5th International Conference on Advanced Research Methods and Analytics (CARMA2023)
Universidad de Sevilla, Sevilla, 2023
DOI: http://dx.doi.org/10.4995/CARMA2023.2023.16456

Newspapers, images and income support policy

Pietro Cruciata[1,2], Chiara Perfetto[2], Giuliano Resce[2]
[1]Polytechnique Montréal, Canada [2] Department of Economics, University of Molise, Italy

Abstract

To what extent do different newspapers have different kinds of images associated with articles on the same topic? We investigate this research question by considering one of the most important Income Support Policies implemented in Italy in recent times ('Reddito di cittadinanza' – RdC) which generated a strong debate in public opinion. Focusing on the national wide media, we downloaded images associated with articles about RdC and by means of Image Captioning algorithms, we generate the description of them. Results show that different newspapers have images containing different objects. Some topics emerging from images published by newspapers are very exclusive and the sentiment associated with the text extracted from the images has a wide heterogeneity. Furthermore, right-hand newspapers show a lower sentiment compared with left-hand newspapers. Overall, the results confirm that the ideological stance associated with different media outlets is reflected also in the images associated with articles and that the integration of Image Captioning algorithms and Natural Language Processes is very promising in this research area.

Keywords: *Image Analysis; Text Mining; Political Debate; Income Support.*

1. Introduction

It has been widely shown that media outlets have their own ideological stance, which implies a bias in the spreading of news, such as how the topic is covered and how it is presented and discussed (Le Moglie, Turati, 2019; Gentzkow, Shapiro, 2010, Mullainathan, Shleifer, 2005). The heterogeneity in the media's ideological stance has a crucial role in the quality of a democracy. Still, it has also been shown that, since profits are driven by the number of readers or viewers, the supply and the discussion of some news could be caused by users' preferences, leading to the reduction in the supply of information more relevant for the accountability of the political system and less attractive for the public (Ho, Liu, 2015, Sen, Yildirim, 2015).

In this paper, we are interested in understanding to what extent different newspapers have different kinds of images associated with articles on the same topic. We investigate this research question by considering a topic that generated a strong debate in Italian public opinion: an Income Support Policy called '*Reddito di cittadinanza*' (RdC). We build a new original database containing all the images associated with articles about '*Reddito di cittadinanza*' published by all the Italian national-wide newspapers. There is a specific reason why we expect heterogeneity in the images associated with the articles: the debate on the RdC has been characterized by a high degree of politicization. In the 2022 Italian electoral campaign, the RdC was one of the main characters with the programs of the various parties proposing different actions on the policy. The Five Stars Movement (M5S) wanted to strengthen the current system, the Democratic Party (PD) and the Third Pole propose relevant reforms while the unitary program of the Centre-right plans propose to replace the RdC with alternative measures of social inclusion.

Preliminary results show that the ideological stance associated with different media outlets is reflected also in the images associated with articles and that the integration of Image Captioning algorithms and Natural Language Processes is very promising in these analyses. Different newspapers have images containing different objects, and the sentiment associated with the text extracted from the images has a wide heterogeneity, in particular, right-hand newspapers show a lower sentiment compared with left-hand newspapers.

2. Data and method

2.1. Data

To create the corpus for analysis, we developed a Python web scraper. Our objective was to search and download all images related to the query "reddito di cittadinanza" from Google, while targeting specific newspapers adding to the query the name between quotation marks to get more precise results. We decided to gather images from Google for a maximum of 300

images per newspaper. We limited the collection to a maximum of 300 images per newspaper, as we observed that beyond this threshold, the relevance to our research decreased. From Google, we retrieved both the images and the website that uploaded them, and we ensured that an image is associated with a unique link to the respective article. Then, we filtered out any images that did not originate from the selected newspapers' websites. Finally, our sample includes 474 images from 9 different newspapers and are distributed as follows in Table 1:

Table 1. Image distribution for each newspaper

Newspaper	Number of images
Il Mattino	66
La Repubblica	91
La Stampa	20
Il fatto quotidiano	90
Il Corriere della Sera	79
Il Messaggero	71
Il Sole 24 Ore	31
Libero	20
Tuttosport	6

2.2. Methodology

Image Captioning aims to briefly describe an image to assign it a caption. We use an algorithm developed in this subfield extending the captions generated to have a more complete description of the images studied.

Researchers created Image Captioning algorithms through the combination of state-of-the-art algorithms in two main fields of AI: Computer Vision and Natural Language Processing (NLP). Computer vision algorithms are used to recognize the entities in an image while NLP algorithms are used to generate the description of it.

The algorithm that we propose is the Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation (BLIP) (Li, Li, Xiong, & Hoi, 2022). This model is based on two pre-trained transformers: the image transformer initialized from ViT pre-trained on ImageNet (Dosovitskiy et al., 2020), and the text transformer initialized from BERT base (Devlin, Chang, Lee, & Toutanova, 2018). Moreover, at the core of BLIP there is a multi-task pre-training and flexible transfer learning architecture called Multimodal

mixture of Encoder-Decoder (MED). MED works pre-training at the same time three different vision-language objectives: image-text contrastive learning, image-text matching, and image-conditioned language modeling. Finally, MED is finetuned with Captioning and Filtering: a new dataset bootstrapping method in which a captioner produces synthetic captions given web images, and a filter removes noisy captions from both the original web texts and the synthetic texts.

The model provides the flexibility to set various parameters, including the number of beams used for the beam search, the option to use nucleus sampling, and the minimum and the maximum number of characters generated for each caption. Nucleus sampling and beam search are two types of algorithms that allow the model to generate the caption words. While the beam search generates the most probable sequence of tokens, nucleus sampling has a different approach. It generates a subset of tokens where the sum of their probabilities is greater than a predetermined value. For this pilot study, we generate 20 captions for each image, one using beam search and 19 using the nucleus sampling. Additionally, we set the minimum number of words to 10 and the maximum to 20.

Thus, we proceed with the image captions analysis, which will involve unsupervised text mining and statistical analysis of captions to extract information, as well as sentiment analysis to detect the semantic orientation of the content. The different captions of each image will be aggregate/paste into one big text and will be represented by the term document matrix (TDM). This representation allows the data to be analyzed with vector and matrix algebra, effectively moving from text to numbers. In the TDM the rows correspond to the terms in the caption, columns correspond to the newspapers and cells correspond to the frequency of the terms. Not all terms are equally informative for text analysis (Welbers et al., 2017). One of the first things to remove very common terms is the use of "stopword" lists, but it is not sufficient there may be still other common words, and this will be different between corpora. For that, an additional approach is to assign them variable weights. A popular weighting scheme is the term frequency-inverse document frequency (TF-IDF), which uses information about the distribution of terms in the corpus to estimate how exclusive the association is between a word and a document (Hidalgo and Hausmann, 2008; 2009). In detail, TF measures the term-frequency that is the times that a term occurs in the given document and IDF indicates how common and rare that term is across all documents. Formally:

$$IDF(term) = ln\left(\frac{n_{documents}}{n_{documents\ containing\ term}}\right) \quad (1)$$

A high TF-IDF value indicates that the term is important to the given document and possibly represents key information o that document.

For sentiment analysis of the text coming from the captions, we will use the "syuzhet" package (Jockers, 2017) in R which allows the calculations of polarity scores of a collection

of documents by using different internal dictionaries. We will focus on the "nrc" dictionary developed by Turney and Mohammad (2010). The words in the dictionary are classified into multiple labels corresponding to positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise and trust. The algorithm after the comparison between the words in the text with the words in the dictionary counts them and returns a data frame in which each row represents the captions. The columns include one for each emotion type as well as the positive or negative sentiment valence. The score obtained in each cell is divided by the total number of words in each caption. In this way, each image will correspond to a text and the text will associate with emotions.

3. Preliminary results

3.1. Discussion

Figure 1 reflects the TF-IDF of the terms characterizing each newspaper. The right-hand newspaper like "Libero" focuses on "wallet" and "euro"; "La Repubblica" e "Il Mattino" are the most similar and both focuse on "card" and "credit". "La Stampa" and "Il corriere della sera" report image of the police, while "Il fatto quotidiano" rappresents the "people". The image from "Il Sole 24 Ore," which is an economic-centric journal, appears to be depicting images that are more closely related to politicians. However, these images may not be informative in our analysis, as the model often struggles to recognize the individuals depicted. Similarly, "Tuttosport," a sports newspaper, is characterized primarily by sports-related images, which are not relevant to our research. Therefore, these last two newspapers will not be included in the subsequent sentiment analysis as they are unrelated to the RDC issue as demonstrated by TF-IDF.

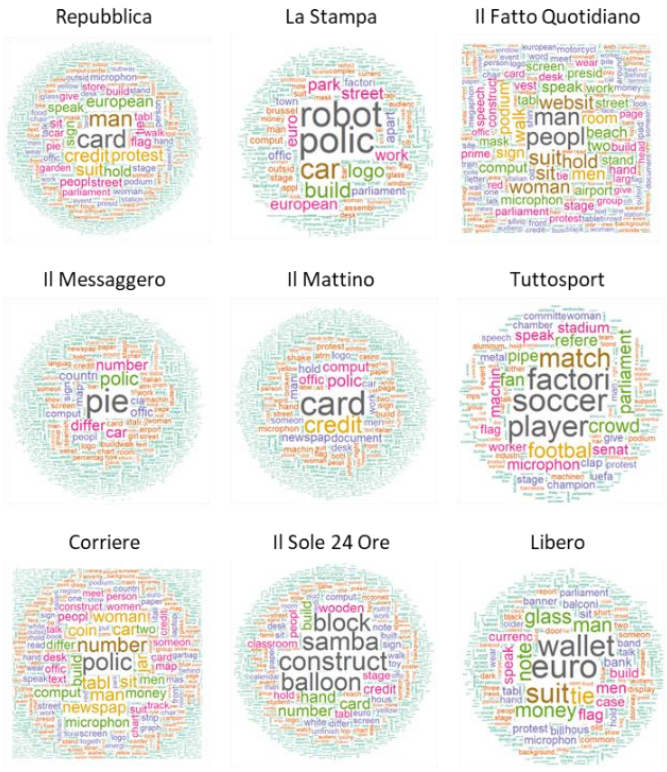

Figure 1. TF-IDF by newspaper

The results obtained from the sentiment analysis are shown in Figure 2 and in Figure 3. Figure 2 shows the percentage of words associated with eight different emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) plus negative and positive general sentiments. The emotions are distributed on the x-axis from the most negative (anger) to the most positive (trust). It is evident from Figure 2 that the emotion "trust" has the longest bar indicating that words associated with this emotion constitute about 20% of all the meaningful words in this text. On the other hand, the emotion of "disgust" shows that words associated with this negative emotion represent about 2% of all the meaningful words in this text. In general, the sentiment is more positive than negative. The specific sentiment related to each newspaper is shown in Figure 3.

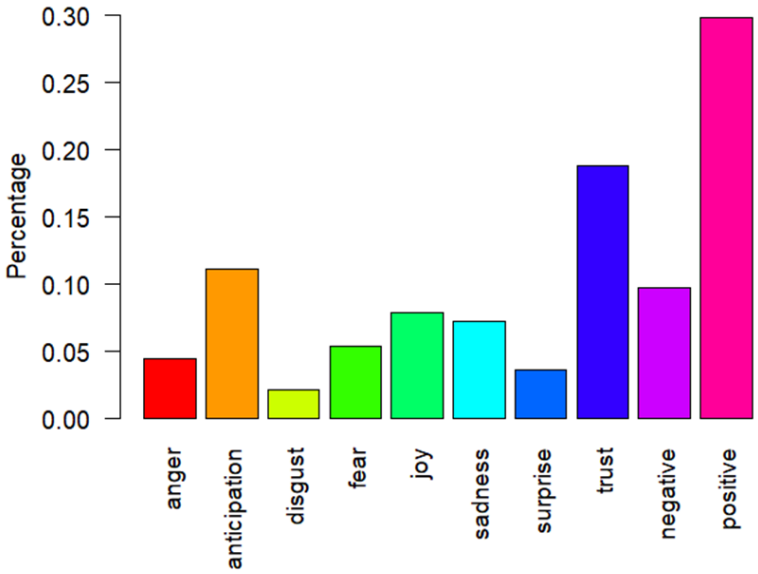

Figure 2. Percentage of words in the text associated with each emotion

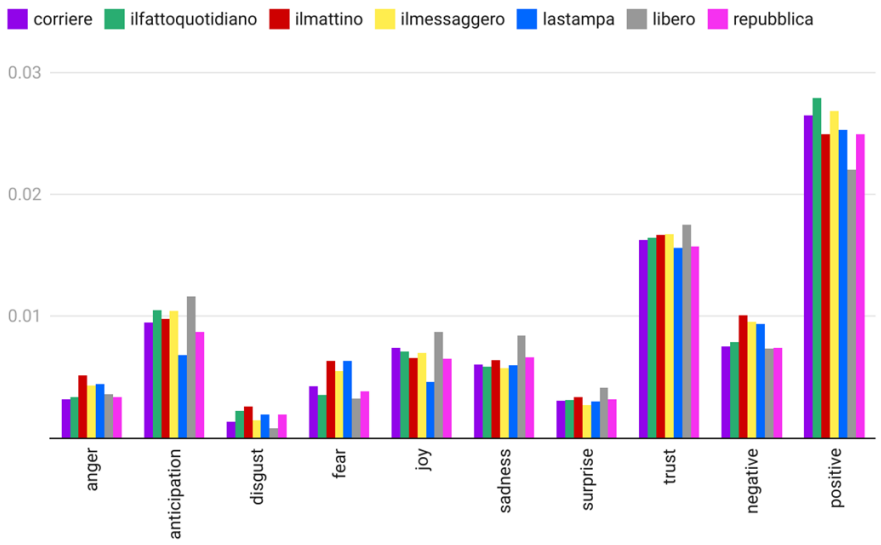

Figure 3. Presence of emotions in each newspaper

Figure 4 shows how the sentiment of each newspaper differs from the average sentiment. The sentiment relating to "Il Fatto Quotidiano" differs positively from the general sentiment average. The lowest sentiments are associated with images in "Il Mattino" and "Libero".

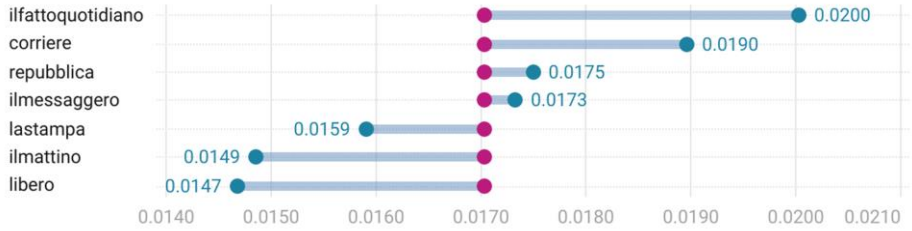

Figure 4. Deviation of the sentiment associated with each newspaper from the average sentiment

In summary, the results confirm that the ideological stance associated with different media outlets is reflected also in the images associated with articles, and in particular the right-hand newspapers show a lower sentiment compared with left-hand newspapers.

3.2. Limitation and Future research

Despite the good results, there are some limitations to this work that surely affect our results. The first limitation is not the best quality of the images. This surely prevents sometimes the ability of the BLIP model to spot all the objects in an image to fully describe it. Another limitation this time is on the language generation part. The model comes already with pre-acquired knowledge that is built in English. Thus, it has difficulties to interpret the image and the context of the topic that we are trying to analyze. Another bias in our methodology stems from the missing data in the information retrieved. Indeed, the images described could bring interesting information to the topic study.

Building from these limitations, there are two main improvements possible to improve the results of our work. The first is to create a pre-trained model that is trained on the subject. This will allow the model to understand the story and the subject involved easing the interpretations. The second improvement possible is to add the date of the image published online to analyze the event that is connected to the different images published by the examined journals.

References

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., … Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv Preprint ArXiv:2010.11929*.

Feinerer I, and Hornik K (2018). tm: Text Mining Package. R package version 0.7-5.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1), 35-71.

Hidalgo, CA, and Hausmann, R (2008). A network view of economic development. *Developing alternatives*, 12(1), 5-10.

Hidalgo, CA, and Hausmann R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26), 10570-10575.

Ho, B. and P. Liu (2015). Herd journalism: Investment in novelty and popularity in markets for news. *Information Economics and Policy* 31, 33–46

Jockers, M. (2017). syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text. R package version 1.0.6.

Le Moglie, M., & Turati, G. (2019). Electoral cycle bias in the media coverage of corruption news. *Journal of Economic Behavior & Organization*, 163, 140-157.

Li, J., Li, D., Xiong, C., & Hoi, S. (2022). Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. *ArXiv Preprint ArXiv:2201.12086*.

Mullainathan, S., & Shleifer, A. (2005). The market for news. *American economic review*, 95(4), 1031-1053.

Mohammad, S., & Turney, P. (2010, June). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. *In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26-34).

Sen, A. and P. Yildirim (2015). Clicks and editorial decisions: How does popularity shape online news coverage? *Mimeo*. Available at SSRN

Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text analysis in R. *Communication methods and measures*, 11(4), 245-265.