

ESTIMATING THE DESTINATION OF UNLINKED TRIPS IN PUBLIC TRANSPORTATION SMART CARD FARE COLLECTION SYSTEMS

Li He, Masters student
Polytechnique Montréal
Department of Mathematics and Industrial Engineering
P.O. box 6079, Station Centre-Ville, Montreal (Quebec), CANADA H3C 3A7
Tel.: (514) 340-4711 ext. 4911, Fax: (514) 340-4173
Email: li.he@polymtl.ca

Martin Trépanier, Ph.D., P.Eng., Full Professor
Polytechnique Montréal
Department of Mathematics and Industrial Engineering
P.O. box 6079, Station Centre-Ville, Montreal (Quebec), CANADA H3C 3A7
Tel.: (514) 340-4711 ext. 4911, Fax: (514) 340-4173
Email: mtrepanier@polymtl.ca

First Draft:
August 1st, 2014

Revision History:
October 20th, 2014

Word Count:
Text (including references): 4534 + 9 figures / tables X 250 = 6784 words

94th Annual Meeting of the Transportation Research Board

ABSTRACT

Smart card automated fare collection (SCAFC) systems have proved effective in collecting data on the travel behavior of users of public transit networks. Since some systems record only the boarding locations, we need a method for estimating destinations. Existing algorithms can find an estimate for most trips. However, unlinked trips, which are trips that are not part of a trip chain during the day, are more difficult to analyze. We present an improvement to the existing model for estimating destinations, in particular for unlinked trips. The method is based on kernel density estimation of the spatial and temporal probabilities of each destination. The data used for this study comes from the Société de transport de l'Outaouais, a mid-size bus service near Ottawa, Canada (for a one-month period in 2009 with 908,303 transactions). Existing algorithms can handle only 80.64% of the trips; our method handles an additional 10.9%. We present an analysis of these results and discuss future research directions.

Key words: public transit, smart card, trip destination

INTRODUCTION

The smart card is versatile and powerful and therefore one of the most exciting technological advances in public transport. Integrated into an electronic fare-collection system, the card allows the continuous collection of individualized transactional data on the use of public transport networks (1, 2). Hence, smart card systems data can be used to analyze the travel behavior of users on a larger scale than surveys permit (3) and can observe this behavior for long periods (4).

In some (*tap-in*) systems, information about the boarding location only is available, because users do not validate (*tap-out*) when they leave the vehicles or stations. Therefore, the planners must derive the possible destinations. The destination is needed for the origin-destination matrices that are used in most public-transit network planning models. Some researchers have proposed algorithms that estimate the destination (5, 6); we discuss these in the next section. These algorithms rely on the chaining of trips during the day, based on the reasonable hypothesis that users will alight at a stop near the next boarding stop. However, some destinations cannot be estimated with this approach because they are not linked to the sequence. These trips are “alone” (unlinked) within the day or they do not satisfy the criteria used by the approach. We propose an extension to the existing algorithms to address these unlinked trips. The aim is to find more destinations, therefore improving the usability of the origin-destination matrices. We use a kernel density estimation to better calculate the probabilities of the destinations of these unlinked trips.

The remainder of this paper is organized as follows. The background section emphasizes the importance of smart-card data use in public transport planning and presents the existing destination-estimation methods. The methodology section presents the information system and describes our algorithm and the concepts underlying our approach. The experiments section analyzes the results of our experiments, which are based on one month of data from a Canadian public transport operator. The conclusion presents further discussion and possibilities for future research.

BACKGROUND

Smart-card data for public transport planning

The operation of a smart card automated fare collection (SCAFC) system is straightforward. When the user enters a vehicle (say a bus), he validates his smart card on a card reader. The card checks if the fare is valid and records entry information on the card. At the end of the day, the data is downloaded from the bus to a central server. This server collects all the validations made in the system (2).

The availability of SCAFC data varies between systems. It is categorized as follows (1): 1) Spatial information (boarding locations, alighting locations, duration, route). 2) Time information (date and time of boarding, date and time of alighting). 3) Structural information (transportation mode such as bus or subway, service details such as line and direction). 4) Personal information about the user: name, gender, age, address, and car ownership. 5) Purchase information: type of ticket, ticket price, purchase location (this location may be different from the boarding location).

The data from the smart card can characterize each of the trips made within the network in a totally disaggregated manner (3). However, as in any information system, there can be errors. Chu et al. (7) developed a procedure to detect suspect values in the data using spatiotemporal logic and the operational concepts of public transit.

Use of data

Smart-card data has a wide range of uses (8). The following list is not exhaustive:

- customer clustering and classification using data mining (9);
- identification of purpose of trips (10);
- investigation of temporal (10) and spatial (11) distribution of transactions;
- calibration of hyperpath choice models (12);
- development of origin-destination matrices (13, 14);
- estimation of waiting and walking times in a subway system (15);
- determination of peak periods (16);
- modeling of loyalty and retention rates of card holders (17);
- estimation of load profile of vehicles (18) and other network performance measures (19);
- linking of commercial transactions made with smart cards to assess card-holders' use of grocery stores (20);
- assessment of impact of weather on transit operations (21);
- investigation of effects of reliability of travel time (22).

Some of these studies need the trip destinations (10, 12, 13, 14, 18, 19, 20). For tap-in networks, the destination must be estimated.

Destination-estimation algorithms

The classical destination-estimation algorithm distinguishes four types of trips (5):

- Trip sequence. This assumes that users alight at the stop where the distance between the current alighting location and the next boarding location is minimal. This distance must be lower than a specified maximum value or the trip is rejected.
- Last trip of day (return home). Users return to the origin of the first trip, which is assumed to be the stop closest to their home.
- First trip of next day. If the above two steps are not successful, we use the first trip of the next day to find the destination of the last trip in the sequence.
- Unlinked trip. As a last resort, we find a similar transaction in the history of the card, i.e., a transaction with the same boarding location at almost the same time. The classical algorithm uses the first such trip found.

For a sequence of trips, there is another way to determine the destination. We can assume that the time between the current alighting and the next boarding is minimal (6). In this case, there is no maximum value on the distance. Moreover, this approach can detect if an activity rather than a regular transfer occurred at the destination.

For some tap-in/tap-out subway networks, we do not know the location of transfers. Si et al. (23) propose a destination-estimation method based on travel time, travel distance, and the number of transfers. They also categorize users based on their occupation, salary, and purpose of travel, and their destination-estimation accuracy is above 85%. In subway systems, we do not know the exact vehicle taken by the user. Sun and Schonfeld (24) tried to identify the exact time of boarding and alighting. They took into account three factors: 1) the time between the transaction and boarding, 2) the possibility of missing a vehicle, and 3) the schedule for the subway and the theoretical transfer time at stations.

The algorithms of (5) can estimate 80% of the destinations, but unlinked trips (called “unit trips” in (5)) are more difficult to analyze. They are the focus of this paper.

METHODOLOGY

We first present the information system used. Then, we define some base hypotheses for the development of the model. We next give a description of the classical algorithm and then our proposed improvements.

Information system

The research was conducted in collaboration with the Société de transport de l'Outaouais (STO), which is a mid-sized transit authority located in Gatineau, Quebec, across the river from Ottawa, Ontario. Gatineau is a city with a population of 240,000 spread over a large area (589 km²). The STO has 200 buses, and all of them were equipped with SCAFC in 2001 (5). We use the month of October 2009, for which there are 903,833 smart-card transactions. The smart-card records are stored in four tables:

- *Transaction* contains individual information about card transactions: identifier, card number, date, line number, direction, boarding stop.
- *Line-stop* contains information on the stops for a given line in a given direction: line direction, ordering of stops, distance between starting point of line and given stop.
- *Stop* contains information on the stops in the network: stop number, name, latitude, longitude.
- *Line* contains information on the lines in the network: line number, direction, real direction.

Hypotheses

To estimate the destinations for the unlinked trips, we use the destinations in the card history. We assume that the historical locations and times of alighting have an impact on the current trip, because users tend to follow the same routines. Hence, users with unlinked trips may have multiple destinations for their historical trips, so we need a better method to estimate the destinations of these trips. In Figure 1, each layer represents a trip. The current layer can collect information from all the previous layers, including the historical origin (HO) and destination (HD). If a destination is sufficiently close to a potential destination (PD) of the current transaction (part of the sequence of stops of the current route), the probability of choosing this location will be higher.

The same logic applies to the temporal dimension. The similarity between the historical arrival time and the expected arrival time helps to determine the likelihood of alighting at this stop. The spatial probability (the distance between the origin of the current trip and the potential destination) and the temporal probability (the arrival time) together determine the probability of alighting.

We use the spatiality and temporality in a kernel estimation to model the probability of discrete variables using continuous variables. We can superimpose the normal probability distribution of each potential destination. In this way, we can obtain a probability curve for all the potential destinations along the route (called the spatial probability).

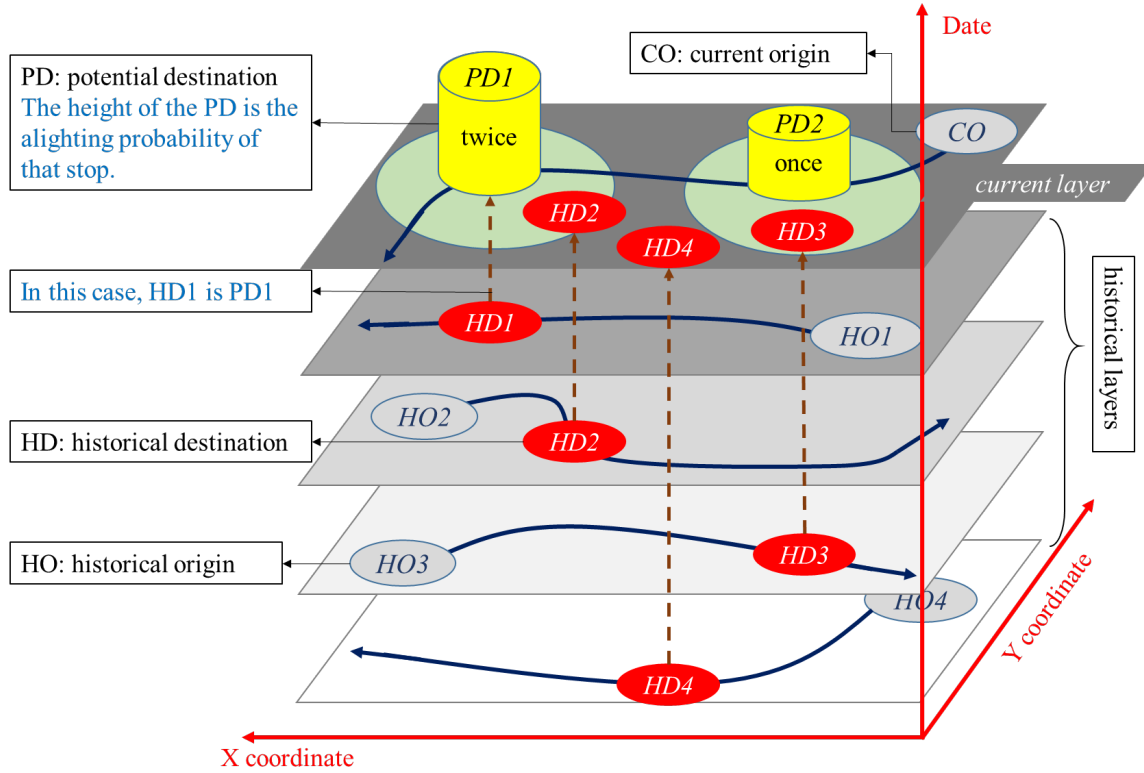


Figure 1: Estimation of the spatial probability of alighting using the historical layers

Figure 2 shows the process of estimating the temporal probability. For every boarding location, we can find several potential destinations for which there are records of historical alighting. For the temporal estimation, we put the time of each transaction along the horizontal axis, which represents the time of day. Then, the alighting instances will form a normal distribution curve, which represents the probability of alighting at a specific time. We can also calculate the planned arrival time for each potential destination. On the alighting probability curve, each planned arrival time corresponds to a probability. The stop with the highest probability is chosen as the estimated destination for the temporal estimation.

We multiply the spatial and temporal probabilities to predict the probability of alighting at a potential stop. The method is thus more likely to choose the correct destination, which should correspond to both the spatial and temporal dimensions of the historical transactions.

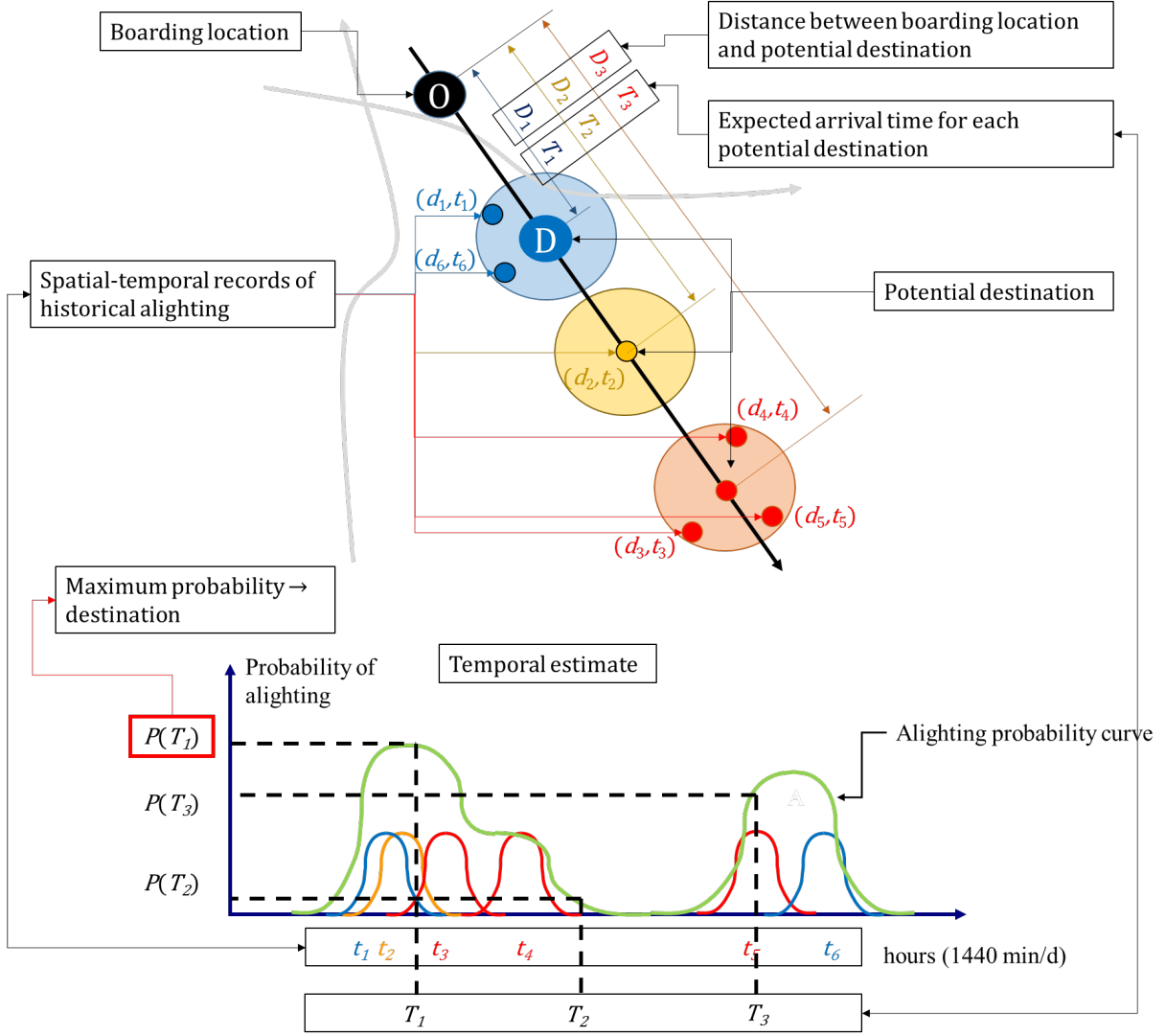


Figure 2: Kernel estimation for temporal processing of unlinked trips

Classical model

Our implementation is based on the classical algorithm proposed by (5). The indices are:

i = transit user (unique smart-card identification)

j = sequence number of transit stop within a route

r = sequence number of route in user's day

k = day

A transit route R (one direction only) is defined by an ordered set (sequence) of stops s :

$$R = \{s^j\}. \quad (1)$$

Then, s_{rik}^j is the j^{th} stop of the r^{th} route of user i for day k . We define the *vanishing route*, V_{rik} , to be the sequence of stops where the user can alight after boarding a given route at stop $j = B$:

$$V_{rik} = \{s_{rik}^j\}, \quad \forall j > B. \quad (2)$$

Each s_{rik}^j is the event where user i , who is on the vehicle on route r , accesses a stop j and time t . The journey J_{ik} of a transit user can be modeled by a sequence of vanishing routes during the day. Let N_k be the number of vanishing routes taken within a single day k . Then

$$J_{ik} = \{V_{rik}^j\}, \quad r = 1 \dots N_k. \quad (3)$$

Let $d(a, b)$ be the Euclidean distance between locations a and b . The model estimates the best location for different values of r in the trip. The goal is to estimate the destination d_{rik} in the vanishing route so that the distance to the boarding stop for the next transaction, $s_{(r+1)ik}^B$, is minimum. We also define a maximum value for the distance (in this case, two kilometers):

$$\begin{aligned} d_{rik} = z &\rightarrow \min_z d(s_{(r+1)ik}^B, z) \quad z \in \{V_{rik}\}, \\ r < N_k &\quad d(s_{(r+1)ik}^B, z) < M_1. \end{aligned} \quad (4)$$

For the last trip of the day, we assume that the distance between the destination of the last transaction and the origin of the first transaction of the day is less than a given value:

$$\begin{aligned} d'_{rik} = z &\rightarrow \min_z d(s_{1ik}^B, z) \quad z \in \{V_{rik}\}, \\ r = N_k &\quad d(s_{1ik}^B, z) < M_1. \end{aligned} \quad (5)$$

If the above assumption does not give a solution, we assume that the distance between the destination of the last transaction of the day and the origin of the first transaction of the next day is less than a given value:

$$\begin{aligned} d''_{rik} = z &\rightarrow \min_z d(s_{1i(k+1)}^B, z) \quad z \in \{V_{rik}\}, \\ r = N_k &\quad d(s_{1i(k+1)}^B, z) < M_1. \end{aligned} \quad (6)$$

Figure 3 illustrates the basic algorithm. We define the *unlinked trips* to be the trips for which no destination can be found using this method.

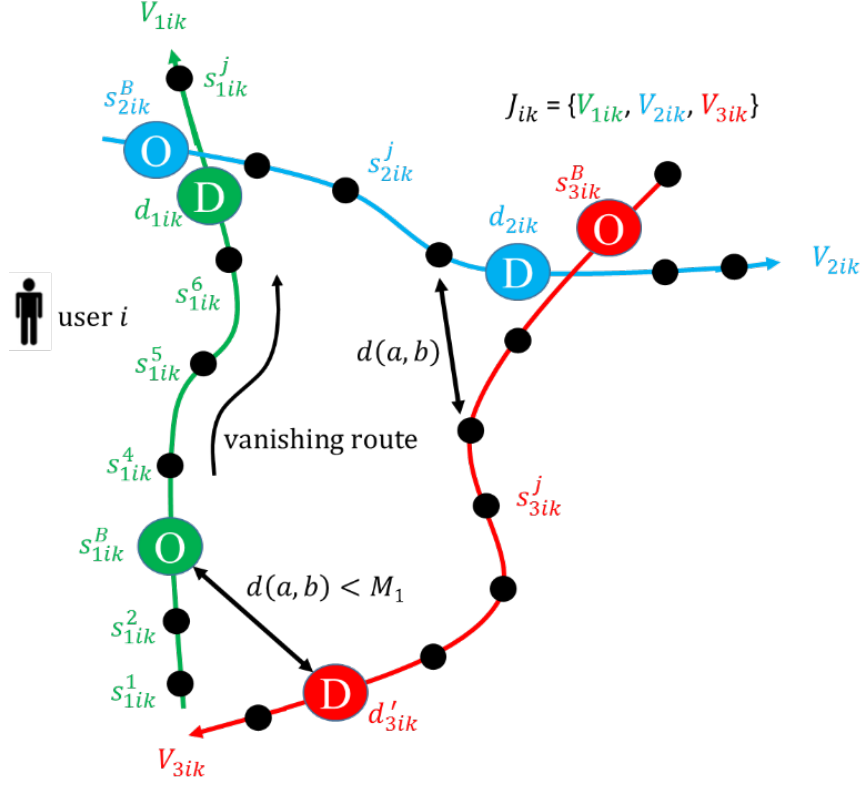


Figure 3: Basic model (inspired by (5))

Proposed improvements

This section presents the improvements that will better estimate the destination of the unlinked trips. We need an additional index:

m = sequence number of historical records.

We find a destination near bus stop j among the historical destinations d_{rik} , d'_{rik} , and d''_{rik} :

$$H_{im}^j = \{d_{rik} \text{ or } d'_{rik} \text{ or } d''_{rik}, \text{ if } d(s_{rik}^B, d_{rik} \text{ or } d'_{rik} \text{ or } d''_{rik}) < M_2\} \quad (7)$$

In this case, if the distance between the historical destination and the potential destination of the current transaction is less than a maximum value (1 km), this historical destination is assumed to have an impact on the current destination.

Let N_m be the number of H_{im}^j of the unlinked trip. Let $E_d(x)$ be the density relative to the distance (spatial). Let $E_t(x)$ be the density relative to the arrival time (temporal). Let $E(x)$ be the general density. The average values of the distance $\overline{d(H_{im}^j, s_{rik}^B)}$ and time $\overline{t(H_{im}^j)}$ are combined to give a parameter named *bandwidth* that governs the degree of smoothness of the estimate. The functions of the *kernel* estimate can now be established. We obtain a probability function $E(x)$ that gives the probability of alighting at each potential stop and time:

$$E_d(x) = \frac{\sum_{m=1}^{N_m} K \left[\frac{x - d(H_{im}^j, s_{rik}^B)}{\overline{d(H_{im}^j, s_{rik}^B)}} \right]}{\sum_{m=1}^{N_m} H_{im}^j \cdot \overline{d(H_{im}^j, s_{rik}^B)}},$$

$$E_t(x) = \frac{\sum_{m=1}^{N_m} K \left[\frac{x - t(H_{im}^j)}{t(H_{im}^j)} \right]}{\sum_{m=1}^{N_m} H_{im}^j \cdot \overline{t(H_{im}^j)}},$$

$$K(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}. \quad (8)$$

In Equation (8), we choose the average of the values (time at the level of spatiality or distance at the level of temporality) as the bandwidth ($\overline{d(H_{im}^j, s_{rik}^B)}$ and $\overline{t(H_{im}^j)}$). This will ensure a relatively smooth curve. In general, the value of the bandwidth function will impact the shape of the probability curve. However, our goal is to find the highest probability from several discrete points (potential destinations), and the value of the bandwidth will have little impact.

We assume that the spatiality and temporality are independent. Our goal is to find the highest probability from several discrete points. We achieve this via multiplication; the specific combination of spatiality and temporality has little impact in practice. This leads to

$$E(x) = E_d(x) \cdot E_t(x). \quad (9)$$

For the current transaction, we can establish a function (see Figure 4) to predict the arrival time at each stop, using the commercial speed v_c :

$$d(z, s_{rik}^B) = v_c \cdot [t(z) - t(s_{rik}^B)]. \quad (10)$$

Given the expected alighting distance and time, we can use the density function to calculate the probability of alighting at each potential destination P_{rik}^j :

$$P_{rik}^j = E_d[d(z, s_{rik}^B)] \cdot E_t[t(z)]. \quad (11)$$

We can now choose the potential destination with the highest probability, d_{rik}''' :

$$d_{rik}''' = z \rightarrow \max_z P_{rik}^j \quad z \in \{V_{rik}\},$$

$$s_{rik}^B = \text{start}(d_{rik}'''). \quad (12)$$

Table 1 shows an example in which there are two potential destinations, 5040 and 5016. For each historical time or distance record, we generate a kernel function. We then calculate the sum of the kernel density for the spatiality ($\sum K_s$) and the temporality ($\sum K_t$). Note that the time is expressed in minutes (e.g., 07:01=7×60+1=421). When we calculate the alighting probability, we remove the denominators $\sum_{m=1}^{N_m} H_{im}^j \cdot \overline{d(H_{im}^j, s_{rik}^B)}$ and $\sum_{m=1}^{N_m} H_{im}^j \cdot \overline{t(H_{im}^j)}$, because these constant factors do not affect the comparison. We also impose that the sum of the probability of all the potential destinations equal 100%. For the two potential destinations, the higher value is 83.38%, so we choose stop 5040 as the destination.

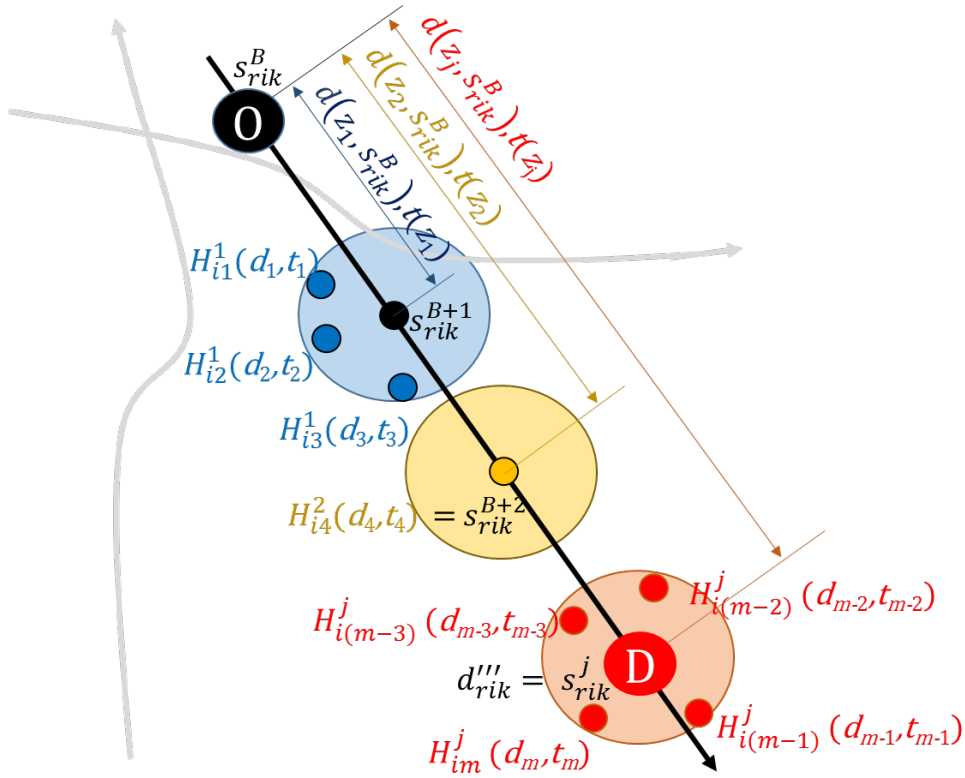


Figure 4: General processing method for unlinked trips

	Historical destination	Historical distance (m)	Spatiality kernel function	Predicted arrival distance	
				Stop 5040	Stop 5016
Spatiality	5040	22287	$K[\frac{x - 22287}{22691.67}]$	$x_1 = 22287$	$x_2 = 23501$
	5040	22287	$K[\frac{x - 22287}{22691.67}]$	$\sum K_s = 0.9559$	$\sum K_s = 0.7167$
	5016	23501	$K[\frac{x - 23501}{22691.67}]$		
Temporality	Historical destination	Historical time	Temporality kernel function	Predicted arrival time	
	5040	7:46 (466)	$K[\frac{x - 466}{464.67}]$	Stop 5040	Stop 5016
				$x_1 = 7:47(467)$	$x_2 = 7:51(471)$
				$\sum K_t = 1.1968$	$\sum K_t = 0.3182$
	5016	7:46 (466)	$K[\frac{x - 466}{464.67}]$		
Alighting probability		$\sum K_s \cdot \sum K_t$ $\sum (\sum K_s \cdot \sum K_t) = 100\%$		Stop 5040	Stop 5016
				83.38% > 16.62%	

Table 1: Numerical example of kernel density estimation

EXPERIMENTS

The improved model has been tested on data for October 2009 from STO. In this section, we use the following codes:

- Code 11: Destination found in the *trip sequence* phase.
- Code 12: Destination found in the *last trip of day* phase (also called *return home*).
- Code 13: Destination found in the *first trip of next day* phase.
- Code 21: Destination found with the improved method, where the unlinked trip has several potential destinations.
- Code 22: Destination found with the improved method, where the unlinked trip has only one potential destination.
- Code 30: No destination found.

One of the above codes is assigned to every smart-card trip of the month.

Overall Results

Figure 5 shows the overall distribution of the codes. The classical algorithm can estimate 80.64% of the destinations. Our improved algorithm can estimate approximately a further 11%, covering 56.30% of those not estimated by the classical algorithm. Just 8.45% of the trips remain unresolved.

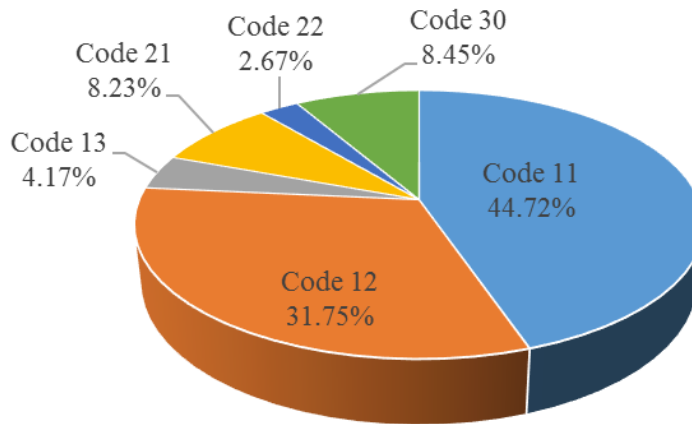


Figure 5: Distribution of codes

Reasons for the nonresolution include: 1) an erroneous route or stop number in the data, 2) an erroneous combination of route and stop, and 3) a lack of historical records. Inference methods, as discussed in the conclusion, may help to resolve these trips.

Analysis by fare type

Different card types usually represent different travel patterns. It is therefore interesting to analyze the results by card type. In Figure 6, the size of each bubble represents the number of corresponding trips. Adults with a regular card represent the largest ridership, especially for codes 11 and 12. For students, codes 11 and 30 dominate; this shows that students have irregular travel. This is also demonstrated by the similarity between codes 21 and 22.

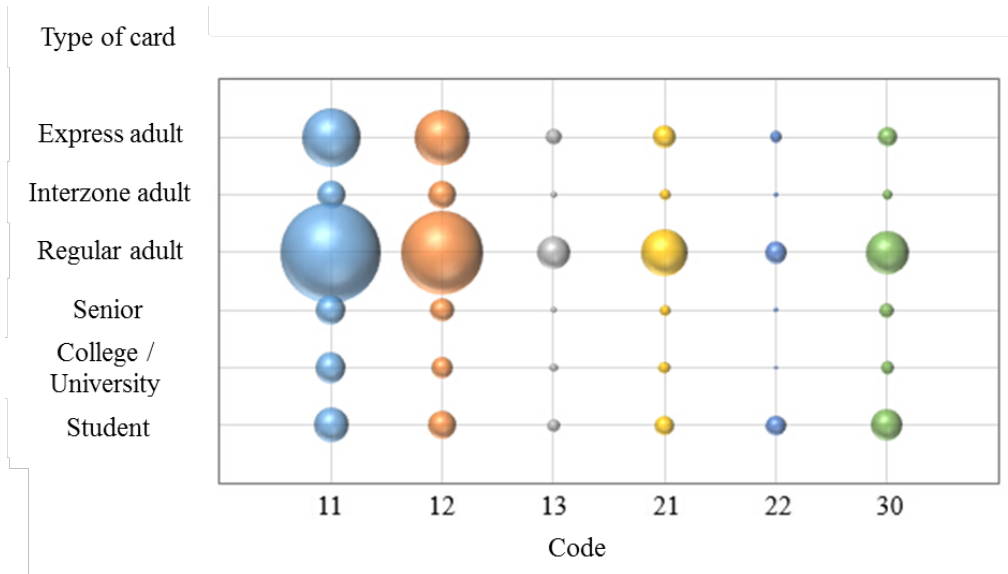


Figure 6: Analysis of estimates by fare type

Analysis by route

The distribution of the codes varies among the bus routes; Figure 7 shows the percentages for a few routes. The similarity of the codes 11 and 12 indicates the importance of transporting commuters between the suburbs and the city center during peak hours. We can therefore infer that routes 35, 39, 40, and 41 connect the suburbs and the city center, and that routes 36, 37, and 38 provide more varied service.

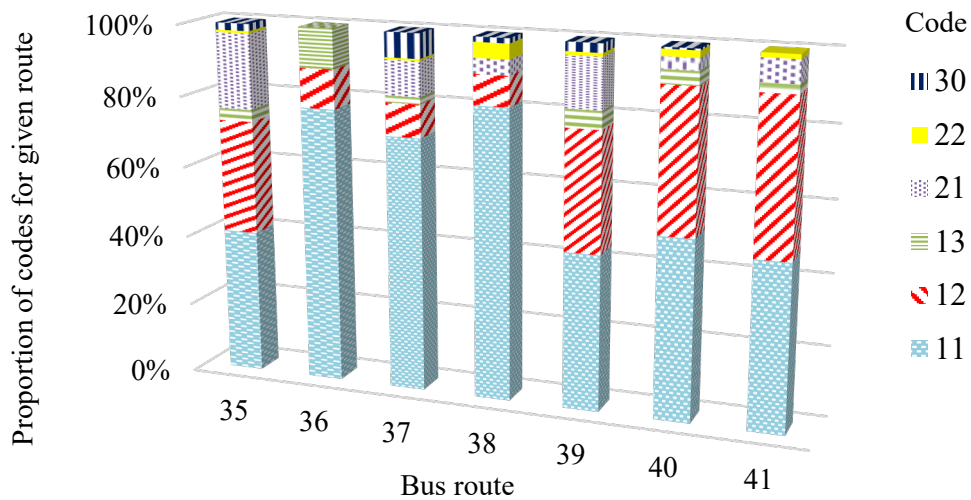


Figure 7: Analysis of estimates by route

Temporal Analysis

Because the algorithm is based on trip sequences, the time of day greatly influences the estimation methods. According to Figure 8, the number of results for code 11 decreases later in the day. The number of results for code 12 begins to increase from 10:00 and the peak hour occurs from 16:00 to 17:00. Code 21 is highest in the morning peak hour and in the evening (after 20:00).

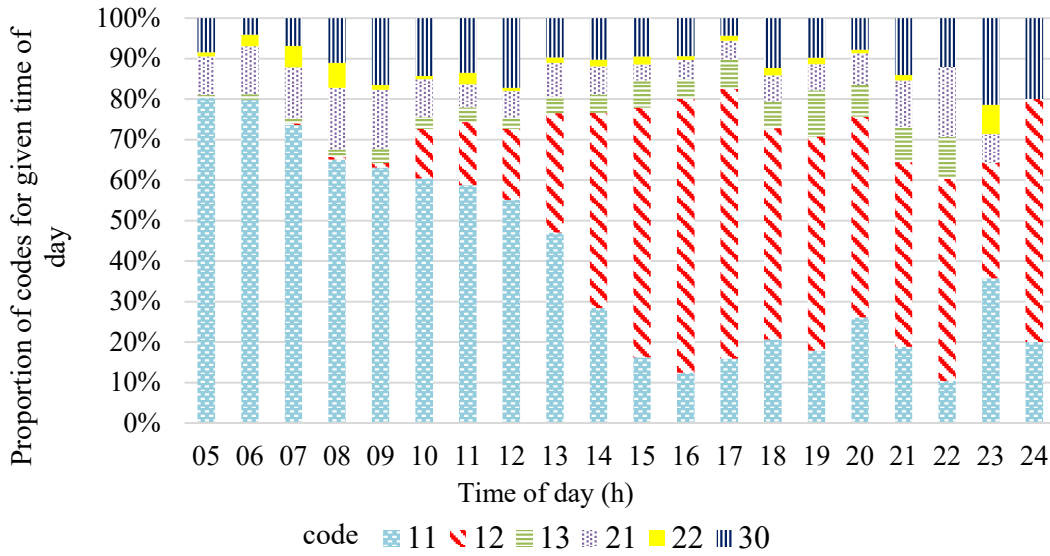


Figure 8: Analysis of estimates by time of day

It is also interesting to analyze the regularity over the week. According to Figure 9, on Saturday, the proportion of code 12 decreases and the proportion of code 30 increases. This indicates that the weekend activities of the users are less regular. On Sunday, the number of results for code 13 increases. This suggests that some users are away from their usual residence on Saturday nights, and it is easier to use their location on Monday morning to determine the destination of their last weekend trip.

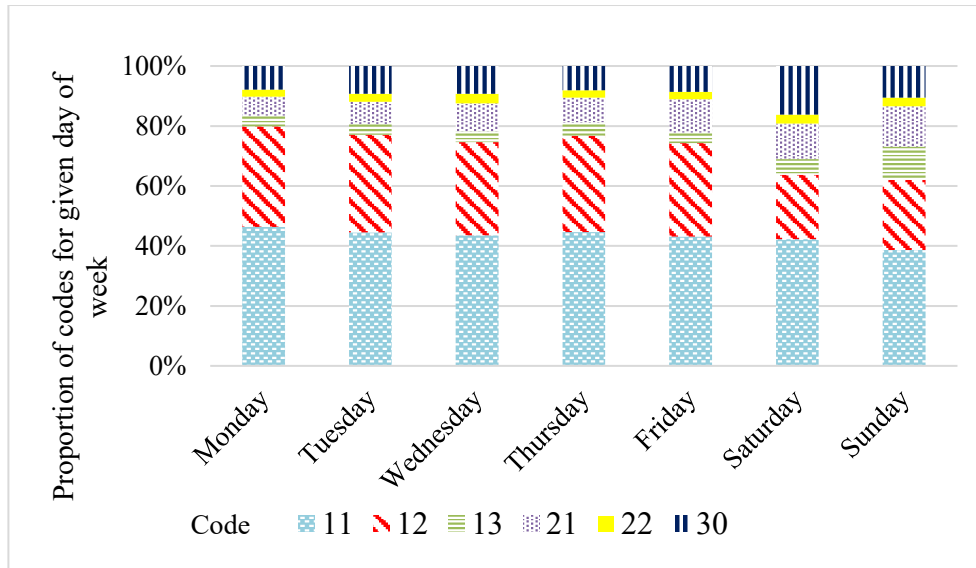


Figure 9: Analysis of estimates by day of week

CONCLUSION

This study has proposed a method to estimate the destination of the unlinked trips in an SCAFC. We have presented a kernel density estimation of the spatial and temporal probabilities of each of the stops in the route, based on the historical record of the card. The results show that our approach resolves an additional 10.9% of the trips for a given month, i.e., 56.3% of the trips that could not be resolved by the classical method.

However, we are not currently able to verify whether or not the destination found is valid, because the STO network is tap-in. We plan to carry out further investigation using data from a tap-in/tap-out network. The tap-out information will allow us to adjust the weighting of spatiality and temporality and to calibrate the parameters, such as the bandwidth in the kernel function and the distance thresholds. With sufficient data, it may be possible to enhance the method via a different weighting of the spatial and temporal probabilities. We will also assess the common-route problem, where users take equivalent routes in their historical data. The use of information from such routes will help to determine destinations.

For the remaining records with no destination (code 30), further inference methods could be explored. We could try to find a passenger whose profile is similar to that of the current passenger and use the destination information of this related passenger. We could also increase the maximum distance and time, but this has to be calibrated. Finally, we could use the general load profile of the route to determine destinations not found by other means.

ACKNOWLEDGEMENTS

The authors wish to thank the Société de transport de l'Outaouais (STO) for providing the data and the Thales group and the Natural Sciences and Engineering Research Council of Canada (NSERC) for providing funding.

REFERENCES

1. Bagchi M, White PR (2004). What role for smart-card data from bus systems? *Municipal Engineer* 157:39–46.
2. Trépanier M, Barj S, Dufour C, Poilpré R (2004). Examen des potentialités d'analyse des données d'un système de paiement par carte à puce en transport urbain. *Congrès de l'Association des transports du Canada*.
3. Bagchi M, White PR (2005). The potential of public transport smart card data. *Transport Policy* 12(5):464–74.
4. Utsunomiya M, Attanucci J, Wilson N (2006). Potential uses of transit smart card registration and transaction data to improve transit planning. *Transportation Research Record: Journal of the Transportation Research Board*, no. 1971. Transportation Research Board of the National Academies, Washington, DC, pp 119–26.
5. Trépanier M, Tranchant N, Chapleau R (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems* 11(1):1–14.
6. Munizaga MA, Palma C (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies* 24:9–18.
7. Chu KKA, Chapleau R, Trépanier M (2009). Driver-assisted bus interview (DABI): Passive transit travel survey using smart card automatic fare collection system and its applications. *Transportation Research Record: Journal of the Transportation Research Board*, no. 2105. Transportation Research.
8. Pelletier, MP, Trépanier M, Morency C (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C* 19:557–568.
9. Morency C, Trépanier M, Agard B (2007). Measuring transit use variability with smart-card data. *Transport Policy* 14(3):193–203.

10. Devillaine F, Munizaga M, Trépanier M (2012). Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record: Journal of the Transportation Research Board* 2276(1):48–55.
11. Spurr T, Chapleau R, Piché D (2014). Discovery and partial correction of travel survey bias using subway smart card transactions. In *Transportation Research Board 93rd Annual Meeting* (No. 14-4665).
12. Shimamoto H (2014). Generation and calibration of transit hyperpath using smart card data. *First International Workshop on Utilizing Transit Smart Card Data for Service Planning*, Gifu, Japan, July 2–3.
13. Nassir N, Khani A, Lee SG, Noh K, Hickman M (2011). Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system. *Transportation Research Record: Journal of the Transportation Research Board* 2263(1):140–150.
14. Oort N (2014). Short term public transport prediction model, based on smartcard data. *First International Workshop on Utilizing Transit Smart Card Data for Service Planning*, Gifu, Japan, July 2–3.
15. Lee S, Ali A (2014). Analyzing subway origin-destination flows by utilizing smart transit card data: Case of Seoul metropolitan area. *First International Workshop on Utilizing Transit Smart Card Data for Service Planning*, Gifu, Japan, July 2–3.
16. Shi X, Hangfei L (2014). The analysis of bus commuters' travel characteristics using smart card data: The case of Shenzhen, China. In *Transportation Research Board 93rd Annual Meeting* (No. 14-2571).
17. Trépanier M, Morency C (2010). Assessing transit loyalty with smart card data. *Twelfth World Conference on Transport Research*, Lisbon, Portugal.
18. Trépanier M, Vassivière F (2008). Democratized smartcard data for transit operators. *Fifteenth World Congress on Intelligent Transport Systems*, New York, USA, November 16–20.
19. Trépanier M, Morency C, Agard B (2009). Calculation of transit performance measures using smartcard data. *Journal of Public Transportation* 12(1):79–96.
20. Nakamura T (2014). Travel and shopping behavior analysis using smart card data. *First International Workshop on Utilizing Transit Smart Card Data for Service Planning*, Gifu, Japan, July 2–3.
21. Trépanier M, Morency C, Agard B, Descoimps É, Marcotte J-S (2012). Using smart card data to assess the impacts of weather on public transport user behavior. *CASPT12 - Conference on Advanced Systems for Public Transport*, Santiago, Chile, July 23–27.
22. Wilson N (2014). Using smart card data and complementary data to improve the performance of public transport systems: A framework and applications. *First International Workshop on Utilizing Transit Smart Card Data for Service Planning*, Gifu, Japan, July 2–3.
23. Si B, Fu L, Liu J, Shiravi S (2014). A multi-class traffic assignment model for predicting transit passenger flows: A case study of Beijing subway network. In *Transportation Research Board 93rd Annual Meeting*, Washington, DC.
24. Sun Y, Schonfeld PM (2014). Schedule-based route choice estimation with automatic fare collection data for rail transit passengers. In *Transportation Research Board 93rd Annual Meeting* (No. 14-0834).