Travellers' activities preference prediction using social media data

Charis Chalkiadakis[1], Panagiotis Iordanopoulos[1], Evangelos Mitsakis, Ph.D. [1] and Eleni Vlahogianni, Ph.D.[2]

[1] Hellenic Institute of Transport – Centre for Research and Technology Hellas, 6th Km Charilaou - Thermi Rd., 57001 Thermi, Thessaloniki, Greece

[2] Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece

Extended Abstract

There is a direct connection between the activities performed by the travellers and the concept of travel demand. A number of studies define that connection between these two concepts. A fragmentation for the further development of methodologies referred to the connection between activities and travel demand was the non-existence of detailed data that could allow for a better understanding of the linkage between the performed activities by the travellers and travel demand (Axhausen 2000, 2008; Carrasco et al. 2008).

Taking into account the rapid development and the wide use of social media from the largest proportion of the population (2.46 billion social networks' users worldwide in 2017) (Statista 2016), there is a source of information which can lead to the development of methodologies for the estimation of travel demand by integrating social media oriented data driven techniques in the well-established transportation concepts. Social media data can have a structured or unstructured form and is characterized by big volume, wide spatial coverage, long observational period and real-time features (Gal-Tzur et al. 2014; Zheng et al. 2015; Ruiz et al. 2016). These characteristics may vary, depending on the source of information. This variability may significantly affect the complexity of the data driven models developed to identify travel patterns and improve the understanding of urban mobility.

There are two factors that need to be taken into account during the estimation of travellers' activities. The first is related to data mining techniques which have to be used in order to collect, process and analyse the necessary data from social media. In general, the algorithms used for data mining purposes first collect Status Update Messages (SUMs - the user's message shared in social networks) from social media and then process the fetched SUMs by applying text mining techniques, in order to assign the appropriate class label to each SUM and analyse the classified SUMs. Through this process the system can notify the presence of a traffic event while adding geographic information is also feasible. There are four main modules, which consist the overall model architecture (Toumpalidis and Karanikolas, 2015, Grau et al.,; Wanichayapong et al. 2011; Gal-Tzur et al. 2014; Chaniotakis and Antoniou 2015; Chaniotakis et al. 2015, 2016; D'Andrea et al. 2015; Sinha et al. 2016; Kuflik et al. 2017; Toumpalidis 2017):

- Search of SUMs and Pre-processing procedures
- Elaboration of SUMs
- Classification of SUMs
- Geo-location

The methodologies for transport-related data mining use a variety of algorithms in order to gather, process, validate and (in some methodologies) geo-locate and represent the outcome of each technique.

The second aspect that needs to be taken into consideration is the methods that utilize the derived from social media data in order to proceed with the travel demand models. There are various studies for the activity-based models and each study provides different approaches upon activity-base models such as those of (Adler

and Ben-Akiva 1979; Ben-Akiva and Bowman 1998, 1999). The study of Adler and Ben-Akiva (Adler and Ben-Akiva 1979) was one of the first contributions in the fields of activity-based modelling, with a model considering the daily activity program. In the other studies (Ben-Akiva and Bowman 1998, 1999) the models are integrated activity-based discrete choice model systems. In the first study an integrated discrete choice model system of a household's residential location choice and its members' activity and travel schedules is presented, while in the second study an integrated activity-based discrete choice model system of an individual's activity and travel schedule, for forecasting urban passenger travel demand is presented. E. Cascetta (Cascetta 2009) describes the methodology for the estimation of trip-based and activity-based demand models. A "reduced" version of the four-step trip-based travel demand model can be used by taking into account historical matrices. There is also the possibility to calculate the attributes of the travellers' destination and the probability of a traveller choosing a specific destination either by taking into consideration the mode and the destination or only the destination.

In this study a methodology based on data mining techniques is proposed with the aim to efficiently collect and validate social media data and incorporate the extracted knowledge to existing activity models. The main source of data for collecting social media data is the Facebook API and the collected information is the number of check-ins per location. Despite the wide use of Twitter in most of the existing studies, Facebook API is much more suitable in terms of privacy issues (the collected data from Facebook API is anonymized data).

The proposed methodology is based on one main dataset coming from the Facebook API. Two additional datasets, one from another social-media platform's API and one from historical matrices, is being utilized for evaluation purposes. After the validation of the data and through the use of an activity-based model, the proposed methodology leads to the activities' preferences.

References

Adler T, Ben-Akiva M (1979) A theoretical and empirical model of trip chaining behavior. Transp Res Part B 13:243–257 . doi: 10.1016/0191-2615(79)90016-X

Axhausen KW (2008) Social networks, mobility biographies, and travel: Survey challenges. Environ Plan B Plan Des 35:981–996 . doi: 10.1068/b3316t

Axhausen KW (2000) Activity-based modelling Research directions and possibilities

Ben-Akiva M, Bowman JL (1999) Activity-based Disaggregate Travel Demand Model System with Activity Schedules. 1–47

Ben-Akiva M, Bowman JL (1998) Integration of an Activity-based Model System and a Residential Location Model. 1–29

Carrasco JA, Hogan B, Wellman B, Miller EJ (2008) Collecting social network data to study social activity-travel behavior: An egocentric approach. Environ Plan B Plan Des 35:961–980 . doi: 10.1068/b3317t

Cascetta E (2009) Transportation Systems Analysis - Models and Applications

Chaniotakis E, Antoniou C (2015) Use of Geotagged Social Media in Urban Settings: Empirical Evidence on Its Potential from Twitter. IEEE Conf Intell Transp Syst Proceedings, ITSC 2015–Octob:214–219 . doi: 10.1109/ITSC.2015.44

Chaniotakis E, Antoniou C, Aifadopoulou G, Dimitriou L (2016) Inferring Activities From Social Media Data. 96th Annu Meet Transp Res Board 1–8

Chaniotakis E, Antoniou C, Mitsakis E (2015) Data for Leisure Travel Demand from Social Networking Services. Hear Conf 6

D'Andrea E, Ducange P, Lazzerini B, Marcelloni F (2015) Real-Time Detection of Traffic from Twitter Stream Analysis. IEEE Trans Intell Transp Syst 16:2269–2283 . doi: 10.1109/TITS.2015.2404431

Gal-Tzur A, Grant-Muller SM, Kuflik T, et al (2014) The potential of social media in delivering transport policy goals. Transp Policy 32:115–123 . doi: 10.1016/j.tranpol.2014.01.007

Grau JMS, Chaniotakis E, Toumpalidis I, et al Big data for transportation analysis and trip generation

Grau JMS, Toumpalidis I, Chaniotakis E, et al A spatio-temporal correlation between digital and physical world, case study in Thessaloniki

Kuflik T, Minkov E, Nocera S, et al (2017) Automating a framework to extract and analyse transport related social media content: The potential and the challenges. Transp Res Part C Emerg Technol 77:275–291 . doi: 10.1016/j.trc.2017.02.003

Ruiz T, Mars L, Arroyo R, Serna A (2016) Social Networks, Big Data and Transport Planning. Transp Res Procedia

18:446–452 . doi: 10.1016/j.trpro.2017.01.122
Sinha M, Varma P, Sivakumar G, et al (2016) Improving Urban Transportation through Social Media Analytics. CODS ’16 Proc 3rd IKDD Conf Data Sci 1–2
Statista (2016) Number of Worldwide Social Network Users 2010-2019. Statista
Toumpalidis I (2017) Physical Spaces and Digital Flows : Navigating through the Informational Matrix
Toumpalidis I, Karanikolas N (2015) Spatial Data Mining from Social Media Services. Aristotle University of Thessaloniki
Toumpalidis I, Karanikolas N Spatial Data Analysis from Social Media Services
Wanichayapong N, Pruthipunyaskul W, Pattara-Atikom W, Chaovalit P (2011) Social-based traffic information extraction and classification. nternational Conf ITS Telecommun 107–112 . doi: 10.1109/ITST.2011.6060036
Zheng X, Chen W, Wang P, et al (2015) Big Data for Social Transportation. IEEE Trans Intell Transp Syst 17:620–630 . doi: 10.1109/TITS.2015.2480157