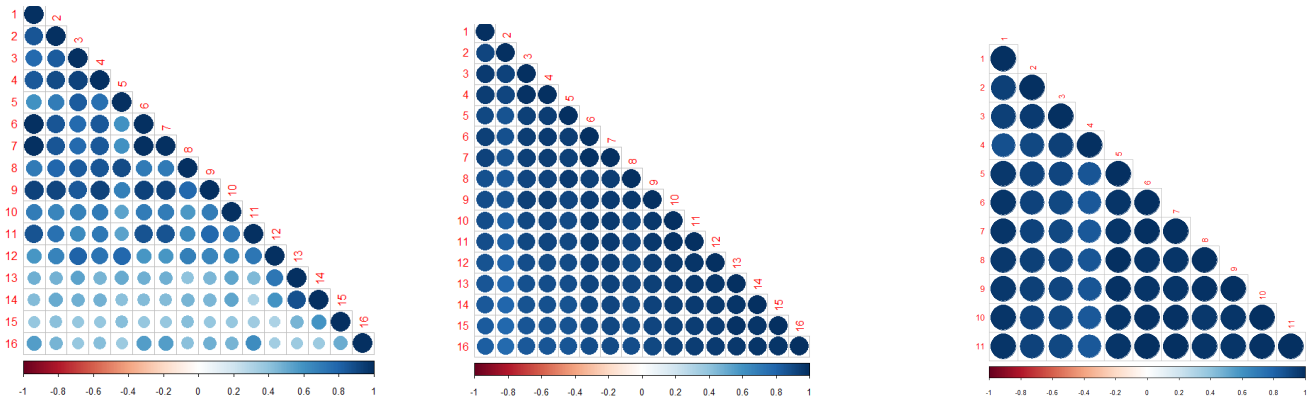


Figure 6.3 Même séquence cyclique de transitions. Le cycle peut commencer à différentes étapes mais suit le même schéma de transition.



(a) ED (b) Markov (c) TMED

Figure 6.4 Résultat de la similarité : (a) La similarité basée sur la distance d'édition (ED) ne peut pas reconnaître la similarité des séquences cycliques. (b) La similarité basée sur la chaîne de Markov peut reconnaître la similarité, comme on s'y attendait, les similarités avoisinent 100%. (c) La similarité basée sur la représentation TMED proposée peut reconnaître les séquences cycliques. Également les similarités avoisinent 100%

appartient le TMED, nous supposons qu'elle est plus sensible et plus susceptible de mieux discriminer des styles différents d'interactions vidéo.

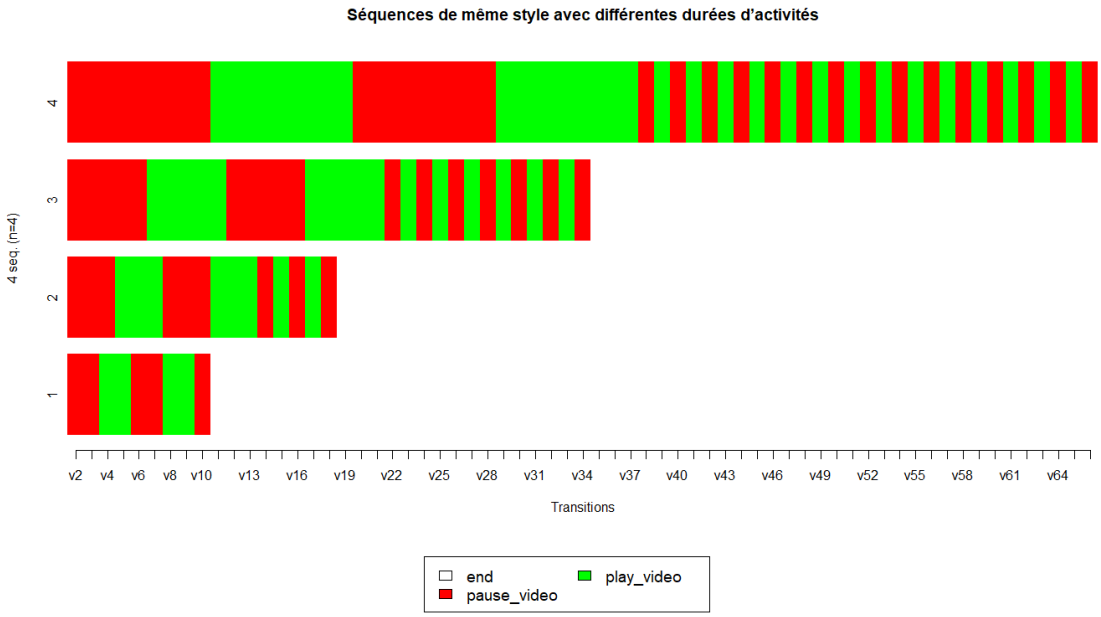
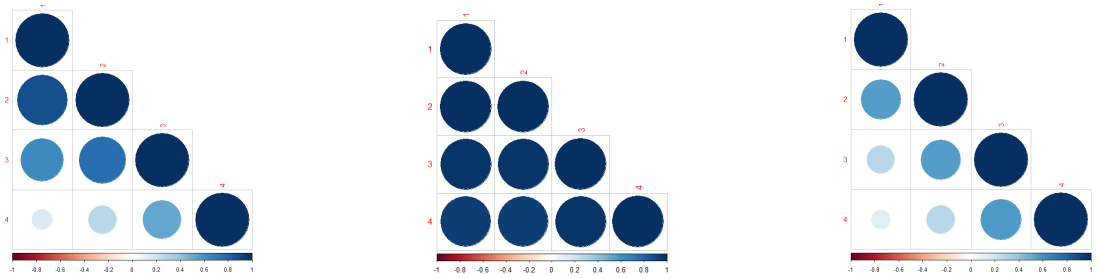


Figure 6.5 Même séquence d'états avec des longueurs différentes.

L'algorithme sélectionne un TMED de chaque étudiant pour le test parmi neuf (9) autres TMED et s'entraîne sur les huit (8) restantes. En général, pour l'identification, 80 % des données sont utilisées pour l'entraînement et 20 % pour le test. Ce qui fait que, chaque étudiant aura au moins une représentation TMED dans l'ensemble test. L'expérience est dans chaque cas répétée 400 fois en utilisant différents groupes d'étudiants à identifier (un nombre de 3 à 15 étudiants à chaque cycle d'identification) et les moyennes des résultats sont rapportées. L'ensemble d'entraînement a ainsi pour chaque étudiant au plus huit (8) de ses TMED et au moins un se trouve dans l'ensemble test à chaque cycle d'identification.

Chaque classificateur (SVM, GBM et KNN) est ensuite entraîné pour chaque étudiant sur l'ensemble d'entraînement. À cette fin, le test devrait identifier, à chaque cycle, jusqu'à quinze (15) étudiants en même temps pour voir dans quelle mesure on peut distinguer l'interaction d'un étudiant en particulier de celle des autres étudiants en utilisant la représentation proposée.

Dans la deuxième partie de la validation, nous faisons une expérimentation similaire à la première expérience mais cette fois-ci pour tester si la représentation TMED peut identifier mieux que les autres quelle est la vidéo avec laquelle interagit un étudiant quelconque. En



(a) ED (b) Markov (c) TMED

Figure 6.6 (a) la similarité basée sur la séquence d'Édition montre une certaine progression mais très déséquilibrée par rapport à ce qui est attendu lorsqu'on observe les séquences. Les degrés de similarité des diagonales ne sont pas les mêmes.(b) la similarité basée sur la chaîne de Markov ne peut pas reconnaître la durée dans chaque état car les probabilités de transition entre les états sont préservées sur la longueur de chaque séquence. Toutes les séquences sont considérées comme identiques, même si elles ne sont pas tout à fait identiques en fonction de la durée dans chaque état. (c) la similarité proposée par TMED peut reconnaître le fait que ces séquences soient identiques mais le niveau de similarité est basé sur le temps passé dans chaque état. Le résultat montre l'augmentation progressive du niveau de similarité. Les diagonales montrent le même degré de similarité, c'est ce qui était attendu vue la structure des interactions.

d'autres termes, nous voulons vérifier l'hypothèse du chapitre 5 selon laquelle une vidéo impose aux étudiants une manière d'interagir avec elle. Nous avons déjà vérifié cette hypothèse avec la représentation SIVS et nous voulons savoir si la représentation TMED peut mieux que d'autres représentations semblables (séquence et chaîne de Markov) corroborer cette hypothèse. Les détails de la procédure de ces deux expérimentations sont présentés dans les sections suivantes.

Ces deux expérimentations de validation de la représentation TMED sont reproduites en utilisant la représentation en chaîne de Markov et la représentation séquentielle pour comparer leurs résultats respectifs. Nous pourrons ainsi voir laquelle des trois représentations est la meilleure représentation dans la discrimination des étudiants et des vidéos.

Les deux expériences se feront sur un ensemble de données réelles d'interaction vidéo pour tester et comparer la capacité de la méthode proposée pour reconnaître (1) l'étudiant derrière une séquence d'interaction, et (2) la vidéo derrière une séquence d'interaction. Bien que cette tâche ne soit d'aucune utilité pratique, puisque la vidéo et l'étudiant associé à une séquence

d'interaction sont déjà connus en général, elle fournit un ensemble de données de vérité de terrain pour évaluer le pouvoir de discrimination de chaque approche.

6.5.3 Identification des séquences d'interaction des étudiants

Les données contiennent les "logs" (les données de connexion) de 4 800 d'étudiants, où chaque étudiant interagit avec neuf (9) vidéos différentes. Ces neuf (9) vidéos correspondent à toutes les vidéos de la première semaine du cours. Tous les 4 800 étudiants ont été sélectionné car ayant interagi avec chacune des neuf (9) vidéos de la première semaine du cours. Chaque interaction vidéo de l'étudiant est une séquence d'interactions dans un premier ensemble de données, une chaîne de Markov dans un second et dans une troisième, sous forme d'une matrice TMED.

Pour l'étudiant cible, les données de huit (8) autres vidéos avec lesquelles les étudiants ont interagi avec, servent de données d'entraînement. Les données d'interaction avec la neuvième vidéo serviront pour la prédiction de l'étudiant. Pour cela, 5 étudiants puis 12 étudiants sur une vidéo non vue doivent être classés comme cible et, bien sûr, un seul enregistrement provient de chaque étudiant cible.

Nous choisissons trois classificateurs bien connus, tels la machine à vecteur de support (SVM), l'arbre de décision (GBM) et le voisin le plus proche (KNN) pour chaque mode de représentation de séquence d'interactions afin de pouvoir identifier l'étudiant derrière la séquence d'interaction. Si une représentation spécifique de la séquence d'interaction de l'étudiant est prévisible en termes d'étudiants qui interagissent avec la vidéo, cela signifie que la représentation est capable de mieux distinguer les différents types d'interactions.

Cette expérience où nous prévoyons l'étudiant auquel appartient la représentation de la séquence, l'algorithme sélectionne une séquence de chaque étudiant pour prédire parmi les neuf (9) mêmes représentations de séquences d'étudiants et s'entraîner sur les huit (8) autres représentations de séquences de l'étudiant. La dimension de la représentation TMED et de chaîne de Markov de chaque séquence est de 25, qui représentent les 25 éléments de la matrice de chaque représentation de séquence comme on l'a décrit dans la section 6.3.1.

6.5.4 Identification des vidéos à partir des interactions des étudiants

Dans la deuxième partie de l'expérience, nous avons utilisé les mêmes représentations des séquences d'étudiants mais au lieu de prédire l'étudiant, on a prédit la vidéo avec laquelle l'étudiant a interagi. Nous avons utilisé les mêmes ensembles d'entrainement (80 % des données) et de test (20 % des données) en nous assurant que dans les données nous avons le même nombre d'étudiants qui ont interagi avec chaque vidéo. Comme chaque étudiant a neuf (9) séquences de représentation d'interaction, le nombre de classes prédites (vidéo 1 à vidéo 9) dans chaque ensemble de données considéré. Ainsi, le nombre des classes des vidéos (neuf) est le même quelque soit le nombre des étudiants considérés. La précision des prédictions aléatoires attendu est alors de $1/9 = 0.11$. Dans ce cas également, à chaque cycle de prédiction, l'algorithme s'assure que chaque représentation de séquence d'étudiant présente dans l'ensemble de données considérées à au moins une de ses représentations de séquence dans l'ensemble test à chaque cycle.

Le même schéma utilisé pour la tâche d'identification des étudiants est utilisé aussi pour la vidéo cible. Les données d'une vidéo non vue sont données en entrée à un classificateur et la tâche consiste à identifier à quelle vidéo correspond l'interaction, sur un total de neuf (9) vidéos à chaque fois. L'entrainement est basé sur les données d'interaction des autres étudiants.

La Figure 6.7, montre comment la méthode de similarité utilisant la représentation TMED a des meilleures précisions aussi bien pour la prédiction des vidéos avec laquelle les étudiants interagissent (première ligne de la figure) que de l'identification des étudiants derrière des interactions (seconde ligne dans la figure). Nous varions nos prédictions qui prennent en compte de trois (3) à quinze (15) étudiants pour vérifier cette capacité dans l'utilisation de la représentation TMED.

6.5.5 Application de la méthodologie de similarité dans l'analyse textuelle

Nous avons choisi également d'appliquer la méthodologie de similarité proposée pour l'analyse de texte. Le but de la recherche était de trouver les points de similarité entre les récits de 15 migrants africains et du Moyen-Orient, arrivés en Europe clandestinement, pour déterminer un récit atypique qui est analysé par un sociologue, en vue de comprendre les raisons

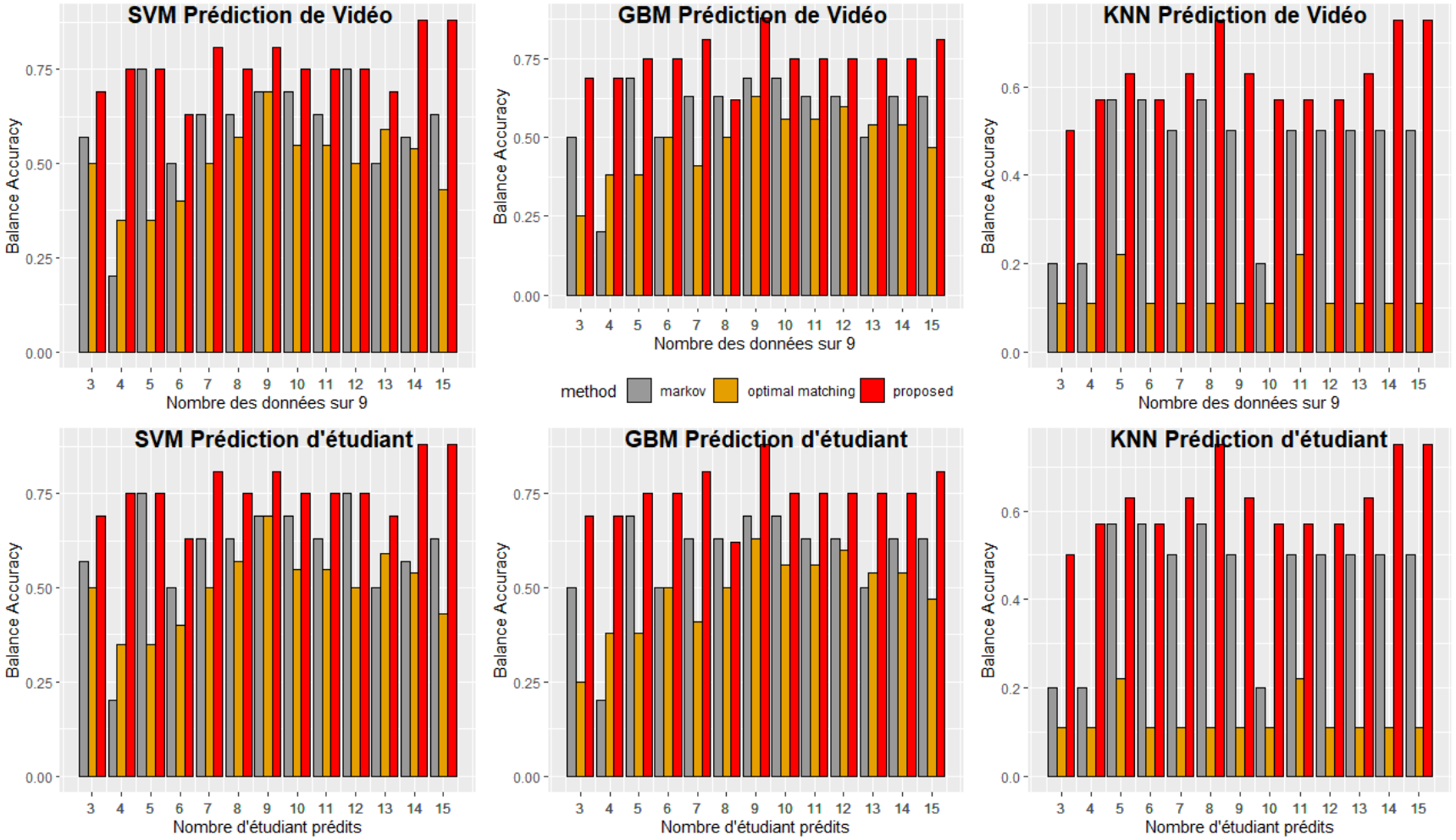


Figure 6.7 Résultats de la comparaison entre la méthode de représentation proposée et d'autres méthodes de prédiction de la vidéo (première ligne) et de l'étudiant (deuxième ligne) à partir des interactions des étudiants. markov= représentation en chaîne de Markov, optimal matching = représentation séquentielle et proposed = représentation TMED.

Prédictions :	45 séquences, 5 étudiants ciblés								
Approches :	SVM			GBM			KNN		
Méthode :	ED	MC	TMED	ED	MC	TMED	ED	MC	TMED
Balanced Acc.	0.75	0.38	**0.88**	0.63	0.38	**1.00**	0.50	0.38	**1.00**
Sensibilité	0.60	0.00	**0.80**	0.40	0.00	**1.00**	0.20	0.22	**1.00**
Spécificité	0.90	0.75	**0.95**	0.85	0.75	**1.00**	0.80	0.81	**1.00**
F_1	0.75	0.00	**0.89**	0.57	0.00	**1.00**	0.33	0.36	**1.00**
Prédictions :	108 séquences, 12 étudiants ciblés								
Balanced Acc.	0.77	0.50	**0.86**	**0.68**	0.60	0.67	0.56	0.45	**0.67**
Sensibilité	0.58	0.20	**0.74**	**0.42**	0.33	0.40	0.11	0.00	**0.40**
Spécificité	0.91	0.84	**0.97**	**0.95**	0.86	0.94	0.56	0.90	**0.93**
F_1	0.73	0.20	**0.78**	0.59	0.50	**0.63**	0.20	0.00	**0.67**

Tableau 6.2 Résultats de la validation croisée vingt fois 400 séries de prédictions d'étudiants de 5 et 12 étudiants utilisant trois méthodes différentes de représentation de l'interaction des étudiants avec des vidéos montrant que la technique de représentation proposée est plus performante que les autres. ED (Distance d'édition), MC (Chaîne de Markov, distance de Frobenius), TMED (Combinaison, distance de Frobenius). Les valeurs de F_1 sont des valeurs moyennes des F_1 de toutes les classes (voir section 6.2.4).

profondes du départ des migrants de leurs pays d'origine. En effet, il est question pour des études sociologiques sur le phénomène migratoire de donner la parole aux premiers concernés au lieu de faire des études sociologiques sur d'autres aspects du phénomène migratoire sans impliquer les migrants directement. Il s'agissait donc de transcrire 15 vidéos différentes de 15 migrants d'horizon diverses venus de divers pays d'Afrique et du Moyen-Orient pour situer la similarité dans leur discours. De cette similarité de discours, trouver un discours atypique de leur cheminement migratoire en trois étapes : avant le départ, le parcours migratoire et l'arrivée en Europe.

La représentation de chaque discours se fait en constituant un corpus des mots du domaine en prenant tous les mots et expressions utilisés par au moins 3 sur 15 migrants. Du corpus des mots fréquents, sont exclus tous les mots de connexion : les articles, les conjonctions de coordination, les ponctuations.

Chaque intervention (témoignage) est représentée sous forme d'un vecteur comptant le nombre des mots utilisés par chaque migrant. L'expression lexicale de ces migrants semble être faible sur le plan du vocabulaire : on observe seulement 606 mots distincts qui sont utilisés communément par au moins 3 des 15 migrants. Le transcrit de chaque intervenant est donc

Prédictions :	45 séquences, 9 vidéos ciblées								
Approches :	SVM			GBM			KNN		
Méthode :	ED	MC	TMED	ED	MC	TMED	ED	MC	TMED
Balanced Acc.	0.50	0.62	**0.75**	0.63	**0.75**	0.56	0.46	0.57	**0.63**
Sensibilité	0.11	0.33	**0.56**	0.33	0.22	**0.56**	0.22	0.22	**0.33**
Spécificité	0.89	0.60	**0.95**	0.71	0.90	**0.95**	0.90	0.99	**0.92**
F_1	0.20	0.50	**0.72**	0.50	0.36	**0.72**	0.36	0.36	**0.50**
Prédictions :	108 séquences, 9 vidéos ciblées								
Balanced Acc.	0.56	0.50	**0.75**	0.50	0.63	**0.75**	**0.57**	0.50	**0.57**
Sensibilité	0.22	0.11	**0.56**	0.11	0.33	**0.56**	0.22	0.11	**0.22**
Spécificité	**0.90**	0.89	0.83	**0.89**	0.85	0.83	0.85	**0.89**	0.85
F_1	0.36	0.20	**0.61**	0.20	0.50	**0.61**	**0.36**	0.20	**0.36**

Tableau 6.3 Résultats de la validation croisée de 400 séries de vidéos de prédiction des vidéos de 45 et 108 enregistrements d'interactions d'étudiants en utilisant trois méthodes différentes de représentation de l'interaction des étudiants avec les vidéos, ED (Distance d'édition), MC (Chaîne de Markov), TMED (Combinaison). Les valeurs de F_1 sont des valeurs moyennes des F_1 de toutes les classes (voir section 6.2.4).

représenté par une ligne de 606 colonnes représentant les 606 mots possibles que l'intervenant a pu utiliser et le nombre de fois dans son discours. Nous obtenons ainsi une matrice de 15 par 606 représentants tous les transcrits des 15 intervenants.

De cette matrice obtenue, la valeur d'une cellule est un (1) lorsqu'un mot est présent une fois dans le discours du migrant et sinon zéro (0). Ayant ainsi l'expression vectorielle de chaque intervention dans l'espace du corpus des mots communs, on applique la méthodologie de la similarité proposée pour vérifier le degré de similarité des discours des migrants que nous pouvons voir de façon numérique dans la figure 6.8 ou de façon visuelle dans la figure 6.9.

Pour le calcul de la similitude entre les discours, nous avons utilisé le principe de la matrice TMED. Avec le vecteur du nombre de parution des mots nous avons considéré comme la matrice de transition dans le cas des TMED. Comme les discours n'avaient pas la même longueur des mots, nous avons normalisé la représentation comme nous le faisons pour la matrice TMED en divisant le nombre d'apparition des mots dans le texte par le nombre total des mots du corpus présents dans le texte de chaque étudiant de telle sorte que la somme de tous les éléments de la représentation donne un. C'est le cas avec la matrice TMED. Et nous pouvons donc calculer les similarités S_{mat} de TMED.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]
[1,]	1.00	0.68	0.91	0.59	0.82	0.82	0.73	0.59	0.41	0.32	0.55	0.45	0.73	0.68	0.73
[2,]	0.68	1.00	0.68	0.64	0.59	0.68	0.68	0.36	0.45	0.55	0.59	0.59	0.59	0.55	0.59
[3,]	0.91	0.68	1.00	0.59	0.73	0.73	0.64	0.50	0.41	0.32	0.55	0.45	0.73	0.68	0.64
[4,]	0.59	0.64	0.59	1.00	0.77	0.59	0.59	0.55	0.45	0.36	0.59	0.41	0.59	0.55	0.59
[5,]	0.82	0.59	0.73	0.77	1.00	0.82	0.73	0.59	0.32	0.32	0.55	0.45	0.73	0.59	0.73
[6,]	0.82	0.68	0.73	0.59	0.82	1.00	0.82	0.59	0.41	0.41	0.64	0.64	0.73	0.68	0.73
[7,]	0.73	0.68	0.64	0.59	0.73	0.82	1.00	0.59	0.50	0.50	0.45	0.64	0.73	0.59	0.73
[8,]	0.59	0.36	0.50	0.55	0.59	0.59	0.59	1.00	0.45	0.36	0.50	0.50	0.50	0.55	0.41
[9,]	0.41	0.45	0.41	0.45	0.32	0.41	0.50	0.45	1.00	0.45	0.41	0.50	0.50	0.36	0.50
[10,]	0.32	0.55	0.32	0.36	0.32	0.41	0.50	0.36	0.45	1.00	0.50	0.59	0.32	0.36	0.41
[11,]	0.55	0.59	0.55	0.59	0.55	0.64	0.45	0.50	0.41	0.50	1.00	0.45	0.45	0.77	0.45
[12,]	0.45	0.59	0.45	0.41	0.45	0.64	0.64	0.50	0.50	0.59	0.45	1.00	0.55	0.50	0.64
[13,]	0.73	0.59	0.73	0.59	0.73	0.73	0.73	0.50	0.50	0.32	0.45	0.55	1.00	0.59	0.64
[14,]	0.68	0.55	0.68	0.55	0.59	0.68	0.59	0.55	0.36	0.36	0.77	0.50	0.59	1.00	0.68
[15,]	0.73	0.59	0.64	0.59	0.73	0.73	0.73	0.41	0.50	0.41	0.45	0.64	0.64	0.68	1.00

Figure 6.8 La similarité entre les discours exprimée en terme de pourcentage de ressemblance. Par exemple 0.55 signifie 55% de ressemblance et 1.0 signifie 100% de ressemblance.

De la matrice de représentation des chaque discours (matrice 15 par 606), on peut donc extraire les mots les plus utilisés par les migrants dans leur discours. Il s'agit en l'occurrence de calculer le nombre total des migrants qui ont utilisé le mot dans leur discours : ici la somme de chaque colonne. On extrait ainsi les mots les plus fréquents. Sont considérés comme mots les plus fréquents, ceux qui apparaissent dans le discours d'au moins la moitié des migrants : les mots utilisés par au moins 7 migrants (47% des migrants). A partir de ces mots communs aux migrants, l'on est capable de rejoindre toutes les phrases des migrants les ayant employés et de diviser les récits en trois parties : avant le départ, pendant le voyage et à l'arrivée en Italie (les interviews des migrants se sont passés en Italie). Nous obtenons alors un discours atypique provenant des mots communs à tous les migrants en trois parties.

6.6 Résultats

Avec les résultats montrant que la méthode utilise la représentation TMED proposée, pour le niveau de similitude, on peut dire à quel point deux séquences arbitraires sont similaires. Grâce aux tests de validation sur des données prototypiques, la méthode proposée donne de meilleurs résultats que les deux autres modes de représentation existants comme on peut le voir à travers les figures 6.4 et 6.6. Pour la même séquence représentée comme une séquence cyclique d'interaction avec différents modes de représentation, la figure 6.3 indique le degré de similarité attendu avoisinant les 100 %, mais la représentation proposée et la chaîne de Markov nous donnent les résultats les plus proches de ceux attendus, comme le montre la figure 6.4.

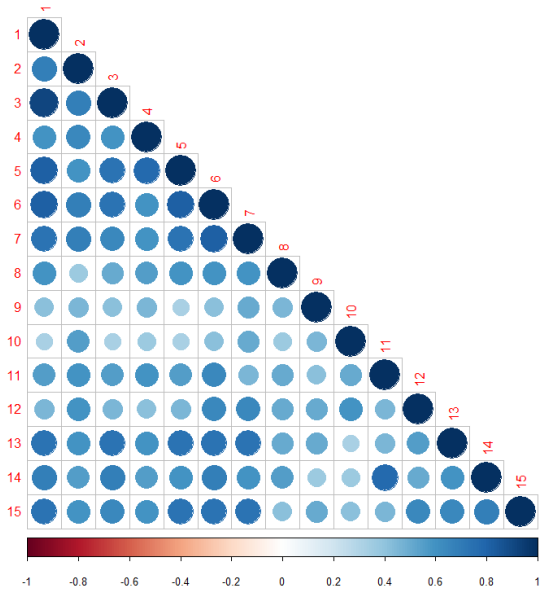


Figure 6.9 La similarité entre les discours de manière visuelle.

Lorsque nous considérons une même séquence d'états avec des durées différentes, comme le montre la figure 6.5, le résultat attendu de la similarité est une augmentation progressive du niveau de similarité en fonction de la longueur de la séquence. La représentation en chaîne de Markov n'a pas pu trouver que la longueur des séquences est différente alors que la méthode utilisant TMED est capable de bien le montrer (Figure 6.6). Le mode de représentation de séquence d'édition quant à elle n'arrive pas non plus à découvrir une similarité progressive et passe à côté du résultat attendu.

L'expérience sur les tâches de données réelles teste la capacité de chaque représentation de l'interaction vidéo à identifier chaque séquence d'interaction en termes de séquences d'étudiants et de vidéo. Les résultats montrent de meilleurs résultats pour TMED que les deux autres (tableau 8.2). Les paramètres de performance sur la prédiction des étudiants en utilisant SVM, GBM et KNN sur la prédiction de cinq (5) étudiants et douze (12) étudiants avec neuf (9) enregistrements de chaque étudiant (où huit (8) au plus enregistrements sont pour l'entrainement et la prédiction d'au moins un enregistrement de chaque étudiant). Les résultats montrent que l'utilisation du TMED proposée est plus performante que les deux autres. Les résultats détaillés des trois représentations sont donnés dans la figure 6.7. Ces résultats montrent que dans chaque cas la méthode proposée utilisant TMED est plus précise que celles utilisant les deux autres représentations (les deux des autres modes).

En prédisant la vidéo, les résultats complets pour quarante-cinq (45) enregistrements de cinq (5) étudiants différents et cent huit (108) enregistrements de douze (12) étudiants en prédisant les neuf (9) vidéos sont présentés dans le tableau 6.3. Nous démontrons que la méthode proposée est également meilleure pour reconnaître à la fois la vidéo et l'étudiant par rapport aux deux autres méthodes de représentation de l'interaction de l'étudiant avec la vidéo.

La comparaison complète des performances sur la prédiction de la vidéo (la première ligne de la figure) et des étudiants (la deuxième ligne de la figure) se trouve dans la figure 6.7. Dans les deux cas, prédisant les vidéos ou les étudiants, la représentation TMED proposée est la plus performante en termes de précision que les représentations existantes. Ces résultats suggèrent que la représentation proposée a une meilleure façon de représenter une interaction des étudiants avec les vidéos d'un côté, et l'autre, de pouvoir discriminer la particularité d'interaction avec une vidéo et peut également être utilisée pour trouver le degré de similarité entre deux interactions différentes.

6.7 Comparaison entre les représentations SIVS et TMED

De par sa structure, avec la représentation SIVS on ne peut pas savoir avec précision si l'on peut reconnaitre des vidéos dans la représentation d'une interaction vidéo à moins d'avoir un ensemble des vidéos de même durée à la seconde près. C'est ce que nous avions fait dans le cadre de la validation en définissant des centroïdes entre les vidéos de même durée. En effet, la taille de la matrice dans la représentation SIVS dépend de la taille de la vidéo à la seconde près. Ainsi cette taille de la matrice peut introduire un biais dans la reconnaissance de la vidéo lorsqu'on a des vidéos de durée différentes à reconnaître dans les interactions étudiantes. Ce biais n'existe pas par contre dans la représentation TMED car les dimensions des matrices TMED sont les mêmes pour toutes les vidéos quelque soit leur durée.

Par contre, de par la structure de la représentation SIVS, l'on peut directement voir les parties de la vidéo écoutées une ou plusieurs fois, pas du tout écoutées ou sautées d'un seul coup d'œil. Elle représente une sorte de tableau de la navigation de l'étudiant dans la vidéo alors que ce n'est pas le cas avec la représentation TMED. Avec la représentation TMED nous ne pouvons pas avoir accès aux détails de navigation de l'étudiant dans la vidéo.

Donc les deux représentations SIVS et TMED ont leurs avantages et leurs limites et peuvent

être utilisées en fonction de ce l'on souhaite analyser.

6.8 Conclusion

La représentation proposée vise à combler une lacune méthodologique sur la représentation et la comparaison des séquences d'interactions vidéo. Elle surmonte les inconvénients des représentations précédentes basées sur la chaîne de Markov et les séquences d'interactions connues sous le nom de "Edit Distance based" (ED). La principale contribution de cette représentation est le fait qu'elle prend en compte le temps passé dans chaque état et le style général de succession des états. Elle offre un nouvel outil aux chercheurs qui souhaitent comparer les interactions des apprenants avec les vidéos entre elles et trouver le degré de ressemblance entre ces interactions vidéo pour notamment pouvoir regrouper des écoutes semblables.

L'utilisation de la méthodologie de similarité proposée dans le cadre des textes montre la capacité de cette méthodologie à pouvoir être utilisable dans d'autres types de données. Dans notre cas, au lieu de l'utiliser simplement dans le cadre des données d'interactions vidéo, nous étions capables de l'utiliser pour repérer la similarité entre des textes à la place de l'utilisation des matrices termes-document.

La représentation proposée est également en mesure de mieux spécifier une séquence d'interaction lors de tâches de classification, comme le montrent les résultats. TMED a une meilleure performance dans identification de l'étudiant et de la vidéo que les deux autres représentations d'interactions vidéo.

La représentation TMED apporte un aspect complémentaire à la représentation SIVS présentée au chapitre 5 dans la mesure où elle peut reconnaître dans la représentation des interactions, les vidéos et les étudiants quelque soit la durée de la vidéo. Cela ne pouvait être possible dans le cas de la représentation SIVS que pour des vidéos de même durée à la seconde près.

La méthodologie de similarité proposée ouvre des nouvelles voies de recherche. Elle pourrait permettre de déterminer de regrouper les interactions vidéo de chaque apprenant en fonction du niveau de similarité des séquences. On peut également utiliser cette méthode pour savoir si chaque style vidéo impose un style d'interaction spécifique aux apprenants. Une autre piste est le lien que l'on peut trouver entre le style vidéo d'interaction et la performance

des étudiants. Les différents styles d'interaction vidéo peuvent-ils induire l'échec ou faciliter la réussite dans un cours en ligne ? Nos futures investigations répondront à certaines de ces questions.

CHAPITRE 7 PRÉDICTION DE L'ÉCHEC ET DU SUCCÈS DES ÉTUDIANTS BASÉE SUR LES MESURES AGGLOMÉRATIVES D'UTILISATION DES VIDÉOS

7.1 Introduction

Dans ce chapitre nous présentons notre quatrième travail (*"Early prediction of success in MOOC from video interaction features."*, In Proceedings of the 21th International Conference on Artificial Intelligence in Education, AIED 2020) qui correspond à la quatrième contribution de cette thèse. L'analyse de l'interaction des étudiants avec les vidéos peut-elle prédire leur succès ou leur échec ? Avec des milliers d'étudiants qui s'inscrivent à des cours en ligne chaque année, il est très intéressant pour les développeurs et les instructeurs de MOOC de prédire, dès les premières semaines du cours, les étudiants qui pourraient abandonner ou qui sont plutôt susceptibles de terminer le cours avec succès. Par exemple, les indicateurs agglomératifs pourraient identifier le nombre des étudiants qui pourraient rester jusqu'à la fin du cours et le pourcentage d'étudiants qui finalement réussiront ou échoueront le cours. Il serait intéressant d'identifier les étudiants qui pourraient réussir le cours en recevant une aide supplémentaire. En concentrant l'aide aux étudiants qui ont une interaction pouvant conduire à l'échec dès la fin de la première semaine du cours, il y a des fortes chances de pouvoir aider à prévenir un échec à la fin du cours pour une partie d'entre eux.

Il est probable que certaines raisons d'échec et d'abandon des cours pourraient également dépendre d'autres facteurs qui pourraient être mis en lumière lors d'une interview des acteurs de l'apprentissage (Guo, Kim et Rubin 2014). Pourtant, en utilisant des données directement issues des comportements des étudiants avec les MOOC, nous pensons pouvoir prédire les succès des étudiants avec objectivité tout en aidant l'étudiant signalé comme pouvant échouer à éviter l'échec à la fin du cours.

Il est très important de s'intéresser aux taux de réussite des étudiants dans les MOOC car les recherches indiquent que les étudiants qui cherchent à apprendre par le biais des systèmes d'apprentissage en ligne ont un taux d'abandon significatif après la première semaine (Hill 2013). Ainsi, les interactions pendant la première semaine de cours sont un point de référence clé pour prédire les succès des étudiants tout au long du cours, si l'étudiant persiste jusqu'au bout. Mais il ne s'agit pas seulement de prédire les succès des étudiants dans les MOOC. L'objectif premier est plutôt de détecter rapidement les étudiants qui sont susceptibles échouer

afin de leur offrir un soutien supplémentaire pour qu'ils puissent persister et améliorer leur façon d'étudier. Il s'agit également d'informer les développeurs et les instructeurs des MOOC des étudiants qui sont susceptibles de persister et de réussir. Plutôt que de se concentrer sur un grand nombre d'inscriptions au détriment des succès des étudiants, il s'agit de s'adapter au fur et à mesure aux besoins du plus grand nombre des apprenants. Par exemple, de nombreuses études sur la prédiction des succès dans les MOOC ont généralement été basées sur des activités qui ont contribué directement aux résultats scolaires de l'étudiant, telles que des quiz, des soumissions de problèmes, et l'interaction avec des forums en ligne (S. Jiang et al. 2014 ; L.-Y. Li et Tsai 2017 ; O. H. Lu et al. 2018).

Le but de ce chapitre est d'étudier les mesures agglomératives dans l'utilisation de la vidéo qui peuvent déterminer si un étudiant risque d'échouer ou de réussir à un stade précoce du MOOC. Il s'agira de voir si, à partir de certaines mesures agglomératives d'écoute, l'on peut efficacement prédire le succès après une semaine d'interactions vidéo, le succès à la fin du cours des étudiants. Nous allons comparer notre approche à celle utilisée par He, Zheng et al. 2018. Il s'agit donc de proposer une nouvelle règle de regroupement dans deux groupes des étudiants (ceux qui vont échouer et ceux qui vont réussir le cours) pour mieux séparer les étudiants en fonction de leur succès ou non finale dans le cours très tôt dans le cours (précoce). Il s'agira donc d'identifier très tôt le groupe d'étudiants qui vont échouer dans le cours pour permettre aux instructeurs de prendre des mesures pour aider ce groupe d'étudiant. C'est ainsi que nous allons classifier en premier les étudiants après la première semaine d'interaction avec les vidéos sur un cours d'une durée de treize (13) semaines. Nous allons également faire un classement après six (6) semaines de cours (mi-parcours) pour vérifier plus attentivement la consistance de la méthode de classification.

L'utilisation d'une telle technique pour prédire les échecs des étudiants pourrait se justifier dans des contextes où l'accès aux données étudiantes est réduit aux interactions vidéo ou dans une recherche rapide de prédiction des échecs. Elle pourrait également être un module de prédiction parmi d'autres dans le cadre d'un système plus robuste de prédiction des échecs qui pourrait alors prendre la majorité des résultats qui convergent venant de plusieurs systèmes de prédiction des échecs.

Pour mieux comparer nos résultats à ceux obtenus par les mesures agglomératives utilisées par He, Zheng et al. 2018, seront utilisées les données de la première semaine du cours (9 premiers vidéos) et, à mis-parcourt du cours après six semaines (79 vidéos) seront utilisées.

Cela permet de reproduire la recherche de He, Zheng et al. 2018 sur nos données et de pouvoir la comparer aux résultats que nous obtenons à en utilisant les mesures que nous proposerons.

7.2 Les mesures agglomératives d'utilisation des vidéos

Nous décrivons dans cette section trois mesures agglomératives qui constituent des paramètres d'écoute vidéo. Nous décrivons ensuite comment elles sont utilisées pour définir les groupes d'étudiants en vue de la prédiction de leur succès. Les mesures agglomératives tel que AR et UR sont inspirées de la publication de He, Zheng et al. 2018.

7.2.1 Taux d'assiduité (*"Attendance Rate"*)

Le taux d'assiduité $AR_{s,c}$ d'un étudiant s sur une semaine donnée c depuis le début du cours, est le nombre de vidéos que l'étudiant a visionnées totalement ou en partie par rapport au nombre total de vidéos disponibles jusqu'à cette période du calendrier des cours.

$$AR_{s,c} = \frac{|W_{s,c}|}{|V_c|} \tag{7.1}$$

Où :
$AR_{s,c}$ est le taux d'assiduité de l'étudiant s jusqu'à la fin de la semaine c.
V_c est le nombre total des vidéos du cours diffusées jusqu'à la fin de la semaine c.
$W_{s,c}$ est la collection des différentes vidéos regardées totalement ou en partie par l'étudiant s jusqu'à la fin de la semaine c.

Si un étudiant a visionné la même vidéo plusieurs fois, cette vidéo est comptée comme une seule vidéo. En d'autres termes, $W_{s,c}$ est le nombre de vidéos uniques visionnées par l'étudiant pendant cette période (du début du cours jusqu'à la fin de la semaine c). La valeur maximale possible de AR est donc de 1, lorsqu'un étudiant a visionné en partie ou en totalité toutes les vidéos diffusées pour la période allant du début du cours jusqu'à la fin de la semaine c.

7.2.2 Taux d'utilisation (*"Utilization Rate"*)

Le taux d'utilisation $UR_{s,c}$ d'un étudiant s jusqu'à la fin de la semaine c depuis le début du cours est la proportion de l'activité de lecture vidéo de l'étudiant (la somme de tous les

temps d'écoute des vidéos) par rapport à la somme de la durée totale de toutes les vidéos jusqu'à la fin de la semaine c.

$$UR_{s,c} = \frac{\sum_{i=1}^{n} W_{ts,i}}{\sum_{j=1}^{N} t(V_j)} \tag{7.2}$$

Où :
$UR_{s,c}$ est le taux d'utilisation de l'étudiant s à la fin de la semaine c.
$t(V_j)$ est la durée de la vidéo j.
N est le nombre total de vidéos diffusées jusqu'à la fin de la semaine c.
$W_{ts,i}$ est la durée totale des sections de la vidéo i jouée par l'étudiant s.
n est le nombre de vidéos uniques que l'étudiant s a joué en totalité ou en partie jusqu'à la fin de la semaine c.

Comme un étudiant peut faire jouer des segments d'une même vidéo plusieurs fois, $Wt_{s,i}$ peut être supérieur à Vt_i, et donc $UR_{s,c}$ peut être supérieur à 1 lorsque la durée totale de lecture de toutes les vidéos par un étudiant est supérieure à la somme totale de la durée des vidéos diffusées jusqu'à cette période. Lorsque $Wt_{s,i}$ est supérieur à 1, il peut être interprété de deux façons : pour certains, c'est un signe de difficultés avec les vidéos (N. Li, Kidzinski et al. 2015 ; N. Li, Kidziński et al. 2015) mais pour d'autre c'est un signe d'engagement de l'étudiant (Draus, Curran et Trempus 2014 ; Cummins, Beresford et Rice 2015). Dans ce cas précis où nous sommes dans un contexte de la libre utilisation des vidéos, la persistance à écouter plusieurs fois certaines sections ou des vidéos entières montre une certaine persistance de l'étudiant. En ce sens, elle peut indiquer plus d'engagement de l'étudiant dans les écoutes vidéo.

7.2.3 Taux de visionnage (*"Watch Ratio"*)

Le taux de visionnage $WR_{s,c}$ d'un étudiant s à la fin de la semaine c est défini comme suit :

$$WR_{s,c} = \frac{UR_{s,c}}{AR_{s,c}} \tag{7.3}$$

Où :
$WR_{s,c}$ est le taux de visionnage de l'étudiant s jusqu'à la fin de la semaine c.
$UR_{s,c}$ est le taux d'utilisation de l'étudiant s à la fin de la semaine c.
$AR_{s,c}$ est le taux d'assiduité de l'étudiant s jusqu'à la fin de la semaine c.

Le $WR_{s,c}$ représente la façon dont un étudiant regarde la vidéo depuis le début d'écoute. Par exemple, si $WR_{s,c} = 1$ cela signifie que l'étudiant s regarde complètement la vidéo du début à la fin. Le fait que $AR_{s,c}$ peut être égal à zéro, le $WR_{s,c}$ ne s'applique qu'aux étudiants ayant un $AR_{s,c} > 0$.

7.2.4 L'index de visionnage (*"Watch Index"*)

Nous introduisons un nouvel index, $WI_{s,c}$. L'index de visionnage d'un étudiant s à la fin de la semaine c :

$$WI_{s,c} = UR_{s,c} \times AR_{s,c} \tag{7.4}$$

Notez que $WI_{s,c}$ peut être supérieur à 1 si le temps total de visionnage est supérieur à la somme des durées des vidéos (UR$\geq$ 1) et si, en même temps, l'étudiant a interagi avec toutes les vidéos en ligne pour cette période (AR =1). Pour que cette situation se produise, l'étudiant écoute plusieurs fois certaines sections ou des vidéos entières. Cette mesure prend en compte à la fois le nombre des vidéos écoutées par l'étudiant et le pourcentage d'écoute par rapport à toutes les vidéos diffusées jusqu'à l'instant de la prise de la mesure. Plus cette mesure est grande, plus l'étudiant a fait jouer les vidéos. Cette mesure va prendre une place importante dans les distributions des groupes d'étudiants et dans la perspective de prédictions que nous proposons.

7.3 Méthodologie de classification

L'objectif général de cette étude est d'utiliser des mesures agglomératives d'interaction vidéo pour évaluer si un étudiant réussit ou échoue le cours. Nous souhaitons évaluer la précision de ces prédictions en fonction du calendrier du MOOC. Les prédictions précoces sont jugées moins précises mais plus utiles à des fins correctives.

7.3.1 Classification en utilisant le taux de visionnage (WR)

He, Zheng et al. 2018 en utilisant des métriques tel que UR, AR, WR sont capables d'évaluer la distribution des modèles vidéo affectant les succès des étudiants. Ils ont réalisé une étude sur la façon dont les différentes distributions de WR de tous les étudiants suivent les performances des étudiants sans les prévoir. Ils ont pu définir des groupes d'étudiants en fonction de la distribution de WR où graphiquement ils ont pu établir les valeurs de WR en fonction de trois (3) groupes :

$WR \leq 0.3, 0.3 < WR \leq 1.3$ *et* $WR > 1.3$

Dans ces groupes, ils ont pu distinguer en majorité les étudiants qui poursuivent leurs études et qui ont obtenu leur diplôme, de ceux qui ont reporté leurs études.

Nous allons reproduire leur méthodologie sur nos données. Dans notre cas, il s'agit de déterminer deux groupes : à savoir ceux qui vont passer le cours et ceux qui vont échouer ou abandonner. A cet effet, nous recourons à la courbe de probabilité cumulative d'écoute en fonction des valeurs de AR et UR (Figure 7.1). La courbe de UR nous montre que 92% des étudiants en moyenne écoutent moins de 40% du contenu des vidéos qu'ils regardent. La courbe de AR nous montre que 45% des étudiants seulement ont interagi systématiquement avec toutes les vidéos de la première semaine. Il faut noter que la particularité du choix des données des étudiants est basée sur le fait que les étudiants retenus pour cette étude sont ceux qui ont interagi avec les vidéos jusqu'à la fin du cours. Nous excluons ainsi ceux qui ont abandonné ou simplement arrêté d'interagir avec les vidéos durant le cours.

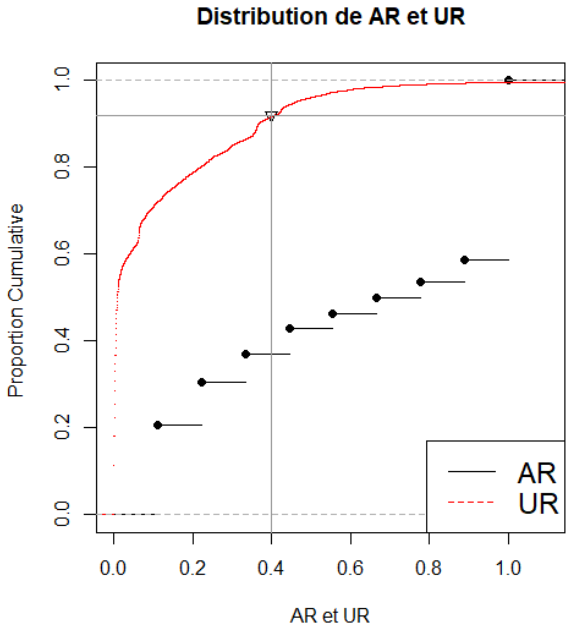


Figure 7.1 Distribution de AR et UR de la première semaine

Dans notre cas la majorité des étudiants ont regardé moins de 40% du contenu des vidéos après la première semaine. Comme le montre la figure 7.1 nous avons 92% des étudiants qui ont regardé moins de 40% du contenu des vidéos. Nous avons retenu les données d'interactions des étudiants ayant interagi avec toutes les vidéos de la première semaine soit 10 424 avec 970 ayant réussi le cours. La distribution descriptive des valeurs de AR, UR, et WR est donnée dans la table 7.1.

Indicateur	Médiane	Moyenne	Déviation Standard
AR	0.78 (0.34)	0.62 (0.37)	0.38 (0.24)
UR	0.01 (0.01)	0.11 (0.09)	0.22 (0.17)
WR	0.01 (0.04)	0.10 (0.15)	0.32 (0.23)

Tableau 7.1 Statistiques descriptives de AR, UR et WR après la première semaine et entre parenthèses à la moitié du cours.

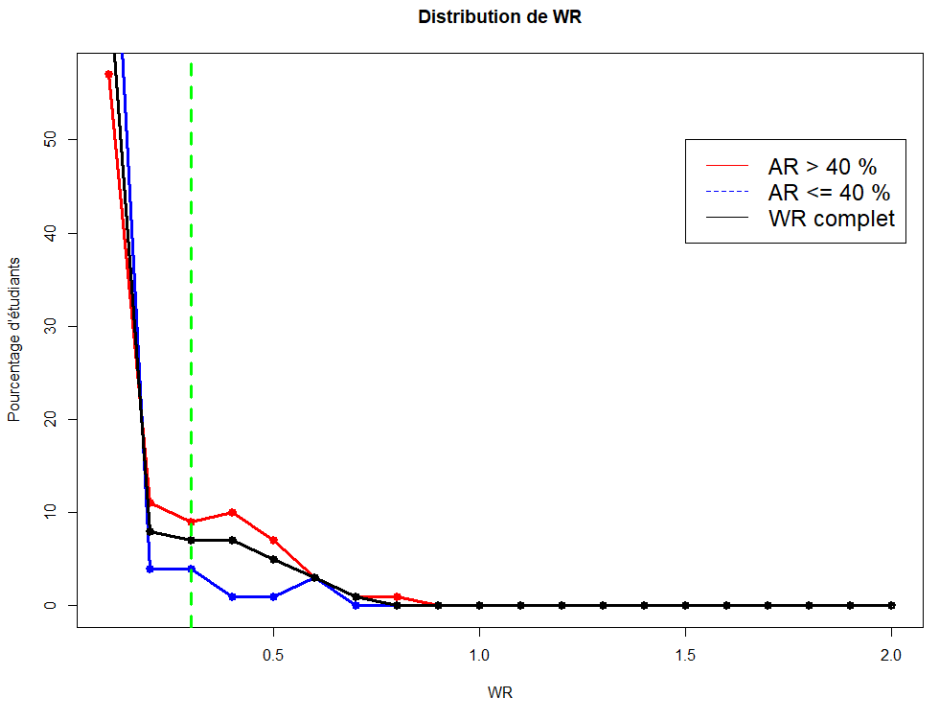


Figure 7.2 Distribution de WR

Dans notre cas la distribution de WR montre un seul sommet qui correspond à ce que les auteurs appellent *"Peak L"* qui nous donne la valeur de WR que nous appelons ici WR_l (voir figure 7.2 et la valeur de WR_l dans notre cas est de 0.3) qui va délimiter la séparation des

groupes selon la règle suivante utilisée par les auteurs pour déterminer le groupe de ceux qui vont échouer (GROUP I) et le groupe de ceux qui vont réussir (GROUP II) :

IF($WR_{s,c} < WR_l$)
THEN {Student s in GROUP I}
ELSE {Student s in GROUP II}

Où $WR_{s,c}$ est le taux de visionnage de l'étudiant s jusqu'à la fin de la semaine c.

Selon la méthodologie proposée par He, Zheng et al. 2018, la table 7.2 donne la répartition du nombre des étudiants qui ont réussis ou échoués dans les deux groupes après la première semaine et à la moitié du cours. Ces résultats vont être comparés à ceux que nous obtiendrons par la méthode proposée basée sur l'index de visionnage ci-dessous.

Tableau 7.2 Répartition par groupe du nombre des étudiants qui ont réussi ou échoué après la première semaine et à mi-parcours (sixième semaine), basée sur WR et telle que proposée par He, Zheng et al. 2018

	Après une semaine		Après six semaines	
	succès	échec	succès	échec
Groupe I	146	851	340	8 792
Groupe II	824	8 603	630	662
Total	970	9 454	970	9 454

7.3.2 Classification proposée des groupes

Les indicateurs des habitudes d'écoute de la vidéo décrits ci-dessus (AR, UR, WR, WI) sont destinés à servir de mesures agglomératives pour prédire la réussite des étudiants en les classant dans des groupes qui, comme nous le montrerons, ont des ratios de réussite nettement différents.

La règle de regroupement et d'évaluation est tirée de la classification d'un seul étudiant par rapport à la valeur moyenne du groupe d'étudiants. L'étudiant dont les résultats sont supérieurs à la valeur moyenne du groupe d'étudiants inscrits au cours a de fortes chances de réussir le cours. La règle de regroupement des étudiants (GROUP I : groupe de ceux qui vont échouer et GROUP II : le groupe de ceux qui vont réussir) en fonction de leurs mesures agglomératives est la suivante :

IF ($WI_{s,c} < \overline{WI}$ OR (étudiant s n'a pas soumis un devoir))
THEN {Student s in GROUP I}
ELSE {Student s in GROUP II}

Où $\overline{WI}$ est la valeur moyenne de WI de tous les étudiants qui interagissent avec les vidéos dans la période considérée. Pour cette règle, nous avons utilisé la valeur moyenne de WI des étudiants pour séparer les deux groupes d'étudiants.

Tableau 7.3 Répartition par groupe du nombre des étudiants qui ont réussi ou échoué après la première semaine et à mi-parcours (sixième semaine) basée sur WI.

	Après une semaine		Après six semaines	
	succès	échec	succès	échec
Groupe I	210	5 697	67	6 276
Groupe II	760	3 757	903	3 178
Total	970	9 454	970	9 454

7.3.3 Préparation des données

Nous avons créé une base de données qui comprenait : l'étudiant par son identifiant, le temps total de lecture de chaque vidéo en secondes et le nombre total de vidéos avec lesquelles chaque étudiant a interagi. Nous avons également les informations sur les problèmes qui ont été essayés par chaque étudiant.

La figure 7.3 indique la répartition des mesures agglomératives à la mi-temps du cours, ainsi que les notes finales de l'étudiant. Notons que trois des quatre histogrammes sont sur une échelle de log_{10}. Le fait que UR (taux d'utilisation) soit inférieur à AR (taux d'assiduité) montre que les étudiants n'ont pas regardé une grande partie du contenu de chaque vidéo et ont sauté une partie importante des vidéos. Mais le fait que AR soit élevé montre que l'étudiant a ouvert la majorité des vidéos. Par ailleurs, la valeur de UR peut montrer que certaines vidéos peuvent contenir des segments que les étudiants ne trouvent pas utiles et qu'ils ont sauté.

Dans le cadre de la classification des étudiants basée sur la valeur de WI, pour la première semaine du cours, les interactions avec les neuf (9) vidéos de la semaine ont été considérées. Pour chaque étudiant, la valeur de son WI ainsi que la moyenne de tous les W. La règle proposée

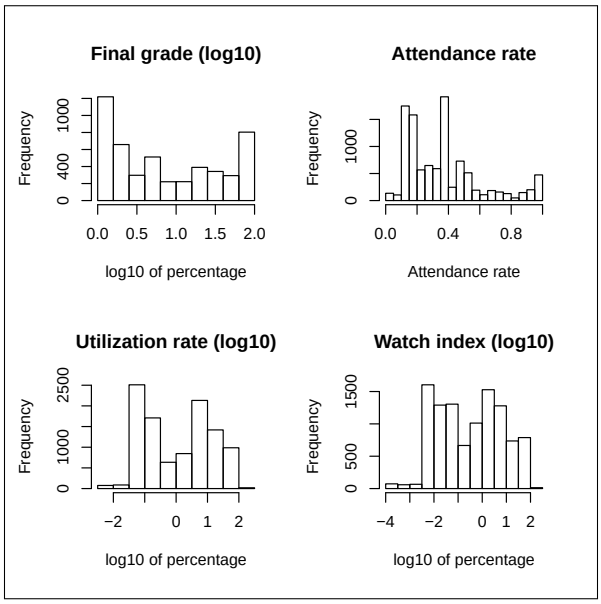


Figure 7.3 Distributions d'indicateurs à la moitié de la durée du cours.

pour la classification a été appliquée et la table 7.3 montre la distribution des étudiants selon cette règle. Un travail similaire a été fait après six semaines (mi-parcours du cours) prenant en compte les interactions avec 54 vidéos. Dans les deux cas les données de 10 424 étudiants ont été prisent en compte dont 970 ont réussi le cours en ayant une note finale de 50% et plus.

Tableau 7.4 Comparaison des résultats de He, Zheng et al. 2018 et de la méthode proposée (Proposition) présentée à des fins de comparaison.

	Après une semaine			
	succès		échec	
	He et al.	**Proposition**	**He et al.**	**Proposition**
Groupe I	146 (15%)	210 (22%)	851 (9%)	5 697 (60%)
Groupe II	824 (85%)	760 (78%)	8 603 (91%)	3 757 (40%)
Total	970 (100%)	970 (100%)	9 454 (100%)	9 454 (100%)
	Après six semaines			
Groupe I	440 (35%)	67 (7%)	8 792 (93%)	6 276 (67%)
Groupe II	630 (65%)	903 (93%)	662 (7%)	3 178 (33%)
Total	970 (100%)	970 (100%)	9 454 (100%)	9 454 (100%)

En supposant que le groupe II est celui des étudiants qui ont passé le cours et que le groupe I de ceux qui ont échoué, nous allons comparer nos résultats obtenus par la règle utilisant WI à ceux obtenus en reproduisant sur nos données la méthodologie proposée par He, Zheng

et al. 2018. Il s'agit donc de déterminer laquelle des deux classifications (table 7.2 ou table 7.3) sépare le mieux les étudiants qui vont réussir de ceux qui vont échouer. Pour cela nous allons calculer pour chacune des deux classifications la probabilité qu'un étudiant classé dans le groupe I échoue et la probabilité pour qu'un étudiant classé dans le groupe II réussisse le cours. Le tableau 7.5 suivant montre les probabilités qu'un étudiant soit classé dans le bon groupe en fonction de sa note finale sachant comme mentionné précédemment que le groupe I est celui des étudiants ayant échoués désigné par E et le groupe II des étudiants ayant réussi désigné par S.

Période	Probabilités	He et al.	Proposition
Première semaine	$P(E/G_1)$	0.90	0.95
	$P(S/G_2)$	0.11	0.18
Après 6 Semaines	$P(E/G_1)$	0.94	0.97
	$P(S/G_2)$	0.54	0.24

Tableau 7.5 Probabilités de classification dans chaque groupe en utilisant les deux méthodes. Ici $P(E/G_1)$= Probabilité d'échouer lorsqu'on est classé dans le groupe I et $P(S/G_2)$ = Probabilité de succès (de réussir) quand on est classé dans le groupe II.

En utilisant les règles de Bayes pour les probabilités conditionnelles à savoir $P(A/B)P(B) = P(B/A)P(A)$ nous sommes en mesure de calculer les probabilités pour qu'un étudiant classé dans le groupe I échoue et la probabilité qu'un étudiant classé dans le groupe II réussisse. Les équations qui ont été utilisées sont :

$$P(E/G_1) = \frac{P(E).P(G_1/E)}{P(G_1)} \tag{7.5}$$

$$P(S/G_2) = \frac{P(S).P(G_2/S)}{P(G_2)} \tag{7.6}$$

Nous pouvons donc constater à travers le tableau 7.5 que la probabilité pour un étudiant qui va échouer à être classé dans le groupe I (groupe des étudiants qui vont échouer) est plus grande avec la règle que nous proposons que celle proposée par la littérature sur la première et après six (6) semaines d'interaction avec les vidéos. Après une semaine, il y a également la probabilité pour qu'un étudiant qui réussit au cours soit classé dans le groupe II (groupe de ceux qui vont réussir) de la règle proposée soit également plus grande que celle de la littérature. Seule après six (6) semaines la méthode de la littérature a une probabilité supérieure à

celle de la règle justifiant ainsi son but à long terme pour identifier les étudiants qui peuvent réussir. Notons que ce qui intéresserait davantage l'instructeur est de pouvoir identifier le plus tôt possible dans le cours les étudiants susceptibles d'échouer afin de leur apporter l'aide nécessaire pour leur éviter l'échec.

Le tableau 7.4 montre les résultats du classement des étudiants en deux groupes selon, d'un côté, la méthodologie proposée par He, Zheng et al. 2018 (Tableau 7.2) et, de l'autre, la règle définie ci-dessus (Tableau 7.3), tout en précisant entre parenthèses la proportion que chaque groupe représente par rapport au nombre total des étudiants ayant réussi ou échoué (Tableau 7.4).

Nous constatons qu'après une semaine, les étudiants qui ont réussi le cours tombent en grande majorité dans le groupe II (78%). La majorité des étudiants ayant échoué se retrouvent dans le groupe I (60%). Nous pouvons voir qu'environ 4/5 des étudiants qui réussiront le cours seront dans le groupe II, alors qu'environ 60 % des étudiants qui échoueront seront dans le groupe I. Nous ne pouvons pas faire ce constat clair avec la méthodologie proposée par He, Zheng et al. 2018. En effet les résultats suivant la méthodologie de He, Zheng et al. 2018 montrent que la majorité des étudiants ayant réussi (72%) et la majorité des étudiants ayant échoué (84%) sont tous dans un même groupe, le groupe I.

Cette tendance (la majorité des étudiants ayant réussi au groupe II et la majorité des étudiants ayant échoué au groupe I) est encore plus forte à la mi-parcours, après la sixième semaine pour la méthodologie proposée (93% des étudiants ayant réussi se retrouvent dans le groupe II et 67% des étudiants ayant échoué se retrouvent dans le groupe I). A la mi-parcours nous pouvons tout de même noter que la méthodologie de He, Zheng et al. 2018 est en mesure de faire passer la majorité de étudiants ayant réussi dans le groupe II (67%) et la majorité des étudiants ayant échoué dans le groupe I (87%).

Ces résultats montrent que la méthode proposée apporte une amélioration par rapport à l'étude précédente qui utilisait le même type de mesures agglomératives pour signaler le risque d'échec des étudiants. Après la première semaine d'interaction déjà, l'on peut avoir deux groupes où la majorité des étudiants qui vont réussir ou échouer se retrouvent dans les deux groupes différents. Cette tendance se maintient à mi-parcours du cours.

Dans les résultats obtenus en reproduisant la méthode de He, Zheng et al. 2018 nous ne

pouvons dégager dans aucun groupe la majorité des étudiants ayant réussi ou échoué après la première semaine (la majorité des étudiants qui ont réussi ou échoué se retrouvent dans le groupe I en même temps). A la mi-parcours du cours, il faut noter tout de même une nette amélioration de performance dans la méthode de He, Zheng et al. 2018 on peut remarquer que la majorité des étudiants qui ont échoué (84%) se retrouvent dans le groupe I, alors que la majorité des étudiants qui ont réussi (67%) se retrouvent dans le groupe II, comme le montre la table 7.4. Même si nous constatons que la grande proportion des étudiants qui avait échoué dans le groupe I, à la mi-parcours du cours de la méthodologie de He, Zheng et al. 2018 est supérieure à celle de la méthode proposée, nous pouvons voir qu'elle est n'est pas en mesure de dégager une nette majorité des étudiants ayant réussi dans le groupe II, en tenant compte que les données sont disproportionnées entre les étudiants ayant réussi et ceux ayant échoué (9 454 étudiants ayant échoué soit 90% des étudiants pour 970 ayant réussi, soit seulement 10% des étudiants).

Ainsi la méthode de He, Zheng et al. 2018 fonctionne mieux pour nos données pour la prédiction des réussites (Tableau 7.5 des probabilités pour réussir lorsqu'un étudiant est classe dans le groupe II) qu'à partir de la mi-parcours du cours (après 6 semaines d'un cours de treize semaines). Il semblerait donc que leur méthode serait plus adaptée pour une plus longue période que pour une période courte comme une première semaine de cours. Cela justifie bien le fait qu'ils aient proposé leur méthode pour prédire des réussites, abandons et échecs dans le cadre d'un programme de cours sur quatre ans. Contrairement à leur méthode, celle que nous proposons, basée sur la moyenne de l'index de visionnage marche aussi bien pour des périodes plus courtes et pourrait donc être utilisée pour la prédiction précoce des étudiants à risque (échec) dans le cadre d'un cours en ligne. Ceci permettrait d'identifier des étudiants à risques pour éventuellement les aider et prévenir les échecs.

7.4 Conclusion

Grâce à une analyse quantitative comprenant des mesures agglomératives, tel le taux d'assiduité (AR), le taux d'utilisation (UR) et l'indice de visionnage (WI), tel que définis dans ce chapitre, il est possible d'identifier les patterns d'échec des étudiants qui échoueront le cours en se basant sur la première semaine d'interaction des étudiants avec les vidéos du MOOC, sur la base d'une durée totale du cours de treize (13) semaines avec une erreur de 7% à mi-parcours. De même, à mi-parcours du cours, les deux tiers (67%) des étudiants ayant échoué ont été identifiés comme susceptibles d'abandonner le cours.

Ainsi, la contribution de cette recherche est qu'en utilisant les mesures agglomératives analytiques telles que définies, les institutions éducatives peuvent identifier les étudiants qui vont échouer en se basant sur les interactions de l'étudiant avec les vidéos d'apprentissage dans les premières étapes du cours. De tels résultats devraient aider les développeurs du MOOC à mieux identifier les étudiants qui pourraient échouer le cours et définir quelles actions pourraient être prises, au moment approprié, pour aider les étudiants qui ont besoin d'un soutien académique supplémentaire pour améliorer leurs performances et leur poursuite dans le MOOC.

Avec de tels résultats des prédictions précoces des succès des étudiants, on pourrait se demander à juste titre si l'on ne peut pas trouver d'autres méthodes de prédictions précoces de la réussite des étudiants. Partant du fait que la représentation d'interactions vidéo des étudiants peut révéler le style particulier d'interaction de ces derniers comme nous l'avons montré au chapitre 6 avec la représentation TMED, nous pouvons raisonnablement penser que la représentation TMED pourrait également distinguer les styles d'interactions conduisant au succès ou à l'échec de l'étudiant. Le prochain chapitre va utiliser la représentation TMED, plus en mesure de discriminer les écoutes pour pouvoir prédire les succès des étudiants et comparer les performances de prédiction obtenues à celles de ce chapitre.

CHAPITRE 8 PRÉDICTION PRÉCOCE DU SUCCÈS DES ÉTUDIANTS BASÉE SUR LA REPRÉSENTATION D'INTERACTIONS VIDÉO TMED

8.1 Introduction

Dans les cours en ligne, la détection précoce des étudiants susceptibles d'échouer et de ceux qui vont réussir a été un sujet d'intérêt soutenu dans l'analyse des données d'apprentissage en ligne. Par exemple, le fait de savoir quels étudiants risquent d'échouer pourrait aider les instructeurs à les orienter vers des services spéciaux ou à leur fournir une aide supplémentaire pour prévenir l'échec.

Le problème de la prédiction précoce des étudiants à risque a été étudié à partir de sources multiples : amélioration du comportement des clics (Wolff et al. 2013), devoirs et interactions sociales (S. Jiang et al. 2014 ; R. S. Baker et al. 2015), l'historique des étudiants (Kloft et al. 2014), cours mixtes avec composante en salle (Sheshadri et al. 2019 ; R. Raga et J. Raga 2019 ; Sinha et Cassell 2015). Or, l'importance des vidéos dans la diffusion des contenus de cours en ligne ne cesse de croître. Les traces des interactions des étudiants sont des sources d'informations qui s'avèrent très utiles pour la prédiction de leurs succès. Cette recherche examine comment, à partir de l'interaction de l'étudiant avec la vidéo, on peut prédire leurs succès dans le cours.

Dans le but d'une prédiction précoce du succès ou non des étudiants, nous utiliserons la représentation TMED proposée au chapitre 6 basée sur une matrice de transition transformée (à partir des séquences d'activités). Nous avons démontré la sensibilité de cette représentation en testant sa capacité à discriminer l'interaction vidéo de l'étudiant et à reconnaitre la vidéo à partir de la représentation TMED des interactions des étudiants. Notre hypothèse est qu'elle est également assez puissante pour détecter si ces interactions sont indicatives d'une personne qui réussira ou échouera le cours. Par conséquent, du fait que la représentation TMED a la capacité de reconnaître une interaction particulière de l'étudiant avec la vidéo, elle peut également identifier les patterns d'interaction conduisant à la réussite ou à l'échec d'un étudiant. A cet effet, nous chercherons à savoir si la représentation TMED, combinée à un algorithme de classification standard, peut prédire les succès de l'étudiant à la fin du cours, en termes de réussite ou d'échec. La contribution de ce chapitre sera soumise pour publication à la conférence EDM 2021(*"Edicational Data Mining"*) sous le titre : *"Video interaction based student performance prediction in MOOC."*.

Dans ce chapitre, nous gardons l'objectif du chapitre 7 de prédire de façon précoce l'échec ou le succès de l'étudiant afin de permettre aux instructeurs d'agir en conséquence. C'est ainsi que les données d'interaction vidéo de la première semaine vont être principalement l'objet de nos prédictions dans ce chapitre. Nos prédictions vont prendre en considération les neuf (9) vidéos de la première semaine de cours. A partir des TMED de chaque étudiant, avec les vidéos de la première semaine du cours, nous serons en mesure de prédire si l'étudiant réussira ou pas le cours. De plus, nous étendons notre étude de prédiction jusqu'à la mi-parcours du cours pour vérifier la consistance de la méthode de prédiction dans l'amélioration de sa performance au fur et à mesure que l'on augmente les données. Cette extension au soixante-dix-neuf (79) vidéos jusqu'à la mi-parcours du cours nous permet également de comparer les résultats de cette recherche a deux recherches précédentes ayant testé les performances de prédictions de leur méthode après la première semaine et à mi-parcours du cours comme présenté dans le chapitre 7 de cette thèse.

Dans cette recherche, nous sélectionnons les données parmi les 10 424 étudiants qui ont interagi avec au moins la moitié des soixante-dix-neuf (79) vidéos jusqu'à la mi-parcours du cours. C'est ainsi que nous obtenons 4 800 étudiants qui ont significativement interagi avec les vidéos jusqu'à la mi-parcourt du cours. Donc, nous reproduisons ici l'étude du chapitre 7 aux 4 800 étudiants et le comparons aux résultats obtenus avec les mêmes étudiants en utilisant la représentation TMED.

8.2 Méthodologie

Cette étude porte sur les traces des 4 800 étudiants qui ont interagi avec des vidéos de façon significative jusqu'à la sixième semaine d'un cours de treize semaines au total. Nous avons d'abord exclu de cette étude, les étudiants qui n'ont pas interagi avec toutes les vidéos de la première semaine mais également ceux qui ont interagi avec moins de la moitié des vidéos jusqu'à la mi-parcours du cours. Parmi les étudiants retenus pour cette étude, 722 ont réussi le cours et 4 078 ont échoué. Le choix des vidéos de la première semaine a été motivé par l'objectif de détection précoce des échecs.

Comme nous allons utiliser des classificateurs pour prédire à partir du TMED de chaque vidéo, nous avons pour besoin de comparaison deux scénarios. Le premier scénario est celui où nous avons des données balancées avec le même nombre des étudiants qui ont réussi le

cours et ceux qui ont échoués. En effet, dans ce scénario, nous gardons tous les étudiants qui ont réussi soit 722 étudiants et choisissons au hasard 722 étudiants parmi ceux qui ont échoué. Pour second scénario, nous gardons les données non balancées en gardant tous les 4 800 étudiants sélectionnés pour cette étude avec 85% des étudiants ayant échoué et 15% ayant réussi (4078 versus 722 étudiants). Nous obtenons ainsi, un ensemble des données balancées de 1444 étudiants et un autre non balancées de 4 800 étudiants pour nos prédictions. Nous avons entrainé nos classificateurs pour chaque vidéo sur 80% des données représentatifs des deux catégories des étudiants (80% des étudiants ayant réussi et autant ayant échoué).Cette prédiction est faite pour chacune des vidéos en gardant les données des mêmes étudiants comme ensemble test (un ensemble test pour les données balancées avec 288 étudiants dont 144 de chaque groupe et un autre ensemble test pour les données non balancées composé de 960 étudiants dont 144 étudiants ayant réussi et 816 étudiants ayant échoué). A la fin cette phase de prédiction pour chaque classificateur, l'on obtient pour chaque étudiant de cet ensemble test, des prédictions de succès ou d'échec.

Notre prédiction finale va prendre en considération les prédictions de chaque vidéo en prenant la majorité des prédictions (succès ou échec). Par exemple pour une prédiction d'échec aux vidéos 2, 4, et 7 (3 vidéos) et succès aux vidéos 1,3,5,6,8 et 9 (6 vidéos) avec un classificateur, cet étudiant sera considéré comme pouvant réussir comme prédiction finale. Pour vérifier la consistance de cette procédure, nous avons regardé les prédictions intermédiaires de chaque étudiant après 3 vidéos, 5 vidéos, 7 vidéos et enfin 9 vidéos (en gardant le principe de la majorité). Si notre procédure est consistante, l'on s'entendrait à une amélioration dans la prédiction au fur et à mesure que les données augmentent. Nous présenterons ces résultats intermédiaires pour voir cette évolution dans l'amélioration de performance de prédiction au fur et à mesure que nous augmenterons les données. La séquence des opérations pour la prédiction du succès ou échec d'un étudiant se retrouve dans le schéma 8.1.

Pour obtenir les TMED, chaque ensemble d'interactions des étudiants a été organisé en une séquence d'activités. Ces séquences d'activités ont toutes été converties sous la forme de la représentation TMED, comme expliqué au chapitre 6. Nous avons alors obtenu pour chaque étudiant des matrices TMED différentes correspondant aux interactions avec chaque vidéo. Ainsi, notre base de données pour les prédictions était constituée de 4 800 matrices TMED obtenu auprès de ces 4 800 étudiants pour chaque vidéo lorsque tous les étudiants ont interagi avec ladite vidéo. Nos données pour cette étude sont composées des vingt-cinq (25) cellules de la matrice TMED transformée sous forme de vecteur de prédicteur pour chaque classificateur. Le facteur que nous voulons prédire pour chaque TMED est le succès ou l'échec de

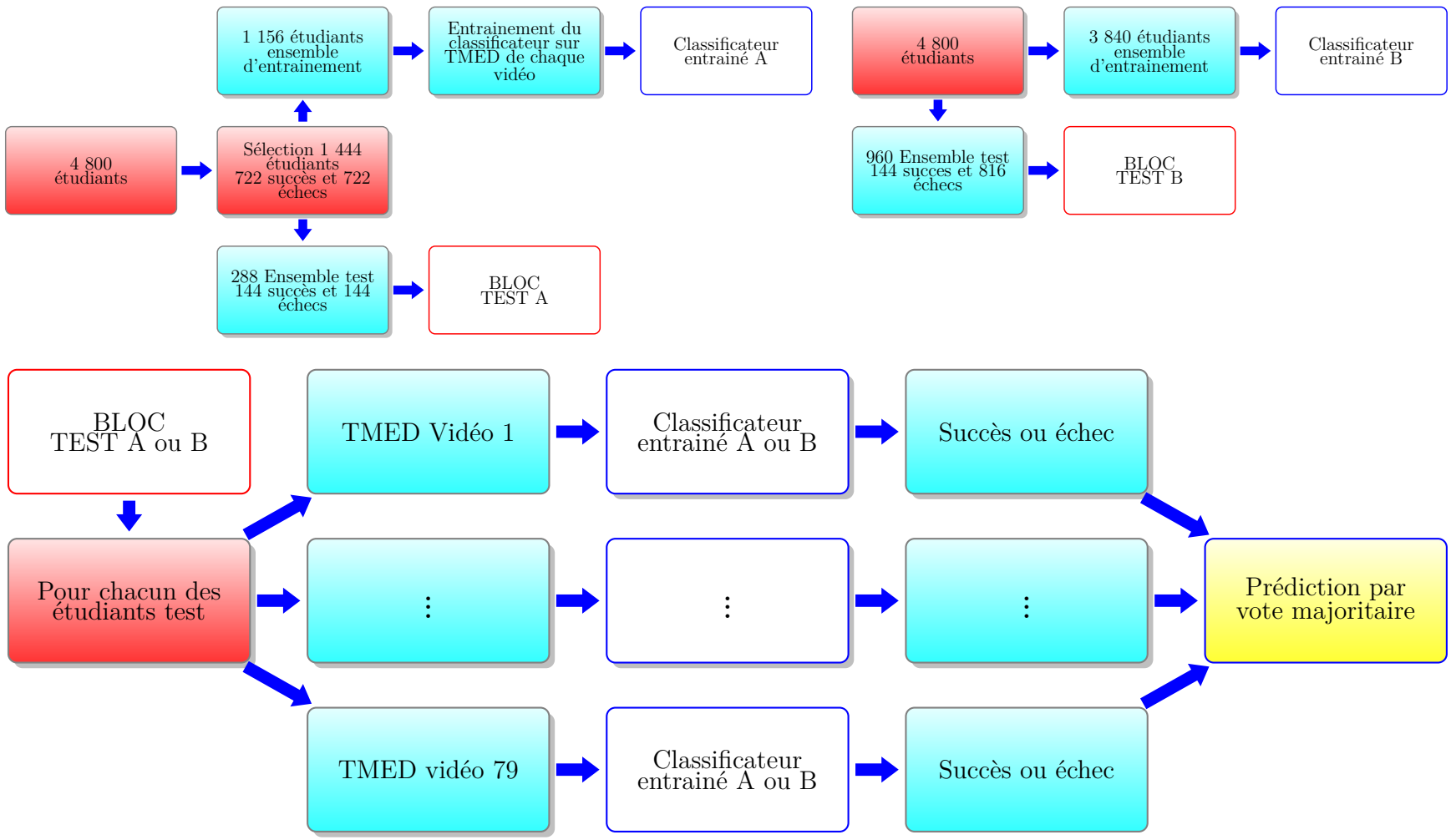


Figure 8.1 Flux de prédiction de la réussite ou de l'échec d'un étudiant. Le classificateur peut être SVM, GBM, KNN ou RF. La sélection de la prédiction finale pour chaque étudiant se fait en choisissant la majorité des prédictions des 3, 5, 7, 9 ou 79 vidéos selon l'étape de prédiction. Nous prédisons ici pour les données balancées (A) et les données non balancées (B).

l'étudiant. La même procédure est utilisée pour le reste des vidéos jusqu'à la mi-parcours du cours.

8.2.1 Classificateurs

La prédiction de la réussite ou de l'échec est effectuée à l'aide de quatre classificateurs différents avec validations croisées à cinq (5) replis. Les quatre classificateurs sont la Machine à vecteur de soutien (SVM), les Arbres de décision (GBM), les Voisins les plus proches (KNN) et la Forêt aléatoire (RF). Pour le classificateur Forêt aléatoire (RF), nous effectuons une modélisation avec une validation croisée répétée 10 x 10. Nous avons utilisé la technique de sur-échantillonnage avec ré-échantillonnage de validation croisée. La performance du modèle final est ensuite mesurée sur l'ensemble test. Chaque élément du *TMED* représentant le nombre normalisé de transitions décrit ci-dessus est un descripteur (caractéristiques utilisées pour la prédiction).

Nous avons utilisé le sur-échantillonnage de la classe minoritaire au lieu de la combinaison proposée par Chawla et al. 2002 qui est une combinaison d'une méthode de sur-échantillonnage de la classe minoritaire et de sous-échantillonnage de la classe majoritaire pour obtenir une meilleure performance du classificateur (dans l'espace ROC). Nous n'avons pas également opté pour le sous-échantillonnage de la classe majoritaire comme la technique ROSE mise en œuvre par Lunardon, Menardi et Torelli 2014 pour les données non équilibrées. En fait, nous avons essayé ces deux techniques et nous avons réalisé qu'elles atteignent une plus grande précision mais avec une très faible spécificité, ce qui n'est pas désirable dans une prédiction binaire. Cela signifie que ces méthodes ne permettent pas de prédire la présence d'une petite classe, c'est-à-dire des étudiants qui réussissent le cours dans notre cas. Dans la première expérience, on donne aux classificateurs un vecteur de caractéristique de longueur 25 (la matrice de transition de 5 par 5) et l'étiquette cible est le numéro d'identification de l'étudiant. Dans la deuxième expérience, le vecteur de caractéristique est le même et le label cible est *succès* ou *échec*.

8.2.2 Prédiction des succès des étudiants

L'expérience vise à évaluer la capacité de la représentation TMED à prédire le succès ou l'échec d'un étudiant. Compte tenu des notes des étudiants à la fin du cours, nous avons considéré comme "échec" tous les étudiants qui ont une note finale inférieure à 50 % et comme "réussite" tous les étudiants qui ont une note finale de 50 % et plus.

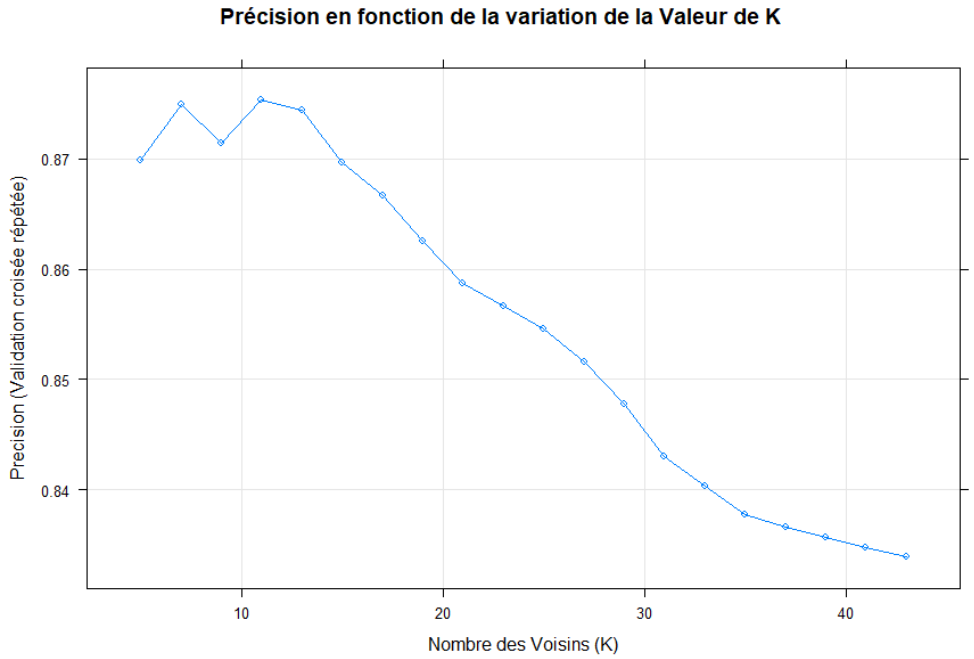


Figure 8.2 Exemple de la variation de la valeur de K pour déterminer la valeur de K avec la meilleure précision de la prédiction en utilisant KNN. Ici c'est l'exemple de la variation de K pour la vidéo 1. Ici k=11 est retenu pour les prédictions.

Nous avons donc gardé le même ensemble d'étudiants pour le test pour chacune des vidéos. La prédiction de la réussite ou l'échec prend en compte toutes les prédictions de la première semaine de l'étudiant pour prédire son succès ou non après la première semaine d'interactions et prend en compte toutes les vidéos avec lesquels l'étudiant a interagi jusqu'à la mi-parcours du cours. Pour cela, nous avons évalué la prédiction basée sur chaque vidéo dans premier temps (voir le bloc "succès ou échec" de la figure 8.1 et les résultats dans la table 8.1 pour les neuf (9) vidéos de la première semaine). Puis, nous avons progressivement étudié la capacité du TMED à prédire le succès ou non d'un étudiant en prenant en compte trois (3), cinq (5), sept (7) et neuf (9) vidéos dans l'ordre de leur apparition dans la première semaine. Nous choisissons les nombres impairs des vidéos pour forcer un vote majoritaire. D'un côté, il s'agit ici de voir si lorsqu'on ajoute des données la performance de la prédiction s'améliore, de l'autre, il s'agit de comparer nos résultats de performance de prédiction après la première semaine (9 vidéos) à celles obtenues pour la même période basée sur les mesures agglomératives du chapitre 7 qui utilise une procédure de prédiction basée sur des règles de classification. Nous gardons cette même procédure (voir figure 8.1) pour prédire si un étudiant va réussir ou échouer en considérant toutes les vidéos avec lesquelles les étudiants ont interagi jusqu'à la mi-parcours du cours.

Dans la présentation de nos résultats, nous allons prendre en considération la capacité de chaque classificateur à prédire les étudiants qui ont réussi en même temps que les étudiants qui ont échoué. Car si un classificateur prédit tous les étudiants de l'ensemble test comme ayant réussi ou échoué uniquement, il serait difficile de juger sa capacité à discriminer les deux types d'étudiants. Nous allons donc prendre en considération la sensibilité et la spécificité de chaque classificateur. La sensibilité est le pourcentage des étudiants ayant échoué le cours et ayant été prédit comme telle par le classificateur alors que la spécificité est le pourcentage des étudiants ayant passé et qui ont été prédit comme tel par le classificateur. Ces deux mesures sont exprimées comme suit :

$$Sensibilité = \frac{[Nombre\ des\ étudiants\ ayant\ échoué\ le\ cours\ prédit\ comme\ tel]}{[Nombre\ total\ des\ étudiants\ ayant\ échoué\ le\ cours]} \tag{8.1}$$

$$Spécificité = \frac{[Nombre\ des\ étudiants\ ayant\ réussi\ le\ cours\ prédit\ comme\ tel]}{[Nombre\ total\ des\ étudiants\ ayant\ réussi\ le\ cours]} \tag{8.2}$$

8.3 Résultats

L'expérience est conçue pour tester si la réussite ou l'échec des étudiants peut être prédit sur la base de la TMED des étudiants à partir de leurs données des interactions vidéo de la première semaine et de la mi-parcours cours.

La première partie de nos expérimentations pour tester prédire le succès ou l'échec de l'étudiant basé sur la représentation TMED des étudiants pour déterminer si un étudiant va réussir ou échouer le cours utilisant uniquement la vidéo. Il s'agit dans un premier temps de se baser sur chacune des vidéos pour prédire le résultat final de l'étudiant. Ici nous ne pouvons pas nous fier aux résultats des performances par vidéo pour jauger la capacité discriminatoire de TMED. En effet, un étudiant peut être engagé sur une seule vidéo pour une raison ou l'autre et se désengager pour la suite des vidéos jusqu'à l'échec du cours. Donc les performances de prédiction en prendre en compte doit tenir compte d'un ensemble successif de vidéos. Dans cette perspective, nous avons présenté les prédictions de 3, 5, 7, 9 et même 79 vidéos successives qui montrent la capacité de TMED à discriminer les étudiants qui vont réussir le cours des autres. Les résultats de prédiction par vidéo se retrouvent dans la table 8.1 pour la première semaine et les résultats de prédictions par ensemble de vidéos dans la

table 8.2. Les résultats proches du hasard pour les données non balancées de la table 8.1 confirment la suggestion des études précédentes selon laquelle l'interaction avec une vidéo seule ne peut servir pour prédire le succès ou l'échec d'un étudiant dans un cours.

Le but de cette étude étant de pouvoir faire de prédiction précoce du succès ou de l'échec de l'étudiant à partir des interactions vidéo, nous avons tenu en compte toutes les vidéos de chaque étudiant pour prédire sa performance dans le cours. Après une semaine d'interaction, l'on peut prédire le succès ou échec des étudiants avec une précision de 63% en utilisant le classificateur SVM, une précision de 67% en utilisant le classificateur KNN, 64% avec GBM et 75% avec RF pour les données balancées. Dans le cas des données non balancées qui sont plus proches de la réalité des données des cours en ligne, l'on est capable, en utilisant la représentation TMED et la méthode proposée dans cette étude, de prédire jusqu'à 54% de précisons avec GBM et KNN, 61% avec SVM et 75% avec RF. Après la mi-parcours du cours ces résultats sont nettement meilleurs dans les deux cas aussi bien pour les données balancées que les données non balancées. Ces résultats qui se retrouvent dans le tableau 8.2 montrent une amélioration de performance de prédiction par rapport aux études précédentes basées sur les mesures agglomératives présentées dans le chapitre 7 dont la meilleure performance précision de prédiction est de 60% après la première semaine et 79% à la mi-parcours du cours pour les données non balancées. Ce résultat montre que la représentation TMED peut être utilisée pour les prédictions précoces de la performance de l'étudiant en termes de succès ou d'échec.

Après six semaines, les prédictions du succès ou d'échec des étudiants basées sur soixante-dix-neuf (79) vidéos montrent une nette amélioration de performance par rapport à la première semaine (voir table 8.3). Par la méthode utilisée l'on est capable à mi-parcours du cours de prédire avec une précision allant jusqu'à 86% avec le classificateur GBM, une précision de 82% avec SVM et 85% avec KNN et RF pour les données non balancées. Cela montre bien qu'au fur et à mesure du parcours du cours les prédictions se précisent.

Prédictions :									
Approche :	Boosted Three : GBM								
Vidéo :	1	2	3	4	5	6	7	8	9
Bal. Accuracy	0.54(0.68)	0.52(0.56)	0.52(0.53)	0.53(0.50)	0.50(0.50)	0.51(0.53)	0.51(0.66)	0.54(0.67)	0.51(0.64)
Sensibilité	0.56(0.78)	0.49(0.33)	0.40(0.58)	0.60(0.58)	0.58(0.52)	0.58(0.64)	0.59(0.83)	0.60(0.82)	0.44(0.84)
Spécificité	0.52(0.58)	0.54(0.79)	0.63(0.49)	0.44(0.42)	0.40(0.49)	0.44(0.42)	0.42(0.49)	0.49(0.53)	0.60(0.44)
F_1	0.48(0.68)	0.44(0.53)	0.41(0.53)	0.41(0.49)	0.45(0.50)	0.46(0.53)	0.46(0.65)	0.42(0.67)	0.42(0.62)
Approche :	Support Vector Machine : SVM								
Bal. Accuracy	0.60(0.68)	0.50(0.55)	0.66(0.52)	0.56(0.51)	0.50(0.50)	0.54(0.52)	0.54(0.63)	0.54(0.61)	0.55(0.63)
Sensibilité	0.63(0.88)	0.48(0.26)	0.25(0.51)	0.22(0.51)	0.45(0.48)	0.56(0.60)	0.54(0.88)	0.20(0.97)	0.57(0.93)
Spécificité	0.41(0.49)	0.51(0.83)	0.91(0.53)	0.92(0.51)	0.47(0.51)	0.39(0.44)	0.33(0.38)	0.90(0.26)	0.42(0.33)
F_1	0.48(0.67)	0.42(0.51)	0.56(0.52)	0.50(0.51)	0.40(0.50)	0.43(0.51)	0.44(0.60)	0.48(0.56)	0.45(0.59)
Approche :	K Nearest Neighbor : KNN								
Bal. Accuracy	0.53(0.69)	0.50(0.52)	0.53(0.51)	0.56(0.51)	0.51(0.49)	0.52(0.55)	0.54(0.60)	0.54(0.60)	0.53(0.62)
Sensibilité	0.55(0.66)	0.48(0.54)	0.51(0.51)	0.52(0.51)	0.48(0.49)	0.48(0.54)	0.57(0.59)	0.64(0.59)	0.51(0.60)
Spécificité	0.50(0.73)	0.51(0.52)	0.55(0.51)	0.60(0.51)	0.54(0.49)	0.56(0.56)	0.51(0.63)	0.43(0.63)	0.54(0.64)
F_1	0.46(0.68)	0.42(0.50)	0.45(0.51)	0.47(0.51)	0.43(0.49)	0.43(0.54)	0.47(0.59)	0.49(0.60)	0.45(0.62)
Approche :	Random Forest : RF								
Bal. Accuracy	0.53(0.73)	0.50(0.73)	0.51(0.53)	0.51 (0.51)	0.54(0.57)	0.53(0.52)	0.51(0.64)	0.52(0.67)	0.50(0.65)
Sensibilité	0.74(0.78)	0.12(0.78)	0.81(0.66)	0.74(0.51)	0.85(0.59)	0.86(0.28)	0.89(0.69)	0.83(0.74)	0.76(0.77)
Spécificité	0.32(0.68)	0.84(0.68)	0.21(0.41)	0.28(0.51)	0.22(0.54)	0.20(0.76)	0.13(0.58)	0.86(0.60)	0.23(0.53)
F_1	0.51(0.73)	0.49(0.73)	0.51(0.53)	0.50(0.51)	0.54(0.57)	0.53(0.49)	0.51(0.64)	0.52(0.67)	0.49(0.65)

Tableau 8.1 Résultats de la validation croisée avec 5 replis de prédiction des succès ou non des étudiants pour chaque vidéo. Les données sont non balancées avec 15% des étudiants qui ont réussi et 85% qui ont échoué par vidéo (soit 722 et 4078 étudiants) et entre parenthèses, les données sont balancées avec 50% d'étudiants qui ont réussi et 50% d'étudiants qui ont échoué par vidéo (soit 722 et 722 étudiants).

Prédictions :					
Approche :	Boosted Three : GBM				
	Première semaine				Après six semaines
Vidéos :	1,2,3	1,...,5	1,...,7	1,...,9	1,...,79
Accuracy	0.52(0.50)	0.56(0.50)	0.58(0.56)	**0.58(0.64)**	**0.86(0.81)**
Sensibilité	0.17(0.49)	0.18(0.48)	0.18(0.57)	**0.18(0.68)**	**0.51(0.82)**
Specificité	0.87(0.47)	0.87(0.48)	0.87(0.56)	**0.86(0.62)**	**0.93(0.80)**
F_1	0.45(0.46)	0.47(0.48)	0.48(0.56)	**0.47(0.64)**	**0.73(0.81)**
Approche :	Support Vector Machine : SVM				
Accuracy	0.59(0.59)	0.60(0.62)	0.61(0.62)	**0.61(0.63)**	**0.82(0.78)**
Sensibilité	0.19(0.58)	0.21(0.61)	0.19(0.64)	**0.20(0.74)**	**0.46(0.74)**
Specificité	0.88(0.59)	0.88(0.63)	0.88(0.60)	**0.89(0.58)**	**0.99(0.82)**
F_1	0.49(0.59)	0.51(0.62)	0.50(0.62)	**0.51(0.60)**	**0.75(0.78)**
Approche :	K Nearest Neighbor : KNN				
Accuracy	0.50(0.58)	0.51(0.60)	0.52 (0.61)	**0.54(0.67)**	**0.85(0.71)**
Sensibilité	0.14(0.58)	0.15(0.58)	0.15 (0.63)	**0.16(0.70)**	**0.49(0.83)**
Specificité	0.85(0.58)	0.85(0.64)	0.86(0.60)	**0.86(0.64)**	**0.90(0.64)**
F_1	0.43(0.58)	0.43(0.60)	0.44(0.61)	**0.46(0.66)**	**0.68(0.71)**
Approche :	Ramdom Forest : RF				
Accuracy	0.70(0.68)	0.70(0.69)	0.71(0.73)	**0.75(0.75)**	**0.85(0.82)**
Sensibilité	0.12(0.69)	0.12(0.73)	0.13(0.75)	**0.12(0.79)**	**0.50(0.95)**
Specificité	0.84(0.67)	0.85(0.66)	0.85(0.71)	**0.84(0.71)**	**0.85(0.75)**
F_1	0.47(0.68)	0.48(0.68)	0.48(0.73)	**0.45(0.74)**	**0.49(0.82)**

Tableau 8.2 Résultats de la prédiction des succès des étudiants en combinant les résultats des données de plusieurs vidéos pour prédire le succès ou non des étudiants. Notez une augmentation de précision à mesure que les données augmentent. Le résultat de la première semaine est celle de la combinaison des neuf (9) vidéos et à mi- parcourt du cours (six semaines) celles des soixante-dix-neuf (79) vidéos. Les résultats entre parenthèses sont des résultats des données balancées avec 722 étudiants de chaque classe. En dehors des parenthèses sont les résultats des données non balancées soit de 4800 étudiants au complet. En gras sont les valeurs de performances de prédiction après la première semaine d'interaction et à la mi-parcours du cours (après six semaines).

8.4 Comparaison des performances de classification de TMED aux études précédentes

Bien que les résultats obtenus montrent que la méthode basée sur la mesure agglomérative WI proposée dans le chapitre 7 a des performances meilleures que celle de la littérature basée sur WR, il serait bon de pouvoir comparer les performances de prédiction de ces deux méthodes (ces deux méthodes utilisent des règles de classification) à l'approche de prédiction

proposée dans ce chapitre basée sur la représentation TMED (qui utilise des classificateurs). Comme les méthodes ont été toutes testées sur les mêmes données, la comparaison des performances peut se faire directement. La présentation des performances des méthodes basées sur les mesures agglomératives d'écoutes vidéo présentées dans ce chapitre seront exprimés en terme la précision balancée (*"Balance Accuracy"*) pour éviter l'influence d'une classe majoritaire (la classe des étudiants qui ont échoué : 4 078 étudiants) par rapport à la classe minoritaire (la classe des étudiants qui ont réussi : 722 étudiants) dans les prédictions utilisant les règles. Ainsi, nous pouvons les comparer aux précisions balancées obtenues par les classificateurs en utilisant les TMED. La table 8.3 montre les performances de toutes les méthodes utilisées pour prédire le succès des étudiants en termes de succès ou échec. Comme dans le cas des prédictions basées sur TMED les données sont de la première semaine du cours et à mi-parcourt, nous nous contenterons de comparer ici les performances de toutes les méthodes sur les prédictions obtenues après la première semaine du cours et à titre de comparaison aux études précédentes à mi-parcourt après six semaines. Ceci nous permet d'identifier la méthode qui donne des meilleurs résultats dans la prédiction précoce du succès ou d'échec de l'étudiant basé sur ses interactions vidéo (table 8.3).

Nous remarquons que de par les valeurs élevées de la spécificité dans les prédictions basées sur TMED (tables 8.2 et 8.3), cette dernière est capable de mieux prédire les étudiants qui ont réussi le cours par rapport aux études précédentes. La prédiction du groupe minoritaire dans le cas des données non balancées est une tâche plus ardue. Dans le cas des données balancées le pourcentage de prédiction des deux groupes sont semblables dans l'utilisation du TMED.

	Après une semaine						Après six semaines					
Méthode :	1	2	3				1	2	3			
Approche :	WR	WI	GBM	SVM	KNN	RF	WR	WI	GBM	SVM	KNN	RF
Type :	Données non balancées											
Accuracy	0.52	0.60	**0.58**	**0.61**	**0.54**	**0.75**	0.66	0.79	**0.86**	**0.82**	**0.85**	**0.85**
Sensibilité	0.86	0.87	**0.18**	**0.20**	**0.16**	**0.18**	0.96	0.97	**0.51**	**0.46**	**0.49**	**0.50**
Specificité	0.14	0.18	**0.86**	**0.89**	**0.86**	**0.85**	0.29	0.29	**0.93**	**0.99**	**0.90**	**0.85**
F_1	0.44	0.49	**0.47**	**0.51**	**0.46**	**0.49**	0.59	0.61	**0.73**	**0.75**	**0.68**	**0.49**
Type :	Données balancées											
Accuracy	0.32	0.39	**0.64**	**0.63**	**0.67**	**0.75**	0.65	0.78	**0.81**	**0.78**	**0.71**	**0.82**
Sensibilité	0.30	0.40	**0.68**	**0.74**	**0.70**	**0.79**	0.64	0.75	**0.82**	**0.74**	**0.87**	**0.95**
Specificité	0.34	0.38	**0.62**	**0.58**	**0.64**	**0.71**	0.66	0.83	**0.80**	**0.82**	**0.64**	**0.75**
F_1	0.32	0.39	**0.64**	**0.60**	**0.66**	**0.74**	0.65	0.78	**0.81**	**0.78**	**0.68**	**0.82**

Tableau 8.3 Comparaison des performances de trois méthodes de prédiction de la réussite des étudiants à la fin du cours en termes de succès ou d'échec après la première semaine d'interaction vidéo et après six semaines pour les données non équilibrées et entre parenthèses les résultats des données équilibrées. Nous comparons ici les performances de deux méthodes différentes de prédiction de la réussite ou de l'échec des étudiants : l'une basée sur des règles de classification et l'autre utilisant des classificateurs. Méthode 1 = méthode proposée dans He, Zheng et al. 2018 basée sur WR (utilisation de règles),
Méthode 2 = méthode proposée dans Mbouzao, Michel C Desmarais et Shrier 2020 basée sur WI (utilisation de règles),
Méthode 3 = méthode proposée dans cette étude basée sur TMED (utilisation de classificateurs)
En gras les performances de prédiction de la méthode proposée basée sur la représentation TMED. Il faut noter ici que pour les données balancées les deux méthodes précédentes performent en déca du hasard après la première semaine mais dans la réalité, les données ne sont toujours pas balancées en général.

8.5 Conclusion et travaux futurs

Ce chapitre aborde la question de la recherche d'une méthode plus efficace pour identifier un modèle de représentation des interactions qui permet de classifier les étudiants qui vont réussir ou non le cours. Tenant en compte le fait que la représentation TMED d'interaction vidéo d'étudiant prend en compte à la fois les transitions d'état de la vidéo et le temps passé dans chaque état, il s'agit de voir sa capacité à discriminer les interactions vidéo de réussite ou non. En particulier, le fait que la représentation TMED a une bonne capacité de discriminer les interactions vidéo des étudiants et les vidéos (comme le montre le chapitre 6), il est logique de présumer son potentiel à mieux discriminer les interactions de réussite des autres. Notre étude montre que la représentation TMED de l'interaction vidéo des étudiants peut être utilisée pour prédire les succès précoces des étudiants.

La deuxième conclusion est qu'en se basant uniquement sur l'interaction vidéo des étudiants de la première semaine d'un cours en ligne de treize semaines, on peut prédire avec une précision raisonnable (mieux que les méthodes basées sur les mesures agglomératives) la réussite des étudiants. Cette conclusion pourrait répondre à l'un des besoins des instructeurs et des développeurs en matière d'identification précoce des étudiants d'un MOOC qui risquent d'échouer, et pourrait fournir une aide supplémentaire pour les empêcher d'échouer.

La comparaison de trois méthodes de prédictions de succès ou d'échec des étudiants, dont une est mentionnée dans la revue de la littérature (He, Zheng et al. 2018) et deux autres méthodes que nous avons proposées, montre clairement que les méthodes que nous proposons dans les chapitres 7 et 8 ont des meilleures performances de prédictions. En fonction de ce que l'on peut récolter en termes des données brutes du serveur, l'utilisation de l'une de ces méthodes proposées pourrait aider à la détection précoce des étudiants à risque dans un cours en ligne.

Dans nos recherches à venir, nous comptons combiner cette source d'information basée uniquement sur les interactions vidéo avec les résultats d'autres études qui s'alimentent de données universitaires et d'autres informations d'enregistrement des interactions avec la plateforme, pour prédire la réussite des étudiants. Il s'agit d'une première tentative de prédiction des performances des étudiants basée uniquement sur les interactions vidéo des étudiants. Les résultats sont prometteurs et doivent être consolidés au cours des investigations futures.

Un autre aspect de la recherche à venir est de se demander pourquoi certains vidéos sont plus aptes à discriminer les étudiants qui vont réussir ou échouer ? La présence ou non des questions post vidéo entrant en compte dans la note finale pourrait justifier en partie cette capacité pour certaines vidéos d'être plus aptes que d'autres à discriminer par les TMED des étudiants qui vont réussir. Il s'agira de déterminer si un tel facteur peut justifier et à quelle hauteur.

Une autre recherche en lien avec la capacité de la représentation TMED à discriminer les étudiants qui vont réussir le cours par rapport d'autres représentations, porterait sur la manière dont les étudiants diffèrent les uns des autres dans leurs interactions vidéo ? Quel est l'élément de la représentation TMED qui permet de différencier l'interaction d'un étudiant par rapport à un autre ? Quel est l'élément de la représentation TMED qui permet de distinguer les étudiants qui vont réussir de ceux qui vont échouer ? Il serait aussi intéressant d'explorer ce qui est à l'origine de ce qui distingue les matrices TMED de deux étudiants différents.

CHAPITRE 9 CONCLUSIONS ET TRAVAUX FUTURS

Le développement des systèmes d'apprentissage en ligne est de plus en plus spectaculaire de nos jours avec l'universalisation de l'accès à internet. Les personnes n'ont plus besoin de quitter leur lieu géographique pour accéder à une formation de bonne qualité. Plus particulièrement, les MOOC (*Massive Open Online Courses*) sont les plus populaires des systèmes d'apprentissage en ligne. Ces cours ouverts en ligne de grande envergure (MOOC) s'appuient souvent sur la vidéo comme premier choix de contenu médiatique. Compte tenu de leur importance, il n'est pas surprenant que de nombreuses études portant sur la manière dont les étudiants utilisent les vidéos dans le cadre des MOOC aient vu le jour ces dernières années.

Pour nos investigations, la collecte des traces d'un cours de McGill sur deux semestres, l'automne 2015 et l'hiver 2016, nous a permis de faire notre recherche dans le cadre de cette thèse. Nous avons essentiellement développé des techniques pouvant aider à mieux analyser les données récoltées principalement dans le cadre de la représentation des interactions vidéo des étudiants et la prédiction précoce de la réussite ou l'échec des étudiants. Ces techniques peuvent aider à fournir le tableau de bord des instructeurs pour une meilleure orientation du cours mais également à améliorer ses versions futures. Ainsi, nous avons développé cinq techniques qui tournent autour de deux grands axes, à savoir la représentation des interactions vidéo pouvant servir à diverses tâches de classification d'écoute vidéo et des techniques pour la prédiction précoce des succès des étudiants permettant éventuellement à l'instructeur de centrer son attention sur les étudiants à risque pour éviter les échecs. Ainsi dans cette thèse nous avons pu apporter des contributions qui répondent à nos questions de recherche suivantes :

1. *QR.1 : La représentation SIVS est-elle plus performante que la représentation cumulative pour discriminer des écoutes de durée semblables ?*

   Comme mentionné dans l'introduction de cette thèse, nous utilisons le terme "mesures cumulatives" pour désigner les mesures d'écoutes vidéo au niveau d'une vidéo à savoir le pourcentage de temps qu'un étudiant à consacrer à jouer une vidéo, à la position pause d'une vidéo par rapport au temps total passer à interagir avec la vidéo puis le nombre de fois des recherches en arrière et en avant dans la vidéo considérée. Nous appelons "mesures agglomératives", les mesures d'interactions qui concernent un ensemble des

vidéos (par exemple les vidéos de la première semaine) pour déterminer divers aspects de l'utilisation de ces vidéos par un étudiant en particulier. Ces contributions peuvent se résumer en cinq axes qui touchent la méthodologie d'analyse d'écoute vidéo à travers l'utilisation des traces.

Une représentation d'interaction vidéo, que nous avons nommé SIVS (*Sequence of Interaction in Vector Space*), est introduite pour l'analyse détaillée des habitudes de visionnement des vidéos par les étudiants et comparée à une approche très connue (l'approche basée sur les mesures agglomératives d'écoute vidéo). La représentation que nous avons proposée encode les séquences d'interaction vidéo dans un espace vectoriel sous forme de matrice. Nous avons défini les mesures de distance entre elles en calculant la norme de Frobenius entre elles. Nous avons également une écoute type de la vidéo que nous avons défini comme centroïde des séquences d'interaction d'une même vidéo pour déterminer une sorte de regroupement des écoutes d'une même vidéo. Nous avons utilisé cette représentation SIVS et mené une étude sur la façon dont les étudiants interagissent avec les vidéos en analysant les distances entre les centroïdes des différentes vidéos. Dans le cadre de la validation de cette représentation, l'analyse de l'influence de la vidéo sur la façon d'interagir des étudiants avec elle est comprise comme une tâche de classification. En utillisant les approches de la machine à vecteur de soutien (SVM), de l'arbre de décision (GBM) et du plus proche voisin (KNN), nous étions capables de classifier les interactions en fonction de la vidéo avec laquelle les étudiants interagissaient. Les résultats révèlent qu'il existe une différence significative dans la façon dont les étudiants interagissent avec chaque vidéo. Cela démontre l'utilité de la représentation SIVS proposée par rapport à la représentation basée sur les calculs agglomératives d'écoute vidéo qui ne pouvait pas distinguer les différences qu'imposent les vidéos dans l'interaction des étudiants.

Les résultats montrent également que la représentation SIVS a l'avantage de fournir un encodage plus précis d'ensemble d'interactions individuelles des étudiants par rapport au modèle basé sur les calculs cumulatifs d'écoute vidéo qui est souvent utilisé dans l'étude des traces d'interaction vidéo des étudiants. Il permet de définir des "clusters" (regroupements) étiquetés, les diverses écoutes vidéo dans notre cas, qui contrastent avec les "clusters" non étiquetés standard qui nécessitent une interprétation des "clusters".

Dans les cours ouverts en ligne (MOOC), les étudiants interagissent avec les vidéos en

faisant des pauses, en cherchant à avancer ou à reculer, en rejouant des segments. La représentation SIVS proposée a l'avantage de pouvoir faire ressortir dans sa structure tous ces éléments d'analyses.

2. *QR. 2 : La présentation TMED est-elle plus performante que les approches séquentielles et de chaînes de transition pour discriminer entre différents types d'écoutes ? La représentation TMED est-elle plus performante que d'autres représentations dans la recherche de similarité entre les écoutes vidéo ?*

   De par la diversité d'interactions vidéo des étudiants dans un cours en ligne, il est raisonnable de supposer que les étudiants ont des façons d'interagir avec les vidéos qui diffèrent. Mais il reste tout de même difficile de comparer les interactions vidéo des étudiants et de mesurer le degré de ressemblance entre deux interactions vidéo. Certaines techniques ont été développées, telles que celle basée sur la chaîne de Markov et celle à partir de la distance d'édition entre les séquences d'activités. Cependant, ces techniques comportent des réserves, comme c'est le cas avec des interactions ayant les mêmes probabilités de transitions entre les états ou les mêmes styles cycliques d'interaction (les cas des prototypes abordés dans cette thèse). Nous avons ainsi proposé une méthodologie de comparaison des séquences d'interaction vidéo basée sur une représentation particulière que nous avons appelée TMED qui tient en compte à la fois le temps passé dans chaque état et la succession des états en calculant la distance entre les matrices de transition des séquences d'interaction vidéo. Les résultats montrent que la méthodologie proposée permet de mieux comparer les interactions des étudiants avec les vidéos en utilisant la représentation TMED.

   La représentation TMED et la méthode de similarité proposées en ce sens comblent une lacune méthodologique sur la représentation pour la comparaison des séquences vidéo d'interaction. La représentation proposée surmonte les limites des représentations précédentes basées sur la chaîne de Markov (difficulté à rendre compte du temps passé dans chaque état) et les séquences d'interactions connues sous le nom de "Edit Distance based" (génération de grande distance dès qu'il y a un décalage de séquence due à une durée un peu plus longue dans un état pour une des séquences comparées). La principale contribution de cette méthode proposée est le fait qu'elle prend en compte le temps passé dans chaque état et le style général de succession des états. Elle offre une technique aux chercheurs qui souhaitent comparer les interactions vidéo des étudiants dans un

système d'apprentissage en ligne et trouver éventuellement un style d'interaction vidéo.

La représentation TMED proposée combine à cet effet les apports de deux styles de représentation de la séquence vidéo d'interaction et calcule la similarité en bénéficiant de l'avantage de chaque style de représentation.

3. *QR. 3 : Un étudiant possède-t-il un style d'écoute qui lui est propre ? Une vidéo possède-t-elle aussi une signature d'écoute ?*

La représentation TMED peut être utilisée pour reconnaitre la façon particulière d'un étudiant à interagir avec les vidéos. Nous avons comparé la capacité de la représentation TMED à rendre compte de la reconnaissance des interactions des étudiants avec d'autres représentations d'interactions vidéo. Les résultats montrent que la représentation TMED est capable de reconnaitre les interactions vidéo des étudiants mieux que les représentations en chaîne de Markov et en séquence d'interactions. La représentation TMED est également capable de mieux représenter une séquence d'interaction lors de tâches de classification, comme le montrent nos résultats. En fait, la représentation TMED est plus performante dans l'identification de la séquence d'interaction de l'étudiant comparée aux deux autres représentations d'interaction vidéo.

La représentation TMED est capable de mieux spécifier une séquence d'interaction lors de tâches de classification, comme le montrent les résultats. En fait, la représentation TMED proposée a une meilleure performance dans l'identification de l'étudiant et de la vidéo que les deux autres représentations d'interactions vidéo (représentation en chaîne de Markov et en séquence d'interactions).

Déjà lors de la validation de la représentation SIVS, nous avons démontré la capacité de SIVS à identifier, à travers une représentation d'interaction, la vidéo avec laquelle l'étudiant interagit mieux qu'avec les autres représentations à l'exemple de la représentation basée sur les mesures cumulatives d'écoute vidéo.

Ces résultats montrent finalement qu'un étudiant a un style d'écoute qui lui est propre et reconnaissable. Donc chaque étudiant possède sa signature d'écoute reconnaissable. D'un autre côté en ayant la possibilité de reconnaitre à travers les interactions la vidéo, nous avons montré que chaque vidéo a sa signature d'écoute également.

4. *QR. 4 : Les mesures agglomératives permettent-elles de prédire avec précision les chances du succès des étudiants ?*

   La popularité de l'apprentissage en ligne, tel celle des systèmes MOOC, continue à augmenter parmi les étudiants. Cependant, le taux d'abandon dans les MOOC reste très élevé. La prédiction précoce des succès ou des échecs des étudiants pourrait alimenter les tableaux de bord des instructeurs et les aider à adapter leur cours. Revoir la structure du cours, réviser le matériel, ou simplement venir en aide au groupe d'étudiants à risque d'échec ou encore l'adaptation des interventions à des groupes spécifiques d'étudiants, est un objectif de recherche. À cette fin, la recherche de He, Zheng et al. 2018 a introduit trois mesures agglomératives (le taux d'assiduité : AR (*"Attendance Rate"*), le taux d'utilisation : UR (*"Utilization Rate"*), et le taux de visionnage : WR (*"Watching Ratio"*)) pour prédire les succès des étudiants. La limite de leur méthodologie est qu'elle dépend des facteurs graphiques qui peuvent être laissés à l'appréciation subjective du chercheur (pour déterminer la valeur seuil de WR). Pour dépasser cette limite, nous introduisons une nouvelle mesure que nous appelons l'indice de visionnage WI (*"Watching Index"*) basée sur AR et UR des étudiants interagissant avec les vidéos du MOOC afin de prédire quel groupe d'étudiants réussira ou échouera le cours. L'apport principal de la mesure WI est de pouvoir simplifier la classification des étudiants et de ne plus dépendre de la composante graphique dans la division des groupes des étudiants (détermination du seuil de WR).

   Grâce à l'analyse quantitative qui comprend des mesures telles que le taux d'assiduité, le taux d'utilisation et l'indice de visionnage, il est possible d'identifier les patterns d'échec jusqu'à 60 % des étudiants qui échoueront le cours en se basant sur les mesures d'interaction vidéo des étudiants durant la première semaine du cours. Sachant que la durée totale du cours considéré pour l'étude est de treize (13) semaines, on peut identifier 93 % des étudiants qui réussiront à partir des mesures de la sixième semaine d'interaction. En utilisant ces mesures, les établissements d'enseignement peuvent signaler les étudiants à risque d'échec, ou d'abandon, du MOOC en fonction des interactions de l'étudiant avec les vidéos d'apprentissage dès les premières semaines du cours. Notre étude montre une meilleure classification par rapport aux résultats de l'étude précédente de He, Zheng et al. 2018 qui n'est pas en mesure de séparer après la première semaine les étudiants qui vont échouer des étudiants qui vont réussir (91% des

étudiants qui vont échouer et 85% des étudiants qui vont réussir se retrouvent dans un même groupe). Par ailleurs, après six semaines, une distinction entre les deux groupes est remarque avec l'identification de 65% des étudiants qui vont réussir et 93% des étudiants qui vont échouer. Une étude sur les probabilités de détection des étudiants qui vont réussir ou échouer a été réalisé sur les performances de ces deux méthodes montrant que la méthode proposée est plus performante dans la prédiction précoce du succès ou échec des étudiants. De tels résultats devraient aider les développeurs des MOOC à mieux identifier les étudiants susceptibles d'abandonner ou d'échouer le cours et donc à prendre des mesures pour l'éviter.

5. *QR. 5 : Comment la représentation TMED se compare-t-elle aux autres méthodes de prédiction de succès ?*

   Les performances de la prédiction du succès des étudiants en utilisant la représentation TMED pour être valide ont été comparé aux résultats obtenus dans les recherches précédentes publiées dans deux conférences différentes basées sur des mesures agglomératives d'écoute vidéo. Cette comparaison montre que les résultats de prédictions obtenus par l'utilisation de la représentation TMED sont plus performants que celles obtenues par ces deux autres méthodes basées sur des mesures agglomératives d'écoute vidéo des étudiants. Notre étude montre ainsi que la représentation proposée TMED de l'interaction vidéo des étudiants peut être utilisée pour les problèmes de classification et notamment de prédiction précoce du succès des étudiants dans le cadre d'un cours en ligne.

   Après une semaine d'interaction, l'on peut prédire le succès ou échec des étudiants avec une précision de 63% en utilisant le classificateur SVM, une précision de 67% en utilisant le classificateur KNN, 64% avec GBM et 75% avec RF dans une donnée balancée (même nombre d'étudiants dans les deux classes : ceux qui ont passé le cours et ceux qui ont échoué). Dans le cas des données non balancées qui sont plus proches de la réalité des données des cours en ligne, l'on est capable, en utilisant la représentation TMED et la méthode proposée dans cette étude, de prédire jusqu'à 54% de précisons avec GBM et KNN, 61% avec SVM et 75% avec RF. Après la mi-parcours du cours ces résultats sont nettement meilleurs dans les deux cas aussi bien pour les données balancées que les données non balancées. Ces résultats montrent une amélioration de performance de prédiction par rapport aux études précédentes basées sur les mesures

agglomératives présentées dans le chapitre 7 dont la meilleure performance précision de prédiction est de 60% après la première semaine et 79% à la mi-parcours du cours pour les données non balancées. Ce résultat montre que la représentation TMED peut être utilisée pour les prédictions précoces de la performance de l'étudiant en termes de succès ou d'échec.

Après six semaines, les prédictions du succès ou d'échec des étudiants basées sur soixante-dix-neuf (79) vidéos montrent une nette amélioration de performance par rapport à la première semaine. Par la méthode utilisée l'on est capable à mi-parcours du cours de prédire avec une précision allant jusqu'à 86% avec le classificateur GBM, une précision de 82% avec SVM et 85% avec KNN et RF pour les données non balancées. Cela montre bien qu'au fur et à mesure du parcours du cours les prédictions se précisent en utilisant la méthode basée sur TMED et la prédiction par vote majoritaire.

Dans les travaux futurs, il serait bon de combiner cette source d'information, basée sur la vidéo, avec les résultats d'autres études qui s'appuient sur des données académiques et d'autres informations d'interaction avec le système en dehors des vidéos pour augmenter la précision de la prédiction de la réussite ou de l'échec des étudiants et pouvoir alimenter les tableaux de bord des enseignants. Il s'agit en termes de prédiction des succès des étudiants d'une tentative basée uniquement sur les interactions vidéo des étudiants. Les résultats sont prometteurs et doivent être consolidés dans les futures investigations prenant en compte plusieurs types de systèmes d'apprentissage en ligne.

Une autre investigation futur, consistera à aller plus loin dans la capacité de la représentation d'interaction vidéo proposée TMED. Nous pouvons dire que la représentation TMED proposée ouvre de nouvelles voies de recherche. Elle pourrait aider à déterminer s'il existe une cohérence dans l'interaction vidéo pour chaque étudiant en fonction du niveau de similarité de leurs séquences. On peut également utiliser cette méthode pour savoir si chaque style vidéo impose un style d'interaction spécifique aux utilisateurs de la vidéo. Une autre piste consiste dans le lien que l'on peut trouver entre le style vidéo d'interaction et le succès des étudiants. Les différents styles d'interaction peuvent-ils conduire à l'échec ou à la réussite dans un cours en ligne ? Des recherches futures répondront à certaines de ces interrogations.

Dans le domaine de classification des séquences d'interaction vidéo des étudiants, l'utilisation de la représentation SIVS doit pouvoir être généralisée dans un contexte de classification

supervisé. Notamment dans les constructions des clusters plus homogènes autour du modèle proposé de centroïde d'écoutes. Un travail futur consisterait à pouvoir comparer la construction des clusters basés sur la représentation SIVS et le modèle de centroïde avec le modèle classique de regroupement supervisé à l'exemple des regroupements hiérarchiques avec un nombre prédefinit des clusters.

La limite de la représentation SIVS repose sur le fait qu'elle est très liée à la durée de la vidéo. Ainsi toutes les études qui utiliseront cette représentation seront limitées à utiliser des données provenant d'une même vidéo ou des données de même durée. Dans les MOOC modernes, la durée des vidéos varie d'une leçon à l'autre ; alors que pouvoir comparer les écoutes des diverses vidéos reste important. Et à cause de la limite de la représentation SIVS liée à la durée de la vidéo, il ne serait pas possible d'utiliser cette représentation pour pouvoir étudier toutes les vidéos d'un MOOC avec des durées différentes.

RÉFÉRENCES

[1] Andrew Abbott. *Time matters: On theory and method*. University of Chicago Press, 2001.

[2] Ma Almeda et al. "Comparing the Factors That Predict Completion and Grades Among For-Credit and Open/MOOC Students in Online Learning." In : *Online Learning* 22.1 (2018), p. 1-18.

[3] Skand Arora et al. "Learner groups in massive open online courses". In : *American Journal of Distance Education* 31.2 (2017), p. 80-97.

[4] Thushari Atapattu et Katrina Falkner. "Discourse analysis to improve the effective engagement of MOOC videos". In : *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 2017, p. 580-581.

[5] Irfan Baig. "Impact of peer-supported video analysis of classroom interactions on teacher understanding of those interactions". Thèse de doct. 2012.

[6] Ryan S Baker et al. "Analyzing Early At-Risk Factors in Higher Education E-Learning Courses." In : *International Educational Data Mining Society* (2015).

[7] Elena Baralis et Luca Cagliero. "Learning from summaries: Supporting e-learning activities by means of document summarization". In : *IEEE Transactions on Emerging Topics in Computing* 4.3 (2015), p. 416-428.

[8] Rebecca Barber et Mike Sharkey. "Course correction: Using analytics to predict course success". In : *Proceedings of the 2nd international conference on learning analytics and knowledge*. 2012, p. 259-262.

[9] Naima Belarbi et al. "User profiling in a SPOC: A method based on user video clickstream analysis". In : *International Journal of Emerging Technologies in Learning (iJET)* 14.01 (2019), p. 110-124.

[10] Yoav Bergner, Deirdre Kerr et David E Pritchard. "Methodological Challenges in the Analysis of MOOC Data for Exploring the Relationship between Discussion Forum Views and Learning Outcomes." In : *International Educational Data Mining Society* (2015).

[11] Suma Bhat, Phakpoom Chinprutthiwong et Michelle Perry. "Seeing the Instructor in Two Video Styles: Preferences and Patterns." In : *International Educational Data Mining Society* (2015).

[12] Sarah Bishara et al. “Revealing Interaction Patterns Among Youth in an Online Social Learning Network Using Markov Chain Principles”. In : Philadelphia, PA: International Society of the Learning Sciences., 2017.

[13] Fernanda Cesar Bonafini. “The Effects of Participants’ Engagement with Videos and Forums in a MOOC for Teachers’ Professional Development.” In : *Open Praxis* 9.4 (2017), p. 433-447.

[14] Mina Shirvani Boroujeni et Pierre Dillenbourg. “Discovery and temporal analysis of latent study patterns in MOOC interaction sequences”. In : *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. 2018, p. 206-215.

[15] Sebastien Boyer et Kalyan Veeramachaneni. “Transfer learning for predictive models in massive open online courses”. In : *International conference on artificial intelligence in education*. Springer. 2015, p. 54-63.

[16] Cynthia J Brame. “Effective educational videos: Principles and guidelines for maximizing student learning from video content”. In : *CBE—Life Sciences Education* 15.4 (2016), es6.

[17] Eric Brangier et Michel C Desmarais. “The design and evaluation of the persuasiveness of e-learning interfaces”. In : *International Journal of Conceptual Structures and Smart Applications (IJCSSA)* 1.2 (2013), p. 38-47.

[18] L. Brantley-Dias et al. “The role of digital video and critical incident analysis in learning to teach science.” In : *In Proceedings of the American Educational Research Association Annual Meeting* (2008).

[19] Nicholas Breakwell et Dara Cassidy. “Surviving the avalanche: Improving retention in MOOCs”. In : *6th International Conference of MIT’s Learning International Networks Consortium (LINC), Cambridge, MA*. 2013.

[20] Lori Breslow et al. “Studying learning in the worldwide classroom: Research into edX’s first MOOC”. In : *Research & Practice in Assessment* 8 (2013).

[21] Christopher G Brinton, Swapna Buccapatnam et al. “Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance”. In : *IEEE Transactions on Signal Processing* 64.14 (2016), p. 3677-3692.

[22] Christopher G Brinton, Mung Chiang et al. “Learning about social learning in MOOCs: From statistical analysis to generative model”. In : *IEEE transactions on Learning Technologies* 7.4 (2014), p. 346-359.

[23] Remi Brochenin et al. “Resource usage analysis from a different perspective on MOOC dropout”. In : *arXiv preprint arXiv:1710.05917* (2017).

[24] Christopher Brooks, Craig Thompson et Stephanie Teasley. “A time series interaction analysis method for building predictive models of learners using log data”. In : *Proceedings of the fifth international conference on learning analytics and knowledge*. 2015, p. 126-135.

[25] Barbara Caputo et al. “Appearance-based object recognition using SVMs: which kernel should I use?” In : (2001).

[26] Wen-Hsuan Chang, Jie-Chi Yang et Yu-Chieh Wu. “A keyword-based video summarization learning platform with multimodal surrogates”. In : *2011 IEEE 11th International Conference on Advanced Learning Technologies*. IEEE. 2011, p. 37-41.

[27] Mohamed Amine Chatti et al. “Video annotation and analytics in CourseMapper”. In : *Smart Learning Environments* 3.1 (2016), p. 10.

[28] Jyoti Chauhan et Anita Goel. “An Analysis of Video Lecture in MOOC.” In : *ICTERI*. 2015, p. 35-50.

[29] Nitesh V Chawla et al. “SMOTE: synthetic minority over-sampling technique”. In : *Journal of artificial intelligence research* 16 (2002), p. 321-357.

[30] Haijian Chen et al. “Classification and analysis of MOOCs learner’s state: The study of hidden Markov model”. In : *Computer Science and Information Systems* 16.3 (2019), p. 849-865.

[31] Liang Chen, Yipeng Zhou et Dah Ming Chiu. “A study of user behavior in online VoD services”. In : *Computer Communications* 46 (2014), p. 66-75.

[32] Liang Chen, Yipeng Zhou et Dah Ming Chiu. “Video browsing-a study of user behavior in online vod services”. In : *2013 22nd International Conference on Computer Communication and Networks (ICCCN)*. IEEE. 2013, p. 1-7.

[33] Qing Chen et al. “Peakvizor: Visual analytics of peaks in video clickstreams from massive open online courses”. In : *IEEE transactions on visualization and computer graphics* 22.10 (2015), p. 2315-2330.

[34] Yishuai Chen et al. “On distribution of user movie watching time in a large-scale video streaming system”. In : *2014 IEEE International Conference on Communications (ICC)*. IEEE. 2014, p. 1825-1830.

[35] Yuanzhe Chen et al. “DropoutSeer: Visualizing learning patterns in Massive Open Online Courses for dropout reasoning and prediction”. In : *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE. 2016, p. 111-120.

[36] Konstantinos Chorianopoulos. “A taxonomy of asynchronous instructional video styles”. In : *International Review of Research in Open and Distributed Learning* 19.1 (2018).

[37] Konstantinos Chorianopoulos et Michail N Giannakos. “Usability design for video lectures”. In : *Proceedings of the 11th european conference on Interactive TV and video*. 2013, p. 163-164.

[38] Krishna Choudhari et Vinod K Bhalla. “Video search engine optimization using keyword and feature analysis”. In : *Procedia Computer Science* 58 (2015), p. 691-697.

[39] Sila Chunwijitra et al. “Authoring tool for video-based content on WebELS learning system to support higher education”. In : *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*. IEEE. 2012, p. 317-322.

[40] Meg Colasante. “Using video annotation to reflect on and evaluate physical education pre-service teaching practice”. In : *Australasian Journal of Educational Technology* 27.1 (2011).

[41] Katherine Renee Comeaux. “Cognitive memory effects on non-linear video-based learning”. In : (2005).

[42] Rianne Conijn, Antoine Van den Beemt et P Cuijpers. “Predicting student performance in a blended MOOC”. In : *Journal of Computer Assisted Learning* 34.5 (2018), p. 615-628.

[43] Stephen Cummins, Alastair Beresford et Andrew Rice. “Investigating Engagement with In-Video Quiz Questions in a Programming Course”. In : (2015).

[44] Shane Dawson et al. *Using technology to encourage self-directed learning: The Collaborative Lecture Annotation System (CLAS)*. 2012.

[45] Jennifer DeBoer et al. “Diversity in MOOC students’ backgrounds and behaviors in relationship to performance in 6.002 x”. In : *Proceedings of the Sixth Learning International Networks Consortium Conference*. T. 4. 2013.

[46] Gerben W Dekker, Mykola Pechenizkiy et Jan M Vleeshouwers. “Predicting Students Drop Out: A Case Study.” In : *International Working Group on Educational Data Mining* (2009).

[47] Erhan Delen, Jeffrey Liew et Victor Willson. “Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments”. In : *Computers & Education* 78 (2014), p. 312-320.

[48] Anant Deshpande et Valeri Chukhlomin. “What makes a good MOOC: A field study of factors impacting student motivation to learn”. In : *American Journal of Distance Education* 31.4 (2017), p. 275-293.

[49] Michel Desmarais et François Lemieux. “Clustering and Visualizing Study State Sequences.” In : *EDM*. 2013, p. 224-227.

[50] Banu Diri et Songul Albayrak. “Visualization and analysis of classifiers performance in multi-class medical data”. In : *Expert Systems with Applications* 34.1 (2008), p. 628-634.

[51] D Dissanayake et al. “Identifying the learning style of students in MOOCs using video interactions”. In : *International Journal of Information and Education Technology* 8.3 (2018).

[52] Prajakta Diwanji et al. “Success factors of online learning videos”. In : *2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014)*. IEEE. 2014, p. 125-132.

[53] Steffi Domagk, Ruth N Schwartz et Jan L Plass. “Interactivity in multimedia learning: An integrated model”. In : *Computers in Human Behavior* 26.5 (2010), p. 1024-1033.

[54] Francis Donkor. “The comparative instructional effectiveness of print-based and video-based instructional materials for teaching practical skills at a distance”. In : *International Review of Research in Open and Distributed Learning* 11.1 (2010), p. 96-116.

[55] Peter J Draus, Michael J Curran et Melinda S Trempus. “The influence of instructor-generated video content on student satisfaction with and engagement in asynchronous online classes”. In : *Journal of Online Learning and Teaching* 10.2 (2014), p. 240-254.

[56] Michael Eagle et Tiffany Barnes. “Modeling student dropout in tutoring systems”. In : *Intelligent Tutoring Systems*. Springer. 2014, p. 676-678.

[57] Michael Eagle et Tiffany Barnes. “Survival analysis on duration data in intelligent tutors”. In : *Intelligent Tutoring Systems*. Springer. 2014, p. 178-187.

[58] Cees H Elzinga. “Sequence analysis: Metric representations of categorical time series”. In : *Sociological methods and research* (2006).

[59] Bruno Emond et Scott Buffett. “Analyzing Student Inquiry Data Using Process Discovery and Sequence Classification.” In : *International Educational Data Mining Society* (2015).

[60] Sean B Eom et Nicholas Ashill. “The determinants of students’ perceived learning outcomes and satisfaction in university online education: An update”. In : *Decision Sciences Journal of Innovative Education* 14.2 (2016), p. 185-215.

[61] Brent J Evans, Rachel B Baker et Thomas S Dee. “Persistence patterns in massive open online courses (MOOCs)”. In : *The Journal of Higher Education* 87.2 (2016), p. 206-242.

[62] Tz-Yung Fang et al. “A Study of Video-based Concordancer on Scene Classification”. In : *2011 IEEE 11th International Conference on Advanced Learning Technologies*. IEEE. 2011, p. 78-82.

[63] Louis Faucon, Lukasz Kidzinski et Pierre Dillenbourg. “Semi-Markov Model for Simulating MOOC Students.” In : *International Educational Data Mining Society* (2016).

[64] Mi Fei et Dit-Yan Yeung. “Temporal models for predicting student dropout in massive open online courses”. In : *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE. 2015, p. 256-263.

[65] Rebecca Ferguson et Doug Clow. “Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs)”. In : *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. ACM. 2015, p. 51-58.

[66] Michael Fire et al. “Predicting student exam’s scores by analyzing social network data”. In : *International Conference on Active Media Technology*. Springer. 2012, p. 584-595.

[67] Joann Fisher et Darrell Norman Burrell. “The value of using micro teaching as a tool to develop instructors.” In : *Review of Higher Education & Self-Learning* 3.11 (2011).

[68] Stephanie Forrest. “Emergent computation: self-organizing, collective, and cooperative phenomena in natural and artificial computing networks: introduction to the proceedings of the ninth annual CNLS conference”. In : *Physica D: Nonlinear Phenomena* 42.1-3 (1990), p. 1-11.

[69] Yoav Freund et Robert E Schapire. “A desicion-theoretic generalization of on-line learning and an application to boosting”. In : *European conference on computational learning theory*. Springer. 1995, p. 23-37.

[70] Jerome H Friedman. “Greedy function approximation: a gradient boosting machine”. In : *Annals of statistics* (2001), p. 1189-1232.

[71] Jerome H Friedman. “Stochastic gradient boosting”. In : *Computational statistics & data analysis* 38.4 (2002), p. 367-378.

[72] Chih-Hsiung Fu et al. “Building video Concordancer supported English online learning exemplification”. In : *Pacific-Rim Conference on Multimedia*. Springer. 2008, p. 731-737.

[73] Alexis Gabadinho et al. “Analyzing and visualizing state sequences in R with TraMineR”. In : *Journal of Statistical Software* 40.4 (2011), p. 1-37.

[74] Julie Gainsburg. “Creating effective video to promote student-centered teaching”. In : *Teacher Education Quarterly* 36.2 (2009), p. 163-178.

[75] Chase Geigle et ChengXiang Zhai. “Modeling MOOC student behavior with two-layer hidden Markov models”. In : *Proceedings of the fourth (2017) ACM conference on learning@ scale*. 2017, p. 205-208.

[76] Michail Giannakos et al. “Video-based learning and open online courses”. In : (2014).

[77] Michail N Giannakos, Konstantinos Chorianopoulos et Nikos Chrisochoides. “Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course”. In : *International Review of Research in Open and Distributed Learning* 16.1 (2015), p. 260-283.

[78] Maggie Celeste Goulden et al. “CCVis: Visual analytics of student online learning behaviors using course clickstream data”. In : *Electronic Imaging* 2019.1 (2019), p. 681-1.

[79] Maja Grgurović et Volker Hegelheimer. “Help options and multimedia listening: Students’ use of subtitles and the transcript”. In : *Language learning & technology* 11.1 (2007), p. 45-66.

[80] Darren K Griffin, David Mitchell et Simon J Thompson. “Podcasting by synchronising PowerPoint and voice: What are the pedagogical benefits?” In : *Computers & Education* 53.2 (2009), p. 532-539.

[81] Charles Miller Grinstead et James Laurie Snell. *Grinstead and Snell’s introduction to probability*. Chance Project, 2006.

[82] Philip J Guo, Juho Kim et Rob Rubin. “How video production affects student engagement: An empirical study of mooc videos”. In : *Proceedings of the first ACM conference on Learning@ scale conference*. ACM. 2014, p. 41-50.

[83] Philip J Guo et Katharina Reinecke. “Demographic differences in how students navigate through MOOCs”. In : *Proceedings of the first ACM conference on Learning@ scale conference*. ACM. 2014, p. 21-30.

[84] S. Halawa, D. Greene et J. Mitchell. “Dropout prediction in MOOCs using learner activity features.” In : (2014).

[85] Sherif Halawa, Daniel Greene et John Mitchell. “Dropout prediction in MOOCs using learner activity features”. In : *Experiences and best practices in and around MOOCs* 7 (2014).

[86] Anna Hansch et al. “Video and online learning: Critical reflections and findings from the field”. In : (2015).

[87] Jiangang Hao, Zhan Shu et Alina von Davier. “Analyzing process data from game/scenario-based tasks: an edit distance approach”. In : *JEDM-Journal of Educational Data Mining* 7.1 (2015), p. 33-50.

[88] Huan He, Bo Dong et al. “VUC: Visualizing Daily Video Utilization to Promote Student Engagement in Online Distance Education”. In : *Proceedings of the ACM Conference on Global Computing Education*. 2019, p. 99-105.

[89] Huan He, Qinghua Zheng et al. “Measuring Student’s Utilization of Video Resources and Its Effect on Academic Performance”. In : *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. IEEE. 2018, p. 196-198.

[90] Khe Foon Hew. “Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS”. In : *British Journal of Educational Technology* 47.2 (2016), p. 320-341.

[91] Phil Hill. “Emerging student patterns in MOOCs: A (revised) graphical view”. In : *WordPress, e-Literate* 10 (2013).

[92] Brahim Hmedna, Ali El Mezouary et Omar Baz. “An approach for the identification and tracking of learning styles in MOOCs”. In : *Europe and MENA cooperation advances in information and communication technologies*. Springer, 2017, p. 125-134.

[93] Brahim Hmedna, Ali El Mezouary, Omar Baz et Driss Mammass. “Identifying and tracking learning styles in MOOCs: A neural networks approach”. In : *International Journal of Innovation and Applied Studies* 19.2 (2017), p. 267.

[94] Tin Kam Ho. “Random decision forests”. In : *1995 IEEE 3th International conference on document analysis and recognition*. IEEE. 1995, p. 278-282.

[95] Elke Höfler, Claudia Zimmermann et Martin Ebner. “A case study on narrative structures in instructional MOOC designs”. In : *Journal of Research in Innovative Teaching & Learning* (2017).

[96] Ching-Kun Hsu et al. “Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices.” In : *Educational Technology & Society* 16.1 (2013), p. 403-414.

[97] Hui Hu et al. “Big data analytics for MOOC video watching behavior based on Spark”. In : *Neural Computing and Applications* (2019), p. 1-9.

[98] Glyn Hughes et Chelsea Dobbins. “The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs)”. In : *Research and practice in technology enhanced learning* 10.1 (2015), p. 1-18.

[99] Christina Ilioudi, Michail N Giannakos et Konstantinos Chorianopoulos. “Investigating differences among the commonly used video lecture styles”. In : (2013).

[100] Suhang Jiang et al. “Predicting MOOC performance with week 1 behavior”. In : *Educational Data Mining 2014*. 2014.

[101] Williams Jiang, Warschauer Schenke et O’dowd. “Predicting MOOC performance with week 1 behavior”. In : *Educational Data Mining*. T. EDM 2014. 2014, p. 3.

[102] Tali Kahan, Tal Soffer et Rafi Nachmias. “Types of participant behavior in a massive open online course”. In : *The International Review of Research in Open and Distributed Learning* 18.6 (2017).

[103] Junzo Kamahara et al. “Behavioral Analysis using Cumulative Playback Time for Identifying Task Hardship of Instruction Video”. In : *2010 5th International Conference on Future Information Technology*. IEEE. 2010, p. 1-6.

[104] Alexandros Karatzoglou et al. “kernlab-an S4 package for kernel methods in R”. In : *Journal of statistical software* 11.9 (2004), p. 1-20.

[105] Mohammad Khalil et Martin Ebner. “Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories”. In : *Journal of Computing in Higher Education* 29.1 (2017), p. 114-132.

[106] Mohammad Khalil et Martin Ebner. “Learning Analytics in MOOCs: Can Data Improve Students Retention and Learning?” In : *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE). 2016, p. 581-588.

[107] Khushboo Khurana et MB Chandak. “Study of various video annotation techniques”. In : *International Journal of Advanced Research in Computer and Communication Engineering* 2.1 (2013), p. 909-914.

[108] Juho Kim, Philip J Guo, Carrie J Cai et al. “Data-driven interaction techniques for improving navigation of educational videos”. In : *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 2014, p. 563-572.

[109] Juho Kim, Philip J Guo, Daniel T Seaton et al. “Understanding in-video dropouts and interaction peaks inonline lecture videos”. In : *Proceedings of the first ACM conference on Learning@ scale conference*. 2014, p. 31-40.

[110] René F Kizilcec, Kathryn Papadopoulos et Lalida Sritanyaratana. “Showing face in video instruction: effects on information retention, visual attention, and affect”. In : *Proceedings of the SIGCHI conference on human factors in computing systems*. 2014, p. 2095-2102.

[111] René F Kizilcec, Chris Piech et Emily Schneider. “Deconstructing disengagement: analyzing learner subpopulations in massive open online courses”. In : *Proceedings of the third international conference on learning analytics and knowledge*. ACM. 2013, p. 170-179.

[112] Severin Klingler et al. “Temporally Coherent Clustering of Student Data.” In : *International Educational Data Mining Society* (2016).

[113] Marius Kloft et al. “Predicting MOOC dropout over weeks using machine learning methods”. In : *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*. 2014, p. 60-65.

[114] Marcus Klüsener et Albrecht Fortenbacher. “Predicting students’ success based on forum activities in MOOCs”. In : *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. T. 2. IEEE. 2015, p. 925-928.

[115] Mirjam Köck et Alexandros Paramythis. “Activity sequence modelling and dynamic clustering for personalized e-learning”. In : *User Modeling and User-Adapted Interaction* 21.1-2 (2011), p. 51-97.

[116] Safak Korkut et al. “Success factors of online learning videos”. In : *International Journal of Interactive Mobile Technologies (iJIM)* 9.4 (2015), p. 17-22.

[117] Max Kuhn. “A Short Introduction to the caret Package”. In : *R Found Stat Comput* 1 (2015).

[118] Max Kuhn et al. “Building predictive models in R using the caret package”. In : *Journal of statistical software* 28.5 (2008), p. 1-26.

[119] Max Kuhn et Kjell Johnson. *Applied predictive modeling*. T. 26. Springer, 2013.

[120] Yu-Chun Kuo et al. “Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses”. In : *The internet and higher education* 20 (2014), p. 35-50.

[121] Yu-Chao Lai, Shelley Shwu-Ching Young et Nen-Fu Huang. “A preliminary study of producing multimedia online videos for ubiquitous learning on MOOCs”. In : *2015 8th International Conference on Ubi-Media Computing (UMEDIA)*. IEEE. 2015, p. 295-297.

[122] F Lemieux, MC Desmarais et PN Robillard. “Motivation et analyse chronologique des traces d’un exerciseur pour l’auto-apprentissage”. In : *Sciences et Technologies de l’Information et de la Communication pour L’Education et la Formation, STICEF* (2013).

[123] François Lemieux, Michel C Desmarais et Pierre-N Robillard. “Analyse chronologique des traces journalisées d’un guide d’étude pour apprentissage autonome”. In : (2014).

[124] Vladimir I Levenshtein. “Binary codes capable of correcting deletions, insertions, and reversals”. In : *Soviet physics doklady*. T. 10. 8. 1966, p. 707-710.

[125] Cen Li et Jungsoon Yoo. “Modeling student online learning using clustering”. In : *Proceedings of the 44th annual Southeast regional conference*. ACM. 2006, p. 186-191.

[126] Jing Li. “Construction of modern educational technology MOOC platform based on courseware resource storage system”. In : *International Journal of Emerging Technologies in Learning (iJET)* 12.09 (2017), p. 105-116.

[127] Liang-Yi Li et Chin-Chung Tsai. “Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance”. In : *Computers & Education* 114 (2017), p. 286-297.

[128] Nan Li, Lukasz Kidzinski et al. “How Do In-video Interactions Reflect Perceived Video Difficulty?” In : *Proceedings of the European MOOCs Stakeholder Summit 2015*. EPFL-CONF-207968. PAU Education. 2015, p. 112-121.

[129] Nan Li, Łukasz Kidziński et al. “MOOC Video Interaction Patterns: What Do They Tell Us?” In : *Design for Teaching and Learning in a Networked World*. Springer, 2015, p. 197-210.

[130] Baker R Li Q. “Understanding Engagement in MOOCs”. In : *Proceedings of the 9th International Conferences on Educational Data Mining* (2016).

[131] Shu-Sheng Liaw. “Investigating students’ perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system”. In : *Computers & education* 51.2 (2008), p. 864-873.

[132] Chih-cheng Lin et Yi-fang Tseng. “Videos and Animations for Vocabulary Learning: A Study on Difficult Words.” In : *Turkish Online Journal of Educational Technology-TOJET* 11.4 (2012), p. 346-355.

[133] Zhongxiu Liu et al. “MOOC Learner Behaviors by Country and Culture; an Exploratory Analysis”. In : *Proceedings of the 9th International Conferences on Educational Data Mining* (2016).

[134] Stephan Lorenzen, Niklas Hjuler et Stephen Alstrup. “Tracking behavioral patterns among students in an online educational system”. In : *arXiv preprint arXiv:1908.08937* (2019).

[135] Owen HT Lu et al. “Applying learning analytics for the early prediction of Students’ academic performance in blended learning”. In : *Journal of Educational Technology & Society* 21.2 (2018), p. 220-232.

[136] Xiaohang Lu et al. “What decides the dropout in MOOCs?” In : *International Conference on Database Systems for Advanced Applications*. Springer. 2017, p. 316-327.

[137] Nicola Lunardon, Giovanna Menardi et Nicola Torelli. “ROSE: A Package for Binary Imbalanced Learning.” In : *R journal* 6.1 (2014).

[138] Xiang Ma, Dan Schonfeld et Ashfaq A Khokhar. “Video event classification and image segmentation based on noncausal multidimensional hidden markov models”. In : *IEEE transactions on image processing* 18.6 (2009), p. 1304-1313.

[139] Jorge J Maldonado et al. “Exploring differences in how learners navigate in MOOCs based on self-regulated learning and learning styles: A process mining approach”. In : *2016 XLII Latin American Computing Conference (CLEI)*. IEEE. 2016, p. 1-12.

[140] Vibhu Malhotra. “PREDICTING STUDENT PERFORMANCE IN BLENDED MOOCS”. In : *Studies in Indian Place Names* 40.33 (2020), p. 42-45.

[141] Mustafa Man, Mohd Hafriz Nural Azhan et Wan Mohd Amir Fazamin Wan Hamzah. “Conceptual Model for Profiling Student Behavior Experience in e-Learning”. In : *International Journal of Emerging Technologies in Learning (iJET)* 14.21 (2019), p. 163-175.

[142] Brian Marsh et Nick Mitchell. “The role of video in teacher professional development”. In : *Teacher Development* 18.3 (2014), p. 403-417.

[143] Matthew Martin, James Charlton et Andy M Connor. “Mainstreaming video annotation software for critical video analysis”. In : *arXiv preprint arXiv:1604.05799* (2016).

[144] Boniface Mbouzao, Michel C Desmarais et Ian Shrier. “Early Prediction of Success in MOOC from Video Interaction Features”. In : *International Conference on Artificial Intelligence in Education*. Springer. 2020, p. 191-196.

[145] Danielle S McNamara et al. *Automated evaluation of text and discourse with CohMetrix*. Cambridge University Press, 2014.

[146] Martin Merkt et Stephan Schwan. “How does interactivity in videos affect task performance?” In : *Computers in Human Behavior* 31 (2014), p. 172-181.

[147] Martin Merkt, Sonja Weigand et al. “Learning with videos vs. learning with print: The role of interactive features”. In : *Learning and Instruction* 21.6 (2011), p. 687-704.

[148] Muhammad Anwar Mohd Kamal et al. “60 Seconds’ Video-based Learning’to Facilitate Flipped Classrooms and Blended Learning at a Malaysian University”. In : *Proceedings of the International Invention, Innovative & Creative (InIIC) Conference, Series*. 2019, p. 118-127.

[149] Gaëlle Molinari et al. “L’engagement et la persistance dans les dispositifs de formation en ligne: regards croisés”. In : *Distances et médiations des savoirs. Distance and Mediation of Knowledge* 13 (2016).

[150] Sylvain Mongy, Fatma Bouali et Chabane Djeraba. “Analyzing user’s behavior on a video database”. In : *Multimedia data mining and knowledge discovery*. Springer, 2007, p. 458-471.

[151] Sylvain Mongy, Chabane Djeraba et Dan A Simovici. “On Clustering Users’ Behaviors in Video Sessions.” In : *DMIN*. 2007, p. 99-103.

[152] Pedro Manuel Moreno-Marcos et al. “Prediction in MOOCs: A review and future research directions”. In : *IEEE Transactions on Learning Technologies* 12.3 (2018), p. 384-401.

[153] Žolt Namestovski et al. “External Motivation, the Key to Success in the MOOCs Framework”. In : *Acta Polytechnica Hungarica* 15.6 (2018), p. 125-142.

[154] Gonzalo Navarro. “A guided tour to approximate string matching”. In : *ACM computing surveys (CSUR)* 33.1 (2001), p. 31-88.

[155] SB Needleman. “Needleman-Wunsch algorithm for sequence similarity searches”. In : *J Mol Biol* 48 (1970), p. 443-453.

[156] Chong-Wah Ngo, Yu-Fei Ma et Hong-Jiang Zhang. “Video summarization and scene detection by graph modeling”. In : *IEEE Transactions on circuits and systems for video technology* 15.2 (2005), p. 296-305.

[157] Ozlem Ozan et Yasin Ozarslan. “Video lecture watching behaviors of learners in online courses”. In : *Educational Media International* (2016), p. 1-15.

[158] Leandro Pardo. *Statistical inference based on divergence measures*. CRC press, 2005.

[159] Amit C Patel et Mia K Markey. “Comparison of three-class classification performance metrics: a case study in breast cancer CAD”. In : *Medical imaging 2005: Image perception, observer performance, and technology assessment*. T. 5749. International Society for Optics et Photonics. 2005, p. 581-589.

[160] Nirmal Patel, Collin Sellman et Derek Lomas. “Mining frequent learning pathways from a large educational dataset”. In : *arXiv preprint arXiv:1705.11125* (2017).

[161] V Mohan Patro et Manas Ranjan Patra. “A novel approach to compute confusion matrix for classification of n-class attributes with feature selection”. In : *Transactions on Machine Learning and Artificial Intelligence* 3.2 (2015), p. 52-52.

[162] Oleksandra Poquet et al. “Video and learning: a systematic review (2007–2017)”. In : *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. 2018, p. 151-160.

[163] Shaojie Qu et al. “Predicting Student Achievement Based on Temporal Learning Behavior in MOOCs”. In : *Applied Sciences* 9.24 (2019), p. 5539.

[164] Rodolfo Raga et Jennifer Raga. “Early Prediction of Student Performance in Blended Learning Courses Using Deep Neural Networks”. In : *2019 International Symposium on Educational Technology (ISET)*. IEEE. 2019, p. 39-43.

[165] Muhamad Izzat Rahim et Sarimah Shamsudin. “Video Lecture Styles in MOOCs by Malaysian Polytechnics”. In : *Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology*. 2019, p. 64-68.

[166] Laxmisha Rai et Deng Chunrao. “Influencing factors of success and failure in MOOC and general analysis of learner behavior”. In : *International Journal of Information and Education Technology* 6.4 (2016), p. 262.

[167] Arti Ramesh et al. “Modeling learner engagement in MOOCs using probabilistic soft logic”. In : *NIPS Workshop on Data Driven Education*. T. 21. 2013, p. 62.

[168] Zhiyun Ren, Huzefa Rangwala et Aditya Johri. “Predicting performance on MOOC assessments using multi-regression models”. In : *arXiv preprint arXiv:1605.02269* (2016).

[169] Jeanine Reutemann. “Video Styles in MOOCs–A journey into the world of digital education”. In : (2016).

[170] Peter J Rich et Michael Hannafin. “Video annotation tools: Technologies to scaffold, structure, and transform teacher reflection”. In : *Journal of teacher education* 60.1 (2009), p. 52-67.

[171] Greg Ridgeway. “Generalized Boosted Models: A guide to the gbm package”. In : *Update* 1.1 (2007), p. 2007.

[172] Evan F Risko et al. “The collaborative lecture annotation system (CLAS): A new TOOL for distributed learning”. In : *IEEE Transactions on Learning Technologies* 6.1 (2012), p. 4-13.

[173] Cristobal Romero et al. “Web usage mining for predicting final marks of students that use Moodle courses”. In : *Computer Applications in Engineering Education* 21.1 (2013), p. 135-146.

[174] Shaghayegh Sahebi, Yun Huang et Peter Brusilovsky. “Predicting student performance in solving parameterized exercises”. In : *Intelligent Tutoring Systems*. Springer. 2014, p. 496-503.

[175] José Miguel Santos-Espino, Maria Dolores Afonso-Suárez et Cayetano Guerra-Artal. “Speakers and boards: A survey of instructional video styles in MOOCs”. In : *Technical Communication* 63.2 (2016), p. 101-115.

[176] Universitätsprofessor Dr-Ing Ulrik Schroeder. “Effective Design of Blended MOOC Environments in Higher Education”. In : ().

[177] Ronald Schroeter, Jane Hunter et Douglas Kosovic. “Vannotea: A collaborative video indexing, annotation and discussion system for broadband networks”. In : (2003).

[178] Daniel T Seaton, Yoav Bergner et al. “Who does what in a massive open online course?” In : (2014).

[179] Daniel T Seaton, Sergiy Nesterko et al. “Characterizing video use in the catalogue of MITx MOOCs”. In : *European MOOC Stakeholders Summit, Lausanne* (2014), p. 140-146.

[180] Tina Seidel, Geraldine Blomberg et Alexander Renkl. “Instructional strategies for using video in teacher education”. In : *Teaching and Teacher Education* 34 (2013), p. 56-65.

[181] Kyle Shaffer. “Predicting Speech Acts in MOOC Forum Posts Using Conditional Random Fields”. In : (2015).

[182] PS Shama et Pattan Prakash. “Textual Description based Video Annotation Methods”. In : *2018 International Conference on Networking, Embedded and Wireless Systems (ICNEWS)*. IEEE. 2018, p. 1-6.

[183] Kshitij Sharma, Patrick Jermann et Pierre Dillenbourg. “Displaying teacher’s gaze in a MOOC: Effects on students’ video navigation patterns”. In : *Design for Teaching and Learning in a Networked World*. Springer, 2015, p. 325-338.

[184] Adithya Sheshadri et al. “Predicting student performance based on online study habits: a study of blended courses”. In : *arXiv preprint arXiv:1904.07331* (2019).

[185] Conglei Shi et al. “VisMOOC: Visualizing video clickstream data from massive open online courses”. In : *2015 IEEE Pacific visualization symposium (PacificVis)*. IEEE. 2015, p. 159-166.

[186] Yuling Shi, Zhiyong Peng et Hongning Wang. “Modeling student learning styles in MOOCs”. In : *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 2017, p. 979-988.

[187] Madhumitha Shridharan et al. “Predictive learning analytics for video-watching behavior in MOOCs”. In : *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*. IEEE. 2018, p. 1-6.

[188] Tanmay Sinha et Justine Cassell. “Connecting the dots: Predicting student grade sequences from bursty MOOC interactions over time”. In : *Proceedings of the second (2015) ACM conference on learning@ scale*. 2015, p. 249-252.

[189] Tanmay Sinha, Patrick Jermann et al. “Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions”. In : *arXiv preprint arXiv:1407.7131* (2014).

[190] Tanmay Sinha, Nan Li et al. “Capturing" attrition intensifying" structural traits from didactic interaction sequences of MOOC learners”. In : *arXiv preprint arXiv:1409.5887* (2014).

[191] Robyn Smyth. “Enhancing learner–learner interaction using video communications in higher education: Implications from theorising about a new model”. In : *British Journal of Educational Technology* 42.1 (2011), p. 113-127.

[192] Marina Sokolova et Guy Lapalme. “A systematic analysis of performance measures for classification tasks”. In : *Information processing & management* 45.4 (2009), p. 427-437.

[193] Doraisamy Gobu Sooryanarayan et Deepak Gupta. “Impact of learner motivation on mooc preferences: Transfer vs. made moocs”. In : *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE. 2015, p. 929-934.

[194] Christian Stöhr et al. “Videos as learning objects in MOOCs: A study of specialist and non-specialist participants’ video activity in MOOCs”. In : *British Journal of Educational Technology* 50.1 (2019), p. 166-176.

[195] Steven Tang, Joshua C Peterson et Zachary A Pardos. “Modelling student behavior using granular large scale action data from a MOOC”. In : *arXiv preprint arXiv:1608.04789* (2016).

[196] Shu-Fen Tseng et al. “Who will pass? Analyzing learner behaviors in MOOCs”. In : *Research and Practice in Technology Enhanced Learning* 11.1 (2016), p. 8.

[197] Frans Van der Sluis, Jasper Ginn et Tim Van der Zee. “Explaining Student Behavior at Scale: The influence of video complexity on student dwelling time”. In : *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM. 2016, p. 51-60.

[198] Amy L Verst. “Outstanding teachers and learner-centered teaching practices at a private liberal arts institution”. In : (2010).

[199] Raphael Viguier et al. “Automatic video content summarization using geospatial mosaics of aerial imagery”. In : *2015 IEEE International Symposium on Multimedia (ISM)*. IEEE. 2015, p. 249-253.

[200] Massimo Vitiello et al. “User Behavioral Patterns and Early Dropouts Detection: Improved Users Profiling through Analysis of Successive Offering of MOOC.” In : *J. UCS* 24.8 (2018), p. 1131-1150.

[201] Ulrike Von Luxburg. “A tutorial on spectral clustering”. In : *Statistics and computing* 17.4 (2007), p. 395-416.

[202] Han Wan et al. “Predicting Performance in a Small Private Online Course.” In : *EDM*. 2017.

[203] Meng Wang et al. “Lazy Learning Based Efficient Video Annotation”. In : *2007 IEEE International Conference on Multimedia and Expo*. IEEE. 2007, p. 607-610.

[204] Wei Wang, Han Yu et Chunyan Miao. “Deep model for dropout prediction in MOOCs”. In : *Proceedings of the 2nd International Conference on Crowd Science and Engineering*. 2017, p. 26-32.

[205] Xu Wang et al. “Investigating How Student’s Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains.” In : *International Educational Data Mining Society* (2015).

[206] Yuan Wang et Ryan Baker. “Content or platform: Why do students complete MOOCs”. In : *MERLOT Journal of Online Learning and Teaching* 11.1 (2015), p. 17-30.

[207] J. Whitehill et al. “Beyond Prediction: First Steps Toward Automatic Intervention in MOOC Student Stopout.” In : (2015).

[208] Jacob Whitehill et al. “Beyond prediction: First steps toward automatic intervention in MOOC student stopout”. In : *Available at SSRN 2611750* (2015).

[209] Annika Wolff et al. “Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment”. In : *Proceedings of the third international conference on learning analytics and knowledge*. 2013, p. 145-149.

[210] Jacqueline Wong et al. “Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course”. In : *Computers & Education* 140 (2019), p. 103595.

[211] Pieter Wouters, Huib K Tabbers et Fred Paas. “Interactivity in video-based models”. In : *Educational Psychology Review* 19.3 (2007), p. 327-342.

[212] Di Wu, Yong Liu et Keith Ross. “Queuing network models for multi-channel P2P live streaming systems”. In : *IEEE INFOCOM 2009*. IEEE. 2009, p. 73-81.

[213] Wanli Xing et Dongping Du. “Dropout prediction in MOOCs: Using deep learning for personalized intervention”. In : *Journal of Educational Computing Research* 57.3 (2019), p. 547-570.

[214] Diyi Yang et al. “Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses”. In : *Proceedings of the 2013 NIPS Data-driven education workshop*. T. 11. 2013, p. 14.

[215] Jie Chi Yang et al. “An automatic multimedia content summarization system for video recommendation”. In : *Journal of Educational Technology & Society* 12.1 (2009), p. 49-61.

[216] Tsung-Yen Yang et al. “Behavior-based grade prediction for MOOCs via time series neural networks”. In : *IEEE Journal of Selected Topics in Signal Processing* 11.5 (2017), p. 716-728.

[217] Cheng Ye et Gautam Biswas. “Early prediction of student dropout and performance in MOOCs using higher granularity temporal information”. In : *Journal of Learning Analytics* 1.3 (2014), p. 169-172.

[218] Ahmed Mohamed Fahmy Yousef, Mohamed Amine Chatti, Narek Danoyan et al. “Video-mapper: A video annotation tool to support collaborative learning in moocs”. In : *Proceedings of the Third European MOOCs Stakeholders Summit EMOOCs* (2015), p. 131-140.

[219] Ahmed Mohamed Fahmy Yousef, Mohamed Amine Chatti et Ulrik Schroeder. *Video-based learning: A critical analysis of the research published in 2003-2013 and future visions*. 2014.

[220] Ahmed Mohamed Fahmy Yousef, Ulrik Schroeder et Marold Wosnitza. *Effective design of blended MOOC environments in higher education*. Rapp. tech. CiL Center for Innovative Learning Technologies, 2015.

[221] Chen-Hsiang Yu, Jungpin Wu et An-Chi Liu. “Predicting Learning Outcomes with MOOC Clickstreams”. In : *Education Sciences* 9.2 (2019), p. 104.

[222] Amelia Zafra et SebastiáN Ventura. “Multi-instance genetic programming for predicting student performance in web based educational environments”. In : *Applied Soft Computing* 12.8 (2012), p. 2693-2706.

[223] Carmen Zahn et al. “How to improve collaborative learning with video tools in the classroom? Social vs. cognitive guidance for student teams”. In : *International Journal of Computer-Supported Collaborative Learning* 7.2 (2012), p. 259-284.

[224] Emily H van Zee. “Using Web-Based”. In : *Issues in Teacher Education* 14.1 (2005), p. 63-79.

[225] Dongsong Zhang et al. “Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness”. In : *Information & management* 43.1 (2006), p. 15-27.

[226] Feng Zhang, Di Liu et Cong Liu. “MOOC Video Personalized Classification Based on Cluster Analysis and Process Mining”. In : *Sustainability* 12.7 (2020), p. 3066.

[227] Xiangyu Zhang et Huiping Lin. “Modeling and Interpreting User Navigation Patterns in MOOCs”. In : *International Conference on Frontier Computing*. Springer. 2017, p. 403-413.

ANNEXE A VIDÉOS ET ACTIVITÉS DU COURS BODY101X PARTIE 1

Unité 1 : Why it is important to be physically activity ?												
SEMAINE 1 & 2 : Module 1 : Introduction to Body 101x Leçon 1 : Overview												
vid./act.	2'33	act.1	6'59	act.2	5'04	7'20	3'05	3'57	act.3	6'14	2'57	2'20	fback
Leçon 2 : The Biggest Public Health Problem Points [/9]													
vid./act.	act.1	1'20	9'08	act.2	4'09	act.3	9'49	act.4	act.5	7'22	fback		
Leçon 3 : Introduction to physical activity Promotion Points [/3]													
vid./act.	7'02	act.1	7'18	7'50	6'45	4'55	4'08	act.2	fback	Project Points [/5]			
Module 2 : Importance of Execise Leçon 1 : Never to late Points [/16]													
vid./act.	10'43	act.1	1'34	act.2	6'12	3'04	act.3	7'05	3'20	act. 4	act.5	6'15	
Leçon 2 : Physical Literacy Points [/8]													
vid./act.	1'51	act.1	8'44	act.2	8'40	6'08	act.3	5'07	4'41	1'46	act. 4	1'53	1'52, fback

act.=activity and fback =feedback

Tableau A.1 Vidéos et activités du cours Body101x (Partie 1)

ANNEXE B VIDÉOS ET ACTIVITÉS DU COURS BODY101X PARTIE 1 SUITE

SEMAINE 3 & 4 : Module 3 : Principle of Training Leçon 1 : How should I train ? Points [/10]												
vid./act.	2'35	act.1,2	4'13	act.3,4	11'41	act.5,6	3'40	3'47	act.7,8	7'16	act.9	fback
Leçon 2 : Execise and Pregnancy Points [/6]												
vid./act.	4'58	act.1	6'55	act.2	9'36	act.3	8'02	2'24	3'33	fback		
Leçon 3 : Too much execise ? Points [/8]												
vid./act.	8'45	act.1	4'24	act.2	6'10	act.3	3'51	act.4	8'02	fback	Quiz Unité 1 : Points [/38]	
Unité 2 : What can I do to avoid injury ?												
SEMAINE 5 : Module 4 : Evaluating the evidence Leçon 1 : Science of evaluating the evidence Points [/9]												
vid./act.	act.1	8'58	8'14, act.2	6'12	act.3	11'10, act.4	9'37,act.5	6'21, act.6	fback			
Leçon 2 : Does stretching prevent injuries ? Points [/7]												
vid./act.	act.1	2'28	act.2	3'36	act.3	11'17,act.4	6'04, act.5	3'57, act.6	act.7, fback			
Leçon 3 : Doping and the physician Points [/6]												
vid./act.	6'52	7'47, act.1,2	11'18	5'31	act.3,4	7'35,act.5	3'22	10'34, act.6	5'05, fback			

act.=activity and fback =feedback

Tableau B.1 Vidéos et activités du cours Body101x (Partie 1 suite)

ANNEXE C VIDÉOS ET ACTIVITÉS DU COURS BODY101X PARTIE 2

SEMAINES 6 & 7 : Module 5 : How did I hurt myself ?									
Leçon 1 : Knee injuries in running and cycling Points [/8]									
vid./act.	5'23	act.1	7'23,act.2	8'34	act.3	8'25	act.4	2'31	act.5
Leçon 2 : Shoulder injuries Points [/16]									
vid./act.	5'15	act.1	9'42	act.2	7'20, act.3	8'08	act.4	10'02	fback
Leçon 3 : Back injuries Points [/9]									
vid./act.	4'01	act.1	7'58	act.2	act.3	5'44	act.4	5'15	fback
SEMAINE 8 : Module 6 : Dance Music and circus arts									
Leçon 1 : Injuries and the Arts									
vid./act.	2'20	7'07	4'07	4'15	7'00	act.1	5'21	fback	
Leçon 2 : When the music stops									
vid./act.	3'22	5'25	5'50	6'38	7'18	8'24	8'03	4'57	act.2, fback
Leçon 3 : Sport versus circus									
vid./act.	6'43	4'42	6'54	1'31	act.1	fback	Quiz Unité 2 : Points [/37]		

act.= activity and fback= feedback

Tableau C.1 Vidéos et activités du cours Body101x (Partie 2)

ANNEXE D VIDÉOS ET ACTIVITÉS DU COURS BODY101X PARTIE 3

Unité 3 : What can I do when i get injured ?										
SEMAINES 9 & 10 : Module 7 : When to seek help Leçon 1 : When do I need a Test ? Points [/9]										
vid./act.	6'40	5'18	act.1	8'22	3'38	act.2	10'56	act.3	fback	
Leçon 2 :Importance of Rehab Points [/15]										
vid./act.	7'53	act.1	3'39	5'55, act.2	5'56	act.3	7'48	4'39	act.4	fback
Leçon 3 :Shoulder Rehab Points [/14]										
vid./act.	6'12	act.1	8'22	act.2	3'50	3'23	act.3	7'07	act.4	fback
SEMAINES 11 & 12 : Module 8 : Return to Activity Leçon 1 : When to go back ? Points [/5]										
vid./act.	7'01	act.1	9'09	act.2	8'24	3'39	act.3	4'44	10'14	act.4, fback
Leçon 2 : Running and Osteoarthristis Points [/10]										
vid./act.	5'40	act.1	8'28	act.2	6'06	act.3	10'43	fback		
Leçon 3 : Sport concussions Points [/19]										
vid./act.	8'58	act.1,2	10'23, act.3	7'40, act.4	4'40	6'03	12'42, act.5	7'57, act.6	12'18	fback
SEMAINE 13 : Module 9 : Wrap up Leçon 1 : Live Lesson										
vid./act.	55'09	38'06	13'55	Quiz Unité 3 : Points [/34]						

act.=activity and fback =feedback

Tableau D.1 Vidéos et activités du cours Body101x (Partie 3)

ANNEXE E DESCRIPTION DES STYLES VIDÉO

De nombreuses études ont défini des styles de vidéo pouvant être utilisés dans les cours en ligne. Ces différents styles vidéo sont utilisés dans les cours de plateformes en ligne tels que Coursera, Udacity, EdX, Khan Academy et TED. Dans la littérature, il existe sept styles vidéo différents : Style tête parlante Ilioudi, M. N. Giannakos et Chorianopoulos 2013 ; Ozan et Ozarslan 2016, style de présentation Ozan et Ozarslan 2016, style image dans l'image Chorianopoulos et M. N. Giannakos 2013, style de voix au dessus de la présentation Griffin, D. Mitchell et S. J. Thompson 2009, style Khan Ilioudi, M. N. Giannakos et Chorianopoulos 2013, style d'interview et projection vidéo de style instructeur. Un résumé des différents styles de vidéo est fourni ci-dessous.

Le style vidéo tête parlante (*"Talking head style"*) : Ce style est également connu sous le nom de capture vidéo de l'enseignement en classe Ilioudi, M. N. Giannakos et Chorianopoulos 2013 ; Ozan et Ozarslan 2016. C'est le style vidéo le plus couramment utilisé dans lequel on voit la tête du professeur expliquer le cours avec ou sans diaporama en arrière-plan. Les exemples les plus connus de ce style sont iTunes U, MIT Open Courseware.

Le style vidéo présentation (*"Presentation style"*) : Voici le style vidéo dans lequel le présentateur et les diapositives de présentation sont montrés en même temps Ozan et Ozarslan 2016.

Le style vidéo image dans l'image (*"Picture-in-picture video style"*) : C'est le style de vidéo où l'on a deux écrans qui s'affichent en même temps. Un écran principal affichant la personne qui explique et un autre petit écran montrant ce qui est expliqué Chorianopoulos et M. N. Giannakos 2013.

Le style vidéo "Interview" (*"Interview style"*) : Le style d'interview est le style dans lequel un intervieweur pose des questions à l'expert en la matière. Il n'est pas important que l'intervieweur soit un expert du domaine. Il peut s'agir simplement d'un étudiant qui pose diverses questions à un expert. L'intervieweur peut aussi être l'instructeur du MOOC qui interroge un expert ou un invité ayant une perspective unique sur le sujet.

Le style vidéo de la voix sur une présentation (*"Voice over presentation style"*) : Généralement, les étudiants reçoivent deux fichiers séparés : l'un est le fichier de présentation PowerPoint et l'autre est le podcast audio MP3. Le commentaire audio n'est pas nécessairement synchronisé avec les transitions des diapositives PowerPoint. Cet effort de synchronisation sera l'effort personnel de l'apprenant Griffin, D. Mitchell et S. J. Thompson 2009.

Le style vidéo "Khan" (*"Khan Style"*) : Ce style vidéo est également connu sous le nom de stylo instructeur. C'est la vidéo qui ne représente que le stylo ou la main de l'instructeur et non le visage Chorianopoulos et M. N. Giannakos 2013. Il s'agit d'un gros plan d'une planche à dessin interactive avec capture vidéo en voix off. Le meilleur exemple est celui d'Udacity et de l'académie Khan qui fournissent une capture vidéo de la planche à dessin avec une simulation de l'écriture de la main par l'instructeur Ilioudi, M. N. Giannakos et Chorianopoulos 2013.

Le style vidéo "Screencast" (*"Screencast style"*) : Ce style est l'enregistrement d'un écran d'ordinateur ou de tablette avec voix off expliquant le contenu de l'écran montré. C'est un moyen de narrer une promenade vidéo pour aider les étudiants à comprendre un sujet ou à trouver une chose. Cela permet d'expliquer des choses sans apparaître sur la vidéo.

ANNEXE F DESCRIPTION DES CLASSIFICATEURS

Support de Vecteur Machine (*Suport Vector Machine*)

Le Support de Vecteur Machine (SVM) est une procédure d'apprentissage machine supervisée qui peut être utilisée pour la classification ou la régression. Habituellement, cet algorithme est utilisé dans des problèmes de classification utilisant un tracé spatial à n dimensions (n étant le nombre de caractéristiques dans les données). Le problème de classification consistait à trouver l'hyperplan qui différencie chaque classe des autres. Les avantages de cet algorithme sont sa capacité à avoir une séparation nette des marges, à être efficace dans les espaces à hautes dimensions, en particulier lorsque le nombre de dimensions est supérieur au nombre d'échantillons, et il est efficace sur le plan de la mémoire car il utilise les points du sous-ensemble d'entraînement dans la fonction de décision. Les limites de cet algorithme résident dans ses mauvaises performances sur de grandes données, car l'apprentissage prend du temps et, en cas de bruits dans les données, il ne fonctionne pas bien. Enfin, le SVM ne fournit pas de probabilité directe et doit être calculé à l'aide d'une validation croisée quintuple coûteuse. Pour mettre en œuvre cette méthode, nous avons utilisé les publications de Kuhn 2015 ; Kuhn et al. 2008 ; Kuhn et Johnson 2013 sur la construction de modèles prédictifs en utilisant le paquet caret dans R.

Pour notre ensemble d'entraînement afin d'adapter le modèle, nous avons utilisé la méthode svmRadial avec 5 comme longueur d'accord. En appelant la fonction $train()$ du paquet caret de R, elle construit divers modèles de prédiction donnant des valeurs sigma, le coût, la précision et kappa. Ensuite, on sélectionne le meilleur modèle pour prédire l'ensemble test.

L'arbre de décision (*Gradian Boosted Machine*)

L'arbre de décision (GBM) est utilisé dans le contexte de l'apprentissage supervisé où les données de formation (avec certaines caractéristiques) X_i pour prédire la variable cible Y_i. Comme problème général de l'apprentissage supervisé, la cible peut être fonction de la fonction objectif avec deux parties : la perte de formation et la régularisation. Pour cette implémentation, nous avons utilisé le package *gbm* (Generalized Boosted Models) dans R expliqué par Ridgeway 2007. Ce paquet *gbm* utilise la fonction de perte exponentielle d'AdaBoost Freund et Schapire 1995 et l'algorithme de descente en gradient de Friedman 2001 ; Friedman 2002. Comme décrit par Ridgeway 2007 voici les paramètres que nous avons

utilisés :

- Une fonction de perte : *preProcess*() fonction du paquet *gbm* pour déterminer les valeurs des transformations des prédicteurs en utilisant l'ensemble de formation et peut être appliquée dans le futur à l'ensemble de test.
- Nombre d'itérations : 400, n.arbres : (1 :10)*25
- La profondeur de chaque arbre : (1 :5)*5
- Le rétrécissement (ou taux d'apprentissage) : 0,5

L'algorithme a suivi la description de Ridgeway 2007 qui est la même approche qu'un arbre unique, mais qui additionne l'importance de chaque itération de boosting.

Le Voisin le plus proche (*K Nearest Neighbor*)

Le plus proche voisin (KNN) est une méthode de classification des modèles utilisant des vecteurs de n dimensions dans un espace euclidien. Tous les points sont classés en fonction de leur proximité euclidienne dans l'espace. En cas de connaissance préalable du nombre de classes à classer, il faut organiser les données en fonction du nombre de classes regroupant les points les plus proches. Dans le cas d'un nombre non prédéfini de classes (ou catégories), chaque point est candidat à devenir le centre d'une classe, alors la recherche de points de fermeture pourrait définir une classe en utilisant la fonction discriminante du plus proche voisin définie comme : Pour chaque vecteur $x = (x_1, x_2, ..., x_n)$ (avec n la dimension des vecteurs) représentant le vecteur séquence centroïde d'un style vidéo et p_i les autres vecteurs pour les autres centroïdes ($i = 1, 2, 3, ...$).

$$g(x) = min(d(x, p_1), d(x, p_2), ..., d(x, p_k)) \tag{F.1}$$

Où $d(x, p_i)$ est la distance euclidienne entre x et P_i, $g(x)$ donne la distance au plus proche voisin de x dans l'espace euclidien. Lorsqu'il faut prendre le k plus proche voisin de x, il faut calculer k fois $g(x)$ en enlevant à chaque fois le plus proche voisin de la liste des points avant de recalculer $g(x)$.
Le style vidéo de x est alors prédit dans notre cas selon le vote majoritaire des 11 plus proches voisins du style x. Ici 11 parce que nous prenons l'impair le plus proche de la racine carrée du nombre total d'enregistrements (ici 100 enregistrements).

Forêt aléatoire (*Random Forest*)

Une forêt aléatoire (*Random Forest*) est un classificateur constitué d'un ensemble d'arbres de décision arborescents $h(x, \Theta_k), k = 1, ...,$ où les Θ_k sont des vecteurs aléatoires indépendants et identiquement distribués et où chaque arbre émet un vote unitaire pour la classe la plus populaire à l'entrée x.

En général, un nombre important d'arbres sont générés et un vote pour la classe la plus populaire pour chaque élément à classifier.

Le terme provient de forêt de décision aléatoire qui a été proposé pour la première fois par Ho 1995 de Bell Labs en 1995 .

Chaque arbre est construit à l'aide de l'algorithme suivant :

1. Soit le nombre de cas dans l'ensemble d'entrainement est de N, et le nombre de variables dans le classificateur est de M. On nous dit que le nombre m de variables d'entrée à utiliser pour déterminer la décision à un nœud de l'arbre ; m doit être bien inférieur à M.
2. Choisir un ensemble d'entrainement pour cet arbre en choisissant n fois avec remplacement parmi tous les N cas d'entrainement disponibles (c'est-à-dire prendre un échantillon bootstrap). Utilisez le reste des cas pour estimer l'erreur de l'arbre, en prédisant leurs classes.
3. Pour chaque nœud de l'arbre, choisir au hasard m variables sur lesquelles fonder la décision à ce nœud. Calculer la meilleure répartition sur la base de ces m variables dans l'ensemble d'entrainement. Chaque arbre est entièrement développé et non élagué (comme on peut le faire pour construire un classificateur d'arbre normal).
4. Pour la prédiction, un nouvel échantillon est poussé en bas de l'arbre. On lui attribue l'étiquette de l'échantillon d'apprentissage dans le nœud terminal où il aboutit. Cette procédure est répétée sur tous les arbres de l'ensemble, et le vote moyen de tous les arbres est rapporté comme une prédiction forestière aléatoire.

Kappa

La valeur Kappa est une mesure de la concordance des données catégorielles lorsque les classes sont fortement déséquilibrées et mesure la concordance par rapport à ce à quoi on devrait s'attendre par hasard. La valeur de Kappa doit être proche de zéro. Pour le SVM et la GBM, elle utilise un paramètre de rotation qui est la fonction d'échelle σ. Pour l'estimation de la valeur σ à partir des données d'entraînement, nous avons utilisé la méthode analytique décrite par Caputo et al. 2001 et par défaut la fonction de train utilise sigest dans les pa-

quets kernlab développés par KARATZOGLOU et al. 2004. Ensuite, kappa s'exprime comme suit :

$$K(a,b) = exp(-\sigma||a-b||^2) \tag{F.2}$$

Où a est la valeur prévue et b la valeur attendue par hasard.
Pour la fonction train, la méthode de rééchantillonnage utilisée est la validation croisée k-fold. En général, 25 itérations de bootstrap sont utilisées, mais en raison du grand nombre d'échantillons dans l'ensemble d'entraînement, nous avons augmenté les itérations à 400.

ANNEXE G MÉTHODES DE CALCUL DES DISTANCES ENTRE LES SÉQUENCES

La similarité entre deux séquences est définie par la distance qui les sépare. Plus cette distance est faible, plus les deux séquences sont proches. La distance est déterminée par le nombre d'insertions, de suppressions et de substitutions pour transformer une séquence en une autre. Pour calculer les similarités entre les séquences, trois algorithmes différents peuvent généralement être utilisés.

Préfixe commun le plus long : LCP (*Longuest Common Prefix*)

Cet algorithme a été proposé par Elzinga 2006 et vise à mesurer la similarité/distance entre les séquences. Il est basé sur la longueur du préfixe commun le plus long (LCP). La formule pour calculer la distance entre deux séquences x et y est donnée par :

$$d_p(x, y) = |x| + |y| - 2|P(x, y)| \tag{G.1}$$

Où :
|x| est la longueur de la séquence x.
P(x,y) est le préfixe le plus long de x et y.

Pour les besoins de cette étude, nous avons essayé cet algorithme et avec un bref exemple de son application, on peut constater qu'il n'est pas adapté à la comparaison de la similarité entre deux séquences entières. Voici l'exemple :

Séquence 1 : P-P-P-P-S-S-P-s-s-P
Séquence 2 : P-P-P-S-P-P-s-p-P-p
Séquence 3 : P-P-p-P-S-P-P-s-S-P

LCP (Séquence 1, Séquence 2) = 14
LCP (Séquence 1, Séquence 3) = 16

Par observation de la similitude entre ces trois séquences, on voit que la séquence 1 est plus proche de la séquence 3 que de la séquence 2. L'algorithme LCP ne peut donc pas être utilisé pour la détermination des similarités tel que voulu dans cette recherche.

La plus longue séquence commune : LCS (*Longest Common Subsequence*)

Le LCS est un autre algorithme des métriques considérées par Elzinga 2006 qui est disponible par la fonction seqdist() de R et calcule la longueur de la plus longue sous-séquence entre deux séquences. La formule utilisée ici pour la distance est la suivante :

$$d_l(x, y) = |x| + |y| - 2A_l(x, y) \tag{G.2}$$

Où :
$|x|$ est la longueur de la séquence x.
$A_l(x, y)$ est la sous-séquence commune de x et de y.

Pour l'exemple précédent de séquences, nous avons :
LCS (Séquence 1, Séquence 2) = 6
LCS (Séquence 1, Séquence 3) = 6

Comme il a été mentionné précédemment, la séquence 1 est plus proche de la séquence 3 que de la séquence 2. Par conséquent, l'algorithme LCS ne peut pas non plus être utilisé dans le but de regrouper des séquences similaires.

Distance d'appariement optimale : OM (*Optimal Matching distance*)

Nous avons utilisé l'algorithme d'appariement optimal (OM) qui génère des distances qui représentent le coût minimal en termes d'insertions, de suppressions et de substitutions pour transformer une séquence en une autre. Le coût de chaque suppression ou insertion est de 1 par défaut et celui de la substitution est de 2. Cet algorithme a été proposé à l'origine par Levenshtein 1966 et a été popularisé dans les sciences sociales par Forrest 1990 et Abbott 2001. L'algorithme mis en œuvre dans TraMineR est celui de Needleman 1970. Le résultat est une matrice de coût de substitution qui donne la distance entre toutes les séquences de la liste. La formule pour obtenir la matrice des coûts est :

$$d_{OM}(x_i, y_j) = \begin{cases} 0, & \text{if } i = 0 \; ou \; j = 0 \\ L(x_i, y_j), & \text{if } i, j > 0 \; et \; x_i = y_j \\ L(x_i, y_j) + 1, & \text{if } i, j > 0 \; et \; x_i \neq y_j \end{cases} \tag{G.3}$$

Où :
$L(x_i, y_j) = \{d_{OM}(x_{i-1}, y_j), d_{OM}(x_i, y_{j-1}), d_{OM}(x_{i-1}, y_{j-1})\}$

Et l'OM nécessite la matrice des coûts de substitution où w_{ij} représente la substitution de coût du symbole tel que :

$$w_{ij} = \begin{cases} 0, & \text{if } i = j \\ 1, & \text{if } i \neq j \ et \ i = 1 \ et \ j = 1 \\ 2, & \text{if } i \neq j \ et \ i, j > 1 \end{cases} \tag{G.4}$$

Pour les besoins de l'étude dans cette thèse, nous avons adapté l'algorithme OM afin que toute substitution, suppression ou insertion ait le même coût (1 par défaut). La distance entre deux séquences devient alors le nombre de substitutions, de suppressions ou d'insertions nécessaires pour transformer une séquence en l'autre. En ce sens, la distance entre la séquence A et B est la même que la distance entre la séquence B et A. C'est donc en utilisant ces distances que l'on considère que deux séquences sont suffisamment proches pour être regroupées ou non.

ANNEXE H MÉTHODOLOGIE DE SIMILARITÉ POUR ANALYSE DE TEXTE

Analyse des transcrits des vidéos

Les transcrits des vidéos des témoignages des 15 migrants montrent une diversité de taille dans l'usage du vocabulaire. Nous obtenons des textes allant de 205 à plus de 2500 mots répartis comme le montre la table H.1 pour les 15 vidéos d'interviews des migrants.

vidéo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Nbre Mots	565	1573	437	807	205	550	538	1515	2018	2594	1019	1484	571	734	596

Tableau H.1 Composition de la taille du vocabulaire des 15 migrants de l'échantillon

L'expression lexicale de ces migrants semble être faible sur le plan du vocabulaire : on observe que seulement 606 mots distincts sont utilisés communément par ces 15 migrants. On appelle mot commun un mot ou une expression utilisée par au moins 3 migrants dans leur intervention. De cette liste des mots fréquents, sont extraits tous les mots de connexion : les articles, les conjonctions de coordination, les ponctuations. Ces éléments ont été supprimés au départ de l'analyse des textes.

Chaque intervention (témoignage) est représentée sous forme d'un vecteur d'une longueur de 606 mots communs aux migrants qui constituent les paramètres des colonnes du vecteur.

Le transcrit de chaque intervenant est donc représenté par une ligne de 606 colonnes représentant les 606 mots possibles que l'intervenant a pu utiliser et le nombre de fois dans son discours. Nous obtenons ainsi une matrice de 15 par 606 représentants tous les transcrits des 15 intervenants. Lorsqu'un mot est présent dans l'intervention d'un migrant, il est représenté par le nombre de son apparition sinon par zéro (0) dans le vecteur qui le représente.

De cette matrice l'on peut donc extraire les mots les plus utilisés par les migrants dans leur discours. Il s'agit en l'occurrence de calculer le nombre total de fois que chaque mot a été utilisé : ici la somme de chaque colonne. On extrait ainsi les mots les plus fréquents. Sont considérés comme mots les plus fréquents, ceux qui apparaissent au moins dans 7/15 du discours des migrants.

Distance entre les textes

Comme chaque texte est représenté par un vecteur de 606 éléments fait des nombres d'apparitions selon que le mot existe ou pas dans le texte.

Nous obtenons la matrice de distance entre les discours à la table H.2 qui montre la distance entre eux en utilisant la distance euclidienne entre les vecteurs.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	170	48	218	97	97	145	218	315	364	242	291	145	170	145
2	170	0	170	194	218	170	170	339	291	242	218	218	218	242	218
3	48	170	0	218	145	145	194	267	315	364	242	291	145	170	194
4	218	194	218	0	121	218	218	242	291	339	218	315	218	242	218
5	97	218	145	121	0	97	145	218	364	364	242	291	145	218	145
6	97	170	145	218	97	0	97	218	315	315	194	194	145	170	145
7	145	170	194	218	145	97	0	218	267	267	291	194	145	218	145
8	218	339	267	242	218	218	218	0	291	339	267	267	267	242	315
9	315	291	315	291	364	315	267	291	0	291	315	267	267	339	267
10	364	242	364	339	364	315	267	339	291	0	267	218	364	339	315
11	242	218	242	218	242	194	291	267	315	267	0	291	291	121	291
12	291	218	291	315	291	194	194	267	267	218	291	0	242	267	194
13	145	218	145	218	145	145	145	267	267	364	291	242	0	218	194
14	170	242	170	242	218	170	218	242	339	339	121	267	218	0	170
15	145	218	194	218	145	145	145	315	267	315	291	194	194	170	0

Tableau H.2 La distance euclidienne entre les textes.

Pour normaliser la distance entre les discours, toutes les distances entre les discours, l'on divise par 606 en suivant les équations 6.9 dans la section 6.3.2 de la méthodologie de similarité.

Degré de similarité entre des textes

Nous obtenons ainsi une matrice de similarité entre les 15 discours de migrants nous montrant leur degré de similarité. Nous pouvons avoir ce degré de similarité en termes de pourcentage de la table H.3 ou visuelle ou la similarité est plus grande au fur et à mesure de degré de la couleur de la ressemblance comme le montre la figure H.1. Les calculs qui ont abouti à la détermination du degré de similarité nous viennent de l'équation 6.12 dans la section 6.3.3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00	0.68	0.91	0.59	0.82	0.82	0.73	0.59	0.41	0.32	0.55	0.45	0.73	0.68	0.73
2	0.68	1.00	0.68	0.64	0.59	0.68	0.68	0.36	0.45	0.55	0.59	0.59	0.59	0.55	0.59
3	0.91	0.68	1.00	0.59	0.73	0.73	0.64	0.50	0.41	0.32	0.55	0.45	0.73	0.68	0.64
4	0.59	0.64	0.59	1.00	0.77	0.59	0.59	0.55	0.45	0.36	0.59	0.41	0.59	0.55	0.59
5	0.82	0.59	0.73	0.77	1.00	0.82	0.73	0.59	0.32	0.32	0.55	0.45	0.73	0.59	0.73
6	0.82	0.68	0.73	0.59	0.82	1.00	0.82	0.59	0.41	0.41	0.64	0.64	0.73	0.68	0.73
7	0.73	0.68	0.64	0.59	0.73	0.82	1.00	0.59	0.50	0.50	0.45	0.64	0.73	0.59	0.73
8	0.59	0.36	0.50	0.55	0.59	0.59	0.59	1.00	0.45	0.36	0.50	0.50	0.50	0.55	0.41
9	0.41	0.45	0.41	0.45	0.32	0.41	0.50	0.45	1.00	0.45	0.41	0.50	0.50	0.36	0.50
10	0.32	0.55	0.32	0.36	0.32	0.41	0.50	0.36	0.45	1.00	0.50	0.59	0.32	0.36	0.41
11	0.55	0.59	0.55	0.59	0.55	0.64	0.45	0.50	0.41	0.50	1.00	0.45	0.45	0.77	0.45
12	0.45	0.59	0.45	0.41	0.45	0.64	0.64	0.50	0.50	0.59	0.45	1.00	0.55	0.50	0.64
13	0.73	0.59	0.73	0.59	0.73	0.73	0.73	0.50	0.50	0.32	0.45	0.55	1.00	0.59	0.64
14	0.68	0.55	0.68	0.55	0.59	0.68	0.59	0.55	0.36	0.36	0.77	0.50	0.59	1.00	0.68
15	0.73	0.59	0.64	0.59	0.73	0.73	0.73	0.41	0.50	0.41	0.45	0.64	0.64	0.68	1.00

Tableau H.3 La similarité entre les discours exprimée en termes de pourcentage de ressemblance. Par exemple 0.55 signifie 55% de ressemblance et 1.0 signifie 100% de ressemblance.

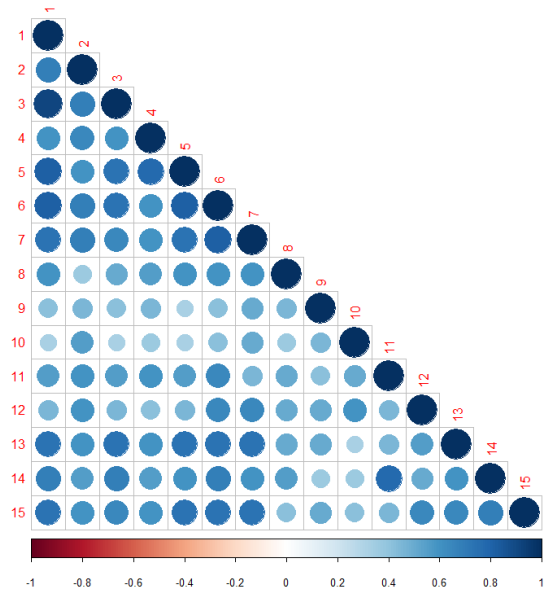


Figure H.1 La similarité entre les discours de manière visuelle.