A Cluster Analysis of Temporal Patterns of Travel Production in the Netherlands: Dominant within-day and day-to-day patterns and their association with Urbanization Levels

Zahra Eftekhar[1], Adam Pel[1], and Hans van Lint[1]

[1]Department of Transport & Planning, Delft University of Technology, The Netherlands

SHORT SUMMARY

Call detailed records (CDR) can show valuable insights in various contexts such as daily mobility motifs (Schneider, Rudloff, Bauer, & González, 2013; Jiang, Ferreira, & Gonzalez, 2017), population movements (Antoniou, Chaniotakis, Katrakazas, & Tirachini, 2020; Szocska et al., 2021), and disaster response (Yabe, Jones, Rao, Gonzalez, & Ukkusuri, 2022). While all the mentioned studies use raw trajectories, acquiring this type of data is difficult. The telecom operators are banned from providing ground truth or contextual information to protect people's privacy at the—inevitable—expense of data utility. Usually, what is available is processed aggregated OD matrices for which the methodology used for processing and up-scaling to the whole population is unclear. This data is a valuable source of information on people's mobility, but its suitability for transportation planning remains unclear. Are these OD matrices representative of the population demand variations and irregular patterns? In many cases, due to limited market share of the telecom operator, only a small and possibly behaviorally biased sample of people are presented in the raw data; How well do they represent the whole population? Mamei, Bicocchi, Lippi, Mariani, and Zambonelli raises the attention to the need for more research and experiments assessing and evaluating the OD matrices derived from these data sources.

To make these processed data useful for urban planning, we must clarify their biases even when facing limited data resources. In this regard, *veracity* of data deals with its representativeness and correspondence to the real world. Essentially, data veracity has two main aspects: data consistency and data trustworthiness (Demchenko, Grosso, De Laat, & Membrey, 2013). Veracity assessment can be performed by using multiple reliable sources, checking data credibility using various quantitative tools, and evaluating the internal data structure (Rubin & Lukoianova, 2013). In this paper we evaluate the veracity of the OD matrices resulting from CDR both by assessing their associated travel demand patterns (representing internal data structure) and by comparison with data from an independent source, in our case land-use characteristics.

Understanding the travel demand patterns is essential for transportation planning and management. Firstly, travel demand analysis plays a significant role in identifying the current problems of transportation systems and helps in modeling the future traffic state (Thakuriah, 2001; Rich & Mabit, 2012). Secondly, the demand patterns help evaluate the impact of transportation infrastructure and management policies and strategies, such as flexible-time work schedules and congestion pricing (Gärling et al., 2002). Thirdly, understanding demand patterns is useful to develop better standards for evacuation plans and responses (Pel, Bliemer, & Hoogendoorn, 2011; Xu, Chen, & Yang, 2017).

Fekih et al. proposes a framework to extract spatiotemporal travel demand patterns from large-scale GSM traces. Their analysis focuses on within-day variations of travel demand. In this paper we build on this work and investigate both within-day and day-to-day production patterns of all the traffic analysis zones (TAZs) in the Netherlands. We tried to capture a more holistic picture of production temporal patterns through normalized heatmaps. After feeding these heatmaps to a deep convolusional neural netwrok (DCNN) and K-means method, three main patterns were discerned. Analysis of these patterns is one of the two methods of evaluating data veracity. The performed temporal analysis of the underlying patterns is valuable for adjusting the demand models and prediction. Later we link these temporal patterns to spatial urbanization levels, through their land-use characteristics, which is beneficial for urban development strategies and policy makers. In fact, this study proposes three urbanization levels for the Netherlands: urban, rural, and other associated with their specific land-use characteristics and travel production pattern within-day and day-to-day, and this study explores the differences in these patterns. This effort is the second method of assessing the data veracity. Furthermore, a OVR-SMOTE-XGBoost ensemble classification method is proposed to investigate the relationship between land-use characteristics and production temporal patterns. Our findings suggest that given the land-use features of each TAZ, their most probable travel production temporal pattern is detectable. The results are beneficial for dynamic demand prediction models. Furthermore, we find indications for the data's ability to represent both dominant patterns and variations correctly in spite of the data's inherent bias. This study will help planners discern and assess the representativeness and suitability of using the processed, up-scaled derivative of GSM traces in traffic planning.

Notwithstanding the relatively limited sample (one month of travel production at a specific spatial-temporal scale), this work establishes a quantitative framework for detecting hourly and daily temporal patterns and how spatial urbanization level helps in demand modeling. Such analysis is required before using the processed demand data for policy making and network development. Further research could also be conducted to determine the effectiveness of using the land-use characteristics (on top of other variables) in improving the demand prediction models.

REFERENCES

Antoniou, C., Chaniotakis, E., Katrakazas, C., & Tirachini, A. (2020). A better tomorrow: towards human-oriented, sustainable transportation systems. *European Journal of Transport and Infrastructure Research*, *20*(4), 354–361.

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In *2013 international conference on collaboration technologies and systems (cts)* (pp. 48–55).

Fekih, M., Bonnetain, L., Furno, A., Bonnel, P., Smoreda, Z., Galland, S., & Bellemans, T. (2021). Potential of cellular signaling data for time-of-day estimation and spatial classification of travel demand: a large-scale comparative study with travel survey and land use data. *Transportation Letters*, 1–19.

Gärling, T., Eek, D., Loukopoulos, P., Fujii, S., Johansson-Stenman, O., Kitamura, R., ... Vilhelmson, B. (2002). A conceptual analysis of the impact of travel demand management on private car use. *Transport Policy*, *9*(1), 59–70.

Jiang, S., Ferreira, J., & Gonzalez, M. C. (2017). Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, *3*(2), 208–219.

Mamei, M., Bicocchi, N., Lippi, M., Mariani, S., & Zambonelli, F. (2019). Evaluating origin–destination matrices obtained from cdr data. *Sensors*, *19*(20), 4470.

Pel, A., Bliemer, M., & Hoogendoorn, S. (2011). Modelling traveller behaviour under emergency evacuation conditions. *European Journal of Transport and Infrastructure Research*, *11*(2).

Rich, J., & Mabit, S. L. (2012). A long-distance travel demand model for europe. *European Journal of Transport and Infrastructure Research*, *12*(1).

Rubin, V., & Lukoianova, T. (2013). Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online*, *24*(1), 4.

Schneider, C. M., Rudloff, C., Bauer, D., & González, M. C. (2013). Daily travel behavior: lessons from a week-long survey for the extraction of human mobility motifs related information. In *Proceedings of the 2nd acm sigkdd international workshop on urban computing* (pp. 1–7).

Szocska, M., Pollner, P., Schiszler, I., Joo, T., Palicz, T., McKee, M., ... others (2021). Countrywide population movement monitoring using mobile devices generated (big) data during the covid-19 crisis. *Scientific reports*, *11*(1), 1–9.

Thakuriah, P. (2001). Urban transportation planning: A decision-oriented approach. *Journal of Transportation Engineering*, *127*(5), 454–454.

Xu, X., Chen, A., & Yang, C. (2017). An optimization approach for deriving upper and lower bounds of transportation network vulnerability under simultaneous disruptions of multiple links. *Transportation research procedia*, *23*, 645–663.

Yabe, T., Jones, N. K., Rao, P. S. C., Gonzalez, M. C., & Ukkusuri, S. V. (2022). Mobile phone location data for disasters: A review from natural hazards and epidemics. *Computers, Environment and Urban Systems*, *94*, 101777.