Demand Modelling for Responsive Transport Systems Using Digital Footprints

Paulo Silva, Francisco Antunes*, Rui Gomes, and Carlos Bento

Faculdade de Ciências e Tecnologia da Universidade de Coimbra
Pólo II, Rua Sílvio Lima, 3030-790 Coimbra, Portugal
pacsilva@student.dei.uc.pt {fnibau,ruig,bento}@dei.uc.pt

Abstract. Traditionally, travel demand modelling focused on long-term multiple socio-economic scenarios and land-use configurations to estimate the required transport supply. However, the limited number of transportation requests in demand-responsive flexible transport systems require a higher resolution zoning. This work analyses users short-term destination choice patterns, with a careful analysis of the available data coming from various different sources, such as GPS traces and social networks. We use a Multinomial Logit Model, with a social component for utility and characteristics, both derived from Social Network Analyses. The results from the model show meaningful relationships between distance and attractiveness for all the different alternatives, with the variable distance being the most significant.

Keywords: Innovative transport modes · public transport operations · transport demand and behaviour · urban mobility and accessibility

1 Introduction

Transportation systems are a key factor for economic sustainability and social welfare, but providing quality public transportation may be extremely expensive when demand is low, variable and unpredictable, as it is on some periods of the day in urban areas. Demand Responsive Transportation (DRT) services try to address this problem with routes and frequencies that may vary according to the actual observed demand. However, in terms of financial sustainability and quality level, the design of this type of services may be complicated.

Anticipating demand by studying users short-term destination choice can improve the overall efficiency and sustainability of the transport services. Traditionally, demand modelling focused on long-term socio-economic scenarios and land-use to estimate the required level of supply. However, the limited number of transportation requests in DRT systems does not allow the application of traditional models. Also, DRTs require a higher resolution zoning, otherwise it can lead to unacceptable inaccuracies. Information coming from various sources should be used effectively in order to model demand for DRTs trips.

* corresponding author: Universidade de Coimbra, fnibau@dei.uc.pt, +351 239790056.

The approach followed in this work analyses users short-term destination choice patterns, with a careful analysis of the available data coming from various different sources, such as, GPS traces and social networks. The theory of utility maximization, usually through discrete choice modelling, is often used to study individual decision-making. We use the Multinomial Logit Model (MNL) with a social component for utility and characteristics, both derived from Social Network Analyses (SNA), where a network is constructed linking the nodes (decision makers) that have social influence over one another (friendship), and the strength of that influence. To measure different ties strength, mutuality, propinquity, mutual friends and multiplexity factors were used.

We review the state of the art in the next section. The methodology is presented in Section 3 and the results in Section 4. The documents ends with the conclusions and possible future lines of work.

2 State of the art

Urban movements profiling has usually relied on traditional survey methods that are expensive and time consuming, giving planners only a picture of what has happened. In contrast, the wide deployment of pervasive computing devices (cell phone, GPS devices and digital cameras) provide unprecedented digital footprints, telling where and when people are. An emerging field of research uses mobile phones for "urban sensing" [1]. Moreover, the past few years have witnessed a huge increase in the adoption of social media and transportation researchers have also realized the potential of SNA for demand modelling [2].

A growing research topic is understanding how trips, trip modes and trip purposes can be derived from GPS data. For instance, [3] propose an approach to predict both the intended destination and route of a person by exploiting personal movement data collected by GPS. GPS data, however, have some limitations, such as (1) GPS signals are usually blocked indoor, (2) GPS devices may get interferences near tall buildings, and (3) continuously collecting GPS data may consume devices energy quickly.

Social networks and human interactions are crucial not only for understanding social activities, but also for travel patterns [2]. [4] connects travel with social networks, arguing that daily life revolves around family, colleagues, friends and shopping. [5] refers to the conformation to social norms, implying that decision-makers are more likely to choose a particular alternative if more peers have already chosen the same alternative. The emergence of geolocated social media seems a good opportunity to address SNA's lack of geographic consideration. For instance, [6] presents a technique to analyse large-scale geo-location data from social media to infer individual activity patterns.

Previously, for travel demand modelling aggregate approaches were used, such as gravity or entropy models. These approaches were gradually replaced by disaggregated models [7]. In discrete choice modelling, the effect of social dimensions was first formalized for the binomial and the multinomial cases in [8] and [9], respectively. Generically, the agents' utility is formed by both private and social components. The private component corresponds to the decision-makers characteristics. The social component represents the strength of social utility and the percentage of others in the neighbourhood selecting the same alternative in the choice set [10].

3 Methodology

3.1 Data gathering

We use GPS data traces provided by TU Delft from 80 individuals over the course of four days, and also data collect from social networks, namely Twitter, Instagram and Foursquare. The data obtained is cleared of personal values as to ensure privacy. To get the geo-located points of interest, we use the FourSquare API, extracting the 50 most popular venues, within a radius of 30 meters for each given point, resulting in a total of 37506 venues, in 489 categories, with their identification, geo-location and total number of check-ins made. The subscription zone for Instagram had a radius of 5 kilometers from the city center. For Twitter, we covered a bigger area in order to get Delft surroundings.

3.2 Social Network Analysis

Friendship To get the friendship, we have to use the user unique identification from the post, and request the users that the user followed and that follow him back. The only significant friendships considered are the ones between users that posted around Delft. The total number of friendships used is 35457. Discrete choice model also had to take into account the strength between users. To get and measure the ties strength, tie mutuality, propinquity, mutual friends and multiplexity factors are used.

Detecting important locations To build the MNL we also need to know the user home and work location, since we are only interested in the user movement patterns before and after work hours. Home and work are the starting points for which the distance to the points of interest are measured. To get these locations, we use a clustering algorithm, namely *DBScan* [11].

3.3 Data preparation

In the data set with the posts and associated venues, i.e., the choice set (CS), there is a large amount of data with no use for us, as it does not provide useful information (for instance, useless categories) or represent work or residential places, for which the demand patterns are well established and can be met by traditional transportation services. The data containing those specific categories was erased from the choice set. Since the number of alternatives is quite big, we grouped those venues in 6 main categories: Appointment (17%), Food (17%), Bar (5%), Shop (24%), Entertainment (27%) and Travel (10%).

If we used these categories as our number of different alternatives for the MNL model, we would only get results concerning each of those 6 alternatives, which are quite generic. However, we want to use the model to predict probabilities of destination choices with a higher resolution, so we generated data for all the venues and then use those categories only to filter unnecessary data.

3.4 Multinomial logit model

Our data corresponds to the observed choices of individuals - revealed preferences data. For each dataset we have the number of alternatives selected in each hour, which is our finite set of alternatives for each individual. The number of alternatives and observations vary significantly along the hours. The variables used for the data-frame are:

- distance : the venue distance to the user central point,
- check-ins : the total number of check-ins in each alternative for each user,
- friendship : the sum of the individual friendship for each alternative,
- choice : the alternative selection.

Since we cannot directly extract user personal information (e.g. age, gender), our data does not contain individual specific variables, and so the alternative specific variables have a generic coefficient, i.e., we consider that the number of check-ins, distance and friendship have the same value for all alternatives. Choice takes values of yes and no, if the alternative was chosen or not by the user. To estimate the MNL we have used the R statistics system with the mlogit package. The following formula was used for our work,

$$Mlogit(choice \sim distance + friendship + attractiveness, CS)$$

where choice is the variable that indicates the choice made for each individual among the alternatives and the distance, friendship and attractiveness being the alternative specific variables with generic coefficients from the choice set CS.

4 Results

We present the results and estimation parameter for one choice set, namely the one representing the choices made at hour 21, which has 24 alternatives and 91 observations.

The model predictions are reasonable good when tested against the user observed choices. Table 1 presents the average probabilities returned by the model against the observed frequency. The results from the MNL model show meaningful relationships between distance and attractiveness for all the different alternatives, being distance the most significant variable, i.e., longer distances almost always reduce the attractiveness of a destination, all else being equal. The same can be said for the attractiveness variable, but the friendship variable does not have the same impact to the individual when choosing an alternative.

Table 1. Average probabilities returned by the model

Venue	Freq.	Avg. Prob.
Stadion Feijenoord	0.032967	0.04490835
EkoPlaza	0.065934	0.03791752
Station Den Haag HS	0.054945	0.05494505
La Mer	0.032967	0.03303908
Diner Company	0.032967	0.03777430
LantarenVenster	0.032967	0.04320755
Station Rotterdam Centraal	0.153846	0.12336128
BIRD	0.043956	0.03794951
Maassilo	0.043956	0.03530117
Emma	0.021978	0.02478593
Station Den Haag Centraal	0.054945	0.05070866
Lucent Danstheater	0.032967	0.02519137
Zaal 3	0.032967	0.03212339
De Banier	0.054945	0.09073518
Randstadrail javalaan	0.032967	0.02411359
Kot Treinpersoneel	0.032967	0.02666024
Spuimarkt	0.032967	0.02543518
Doerak	0.032967	0.04017886
Paard van Troje	0.032967	0.03294463
Ahoy Rotterdam	0.043956	0.03654922
Restaurant Meram	0.043956	0.03907711
Live Tv Show	0.021978	0.04634078
Stadskwekerij den haag	0.010989	0.03433138
Oudedijk 166 A2	0.032967	0.04312563

Table 2. Relationships between variables

Variables	Estimate	Std.error	t-value	p-value
distance	-0.107414	0.022597	-4.7534	2.000e-06
friendship	0.094441	0.081570	1.1578	0.2469
attractiveness	0.342170	0.061907	5.5272	3.254e-08

Fig. 1. Simulation results

Table 2 illustrates these findings. To show the usefulness of the analyses made, we feed the probabilities predicted by our model to a DRT simulator developed in [12]. Figure 1 shows that most origins and destinations found for the time period and travel objective considered lie outside the service area of the different public transport modes (dotted lines) and DRT could satisfy this demand (solid lines).

5 Conclusions

Traditionally, travel demand modelling focused on long-term socio-economic scenarios and land-use to estimate the required transport supply. However, the limited number of transportation requests in demand-responsive flexible transport systems require a higher resolution zoning. We analysed users short-term destination choice patterns, with a careful analysis of the available data coming from GPS traces and social networks. We defined a Multinomial Logit Model (MNL), with a social component for utility and characteristics, both derived from Social Network Analyses. The low frequency of posts with identified locations for each user made it difficult to generate a clear pattern for each user. Nevertheless, the results from the model show meaningful relationships between distance and attractiveness for all the different alternatives, with the variable distance being the most significant.

Since the analyses of the social network done in this work does not produce individual characteristics, like age, gender and socio-economic, it would be interesting for future work to include data mining algorithms to extract some of those values from tweets, and add features specific to each venue, to better understand the motivation behind the choice made.

References

1. Cuff, D., Hansen, M., Kang, J.: Urban sensing: Out of the woods. Communications of the ACM , 51 (3), 24-33 (2008).
2. Carrasco, J., Hogan, B., Wellman, B., Miller, E.: Collecting social network data to study social activity-travel behaviour: an egocentric approach. Environment and Planning B: Planning and Design , 35, 961 - 980 (2008).
3. Chen, L., Mingqi, L., Chen, G.: A system for destination and future route prediction based on trajectory mining. Pervasive and Mobile Computing , 6 (6), 657676 (2010).
4. Axhausen, K.: Social networks, mobility biographies, and travel: survey challenges. Environment and Planning B: Planning and Design , 35, 981996 (2008).
5. Paáez, A., Scott, D.: Social influence on travel behavior: a simulation example of the decision to telecommute. Environment and Planning A , 39 (3), 647-665 (2207).
6. Hasan, S., Ukkusuri, S.: Urban activity pattern classification using topic models from online geo-location data. Transportation Research Part C: Emerging Technologies , 44, 363381 (2014).
7. Ben-Akiva, M., Lerman,S.: Discrete Choice Analysis: Theory and Application to Travel Demand (1985).
8. Brock, W., Durlauf, S.: Discrete Choice with Social Interactions. Review of Economic Studies, 68 (2): 235-260 (2001).
9. Brock, W., Durlauf, S.: A multinomial choice model with neighborhood effects. American Economic Review 92,298303 (2002).
10. Zanni, A.M., Ryley, T.J.: Exploring the possibility of combining discrete choice modelling and social networks analysis: an application to the analysis of weather-related uncertainty in long-distance travel behaviour. International Choice Modelling Conference, Leeds, pp.1-22 (2011).
11. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, AAAI Press, 226-231 (1996).
12. Gomes, R., Sousa, J.P., Galvao, T.: An integrated approach for the design of Demand Responsive Transportation services. Advances in Intelligent Systems and Computing, 232, pp.223-235 (2014).