Psychological Factors in Travel Behaviour Interpretation with Social Media Data

Yanyan Xu*, Neil Yorke-Smith, Serge Hoogendoorn

Delft University of Technology, Delft, Netherlands
Corresponding author: Y.Xu-10@tudelft.nl

Keywords: Nature Language Processing, Travel behaviour interpretation, Social media data

1 Introduction

Social-psychological factors – namely attitude, perception or belief – have been widely acknowledged to reflect user travel mode preferences and travel intentions, highlighting their vital roles in travel decision-making process (Abenoza *et al.*, 2017). Current studies extensively implement them as latent variables in travel choice modelling and travel behaviour prediction (Kroesen *et al.*, 2017) with emphasis on quantitative features such as categorizing attitudes as 'positive or negative' or preference as 'like or unlike'. However, few explore the underlying *reasons*, which is critical to reveal key motivations behind dynamic travel behaviour and reflect social needs. Coincidentally, these underlying causes are also psychological reactions towards external contextual factors in travel decision-making process such as socio-demographic, policy, economic, environment (Biehl *et al.*, 2018). Given that individuals' psychological reaction varies based on their beliefs and values, it is essential to unveil a comprehensive understanding to effectively analyze the impact of external determinants on travel behaviour, rather than the prevalent approach of directly using them in a generalized way.

In reality, users' psychology is not always aligned with behaviour, leading to the important phenomenon of *travel choice dissonance*. Despite directly identifying 'attitude–behaviour' gaps, psychological variables are also capable of explaining the inherent causes of these dissonance, which implies social inequalities and social needs. Meanwhile, the majority of studies are based on survey data which has some drawbacks including relatively static data, time-consuming costly data collection progress, incomplete user concerns extraction and relatively general spatial-temporal data scale. There is a knowledge gap in obtaining dynamic psychological variables and further explore its underlying causes to disentangle specific key determinants in individuals' mode travel choice in order to support dynamic travel behaviour interpretation, efficient travel management and policy implementation.

This study aims to illustrate the critical role of user psychological factors in travel choice and travel behaviour via social media data and natural language progressing methodologies. By integrating social media with survey data to obtain more comprehensive psychological features including dynamic user attitude and user concerns through sentiment analysis and dynamic topic modelling, we are able to demonstrate how these features help in identifying underlying causes and key determinants on both consistence and inconsistency travel behaviour within neighborhood level and different periods. We implement a case study in New York City via Twitter data to capture transit-related psychological factors on five travel modes (cycling, driving, subway, taxi and ride-hailing, walking) from 2019 to 2022. Results show the advanced ability of social media data in capturing dynamic travel-related user attitudes and concerns and further support in travel choice dissonance identification and dynamic travel behaviour explanation among five travel modes. Consequently, the outcome efficiently reflects various social needs, identifies priority areas and provide valuable suggestions for policy-makers and planners to develop more targeted improvement strategies.

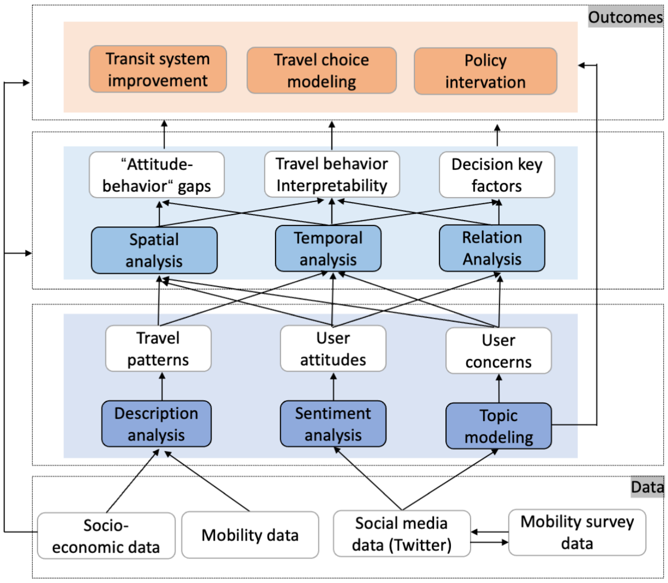

Figure 1 – *Design Framework*

2 Methodology

This paper takes advantage of social media data and natural language processing (NLP) methods to illustrate the extensive utilisation of psychological factors in transportation by identifying underlying causes behind psychological determinants, capturing psychological reactions towards various contextual factors, and recognising social demand according to consistency and inconsistency psychology–behaviour pairs. The conceptual framework is depicted in Figure 1. Sentiment analysis and dynamic topic modelling were first implemented to capture psychological features including attitude, opinions and geo-spatial patterns. Then, the study illustrates how these features could help in identifying dynamic key determinants of travel behaviour from spatial analysis and time series analysis, thereby contributing to policy suggestions.

2.1 Sentiment analysis and dynamic topic modeling

We implement a BERT model in order to classify social media postings (tweets). We first fine-tune the model with domain-specific data to improve the classification performance on our target dataset. Using the PyTorch library, we embed the fine-tuned model into torch tensors. We test the fine-tuned BERT model on our own dataset with 600 rows of manually-labelled tweets, and find accuracy of 0.87, demonstrating relatively good performance on the target dataset.

In order to capture user concerns, we also implement dynamic topic modelling based on TopicBERT on different transit modes (Grootendorst, 2022). Specifically, we embed each document (i.e. tweet) by the semantically-similarity logic via pre-trained language model Sentence-BERT, and convert each of the tweet into numerical representations. Given that cluster models have difficulty in dealing with such high-dimensional data, we use UMAP to reduce the dimension of the embedding data from previous steps while preserve the local and global features of embedded documents. Then, HDBSCAN is applied to cluster the reduced embedding with noise data labels as outliers. Then, each document is labelled as one topic and we use TF-IDF to present term importance for a certain topic. Finally, we create topic representations from clusters using c-TF-IDF:

$$W_{t,c} = \; tf_{t,c} * log(1 + \frac{A}{tf_t}) \tag{1}$$

where t, c is the term t in a class c, c being a cluster collection of documents; $W_{t,c}$ is the weight

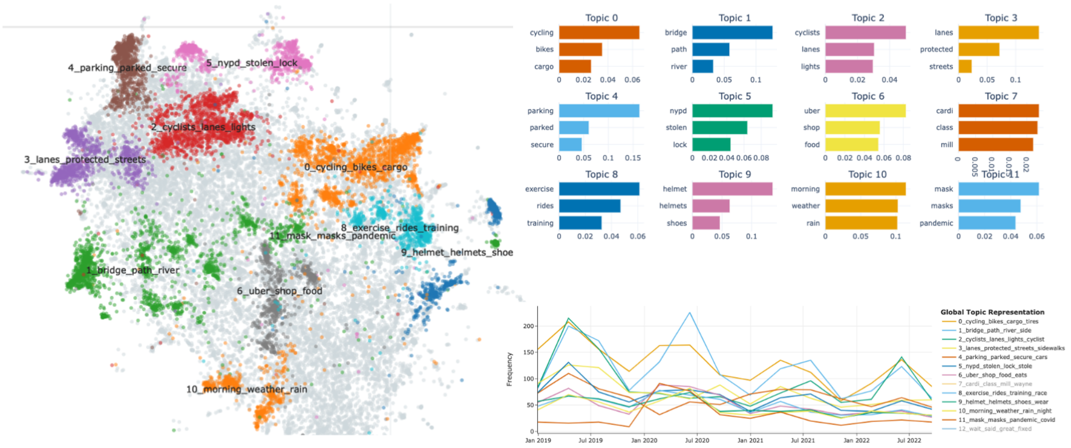

Figure 2 – *Dynamic topic modelling: example of cycling mode*

of each term i in class c; $tf_{t,c}$ is the frequency of each term i in class c; A is the average amount of terms per class; f_i is the frequency of term i among all classes.

2.2 Temporal-spatial analysis

In order to calculate change ratio of certain ridership or attitude within certain time interval, we implement descriptive analysis by comparing these factors in time series within four years from 2019 to 2022. Further, by taking advantage of precise latitude and longitude of each tweet, we implement spatial analysis by aggregating tweet volumes in each neighbourhoods to identify the most changeable areas in both attitude and travel ridership. We also calculate the standardized differential values between attitude and behaviour patterns among different travel modes to identify the areas that contains more 'attitude–behaviour' gaps. Further, we utilize topic modeling results to point out the underlying causes of these areas and indicate the specific user demands towards each travel modes.

3 Results

We analyse five travel modes for New York City, for the period Jan 2019 to Dec 2022. Figure 2 shows the example result for the travel mode of cycling. Specifically, the left sub-figure reveals the global topic clusters with distributions and relations among each topic for the overall four years. The upper-right sub-figure shows the keywords and probabilities of each topic, while the lower-right sub-figure presents the dynamic trend of top topics in time series that used to explain consistency and inconsistency travel behaviour.

We summarise the key topics in Table 1 for the five travel modes. The temporal and spatial analysis results first presents the most changeable travel patterns and the 'attitude–behaviour' gaps temporally and spatially among each travel modes. Figure 3 shows cycling travel choice dissonance patterns in the four years for which there is more gaps during the summer and most of the dissonance areas are concentrated in the downtown Manhattan. We then utilize dynamic topic modelling results to point out key determinants, explain reasons and indicate social needs.

Table 1 – *Summary of overall topic in five travel modes*

Topics	***Public transit***	***Driving***	***Ride***	***Cycling***	***Walking***
Topic 1	time,delay,incident	parking	ride-sharing,pool	cycle lanes	pedestrian,safety
Topic 2	mask,COVID	buying, credit	driver, cancel	parking	food,coffee
Topic 3	route,bus	electric cars	mask,COVID	path	time

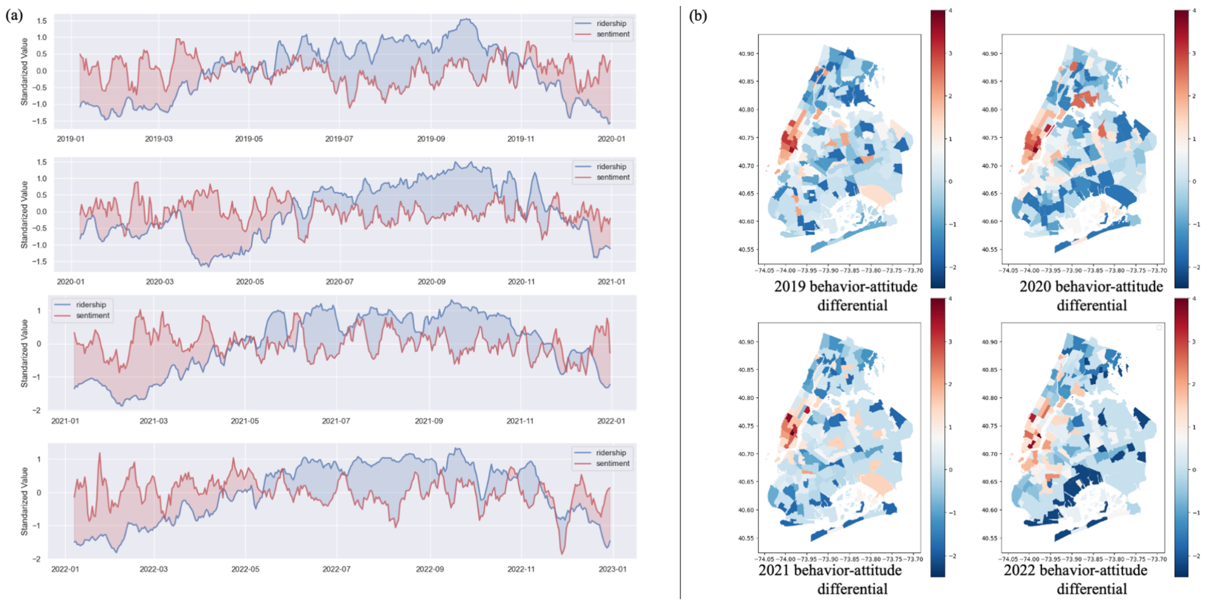

Figure 3 – *'Attitude–behaviour' gaps identification: example of cycling mode*

4 Discussion

This paper illustrates the ability of social media data via sentiment analysis and dynamic topic modelling to capture psychological features while identifying key motivations to explain dynamic travel behaviour and travel choice dissonance in more precise areas and dynamic periods. It further help policy-makers understand social needs and identify priority areas to implement efficient interventions. Further, this paper states how social media data integrate with survey data to obtain a more comprehensive psychological features. The case study also indicates the interactive relations among each travel mode according to user psychological perceptions.

References

Abenoza, Roberto F, Cats, Oded, & Susilo, Yusak O. 2017. Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. *Transportation Research Part A: Policy and Practice*, **95**, 64–84.

Biehl, Alec, Ermagun, Alireza, & Stathopoulos, Amanda. 2018. Modelling determinants of walking and cycling adoption: A stage-of-change perspective. *Transportation research part F: traffic psychology and behaviour*, **58**, 452–470.

Grootendorst, Maarten. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Kroesen, Maarten, Handy, Susan, & Chorus, Caspar. 2017. Do attitudes cause behavior or vice versa? An alternative conceptualization of the attitude-behavior relationship in travel behavior modeling. *Transportation Research Part A: Policy and Practice*, **101**, 190–202.