



HAL
open science

Smart Card clustering to extract typical temporal passenger habits in Transit network. Two case studies: Rennes in France and Gatineau in Canada

Anne Sarah Briand, Etienne Come, Martin Trepanier, Latifa Oukhellou

► To cite this version:

Anne Sarah Briand, Etienne Come, Martin Trepanier, Latifa Oukhellou. Smart Card clustering to extract typical temporal passenger habits in Transit network. Two case studies: Rennes in France and Gatineau in Canada. 3rd International Workshop and Symposium: "Research and applications on the use of passive data from public transport", May 2017, Santiago, Chile. hal-01570726

HAL Id: hal-01570726

<https://hal.science/hal-01570726>

Submitted on 31 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Smart Card clustering to extract typical temporal passenger habits in Transit network. Two case studies: Rennes in France and Gatineau in Canada

Anne-Sarah Briand¹, Etienne Côme¹,
Martin Trépanier², Latifa Oukhellou¹

¹ Université Paris-Est, French Institute of Science and Technology for Transport, Development and Network (IFSTTAR), 14-20 Bd Newton 77420 Champs-sur-Marne, France

² Université de Montréal, Pavillon André-Aisenstadt, CIRRELT, CP 6128, Succursale Centre-ville, Montréal QC H3C 3J7 Canada

Smart card data are increasingly used for the analysis of urban mobility in particular multimodal and/or intermodal mobility. Their richness from spatial, temporal and quantitative point of view made them good material for mobility analysis. Indeed, they can be used to predict passenger flows, as well as to extract relevant information on transport network operation. They can also be used to better understand passenger behaviours. In the latter, statistical models are built to automatically extract a reduced set of clusters, each cluster contains a subset of passengers sharing the same temporal habits in terms of public transport usage.

In this paper we will investigate a statistical modeling to perform the clustering of passengers based on their ticketing logs in the public transport. The aim is partition the passengers into groups on the basis of their travel hours. The clustering is proposed to be performed in unsupervised way by using advanced data partitioning tools, that is dedicated Gaussian mixture models. Doing so, we will be able to extract typical patterns describing different types of transport usage, namely sporadic usage, typical home-work commute behavior, scholar usage etc.

Considering two case studies of the city of Rennes in France and the city of Gatineau in Canada, three main contributions will be illustrated in this paper:

- The first contribution is related to the statistical modeling based on a two-level generative model. Unlike different clustering approaches proposed in the literature that consider a discrete representation of time (e.g. trips are aggregated over pre-defined time slots), the proposed model allows to take into account the continuous nature of timestamps. The obtained distributions offer a direct way to interpret the results in terms of times of typical transport usage as well as in terms of variance around these time peaks of usage.
- The second contribution is related to the way to deal with the type of the day in the statistical model. It will be shown that this choice is closely linked to the public transport usage in the city. A categorization weekday/weekend is enough sufficient in Gatineau whereas a model accounting for all the days of the week is more appropriate to highlight the passenger habits in Rennes. This is mainly due to the temporal organization of scholars in France most of which do not study on wednesday afternoon. Moreover, this is also due to the

usage of public transport for nightlife activities or for leisure activities in the weekend.

- The third contribution is related to the analysis of different temporal activities that could be revealed by such method and link those activities with the city characteristics. By crossing the clustering results with spatial organization of both cities, it is indeed possible to have a better understanding of passenger activity on the transportation network. It will be shown that passengers, whom city is a centre of activity (Rennes) will be more likely to go out at night or during the weekend than passengers, whom city is located in the suburbs, which is the case of Gatineau. The results also highlight cultural differences in the use of public transportation between France and Canada.

Experiments are carried out on two real datasets. The first data set was collected on the public transport network of Rennes Metropole (France) and provided by the STAR network (Service de Transport en Commun de l'Agglomération Rennaise). The network is composed of one metro line (ligne a) and around 135 bus lines. 43 municipalities containing more than 400,000 inhabitants are serviced by this network. The second dataset was provided by the Société de Transport de l'Outaouais (STO) based in Gatineau, Canada. As a medium-sized public transit authority, the STO operates 310 buses and services 291,000 inhabitants. The STO has been operating its smart card system since 2001.

The obtained results exhibits different cluster profiles that can be used to differentiate between passengers with regular travel hours and those with diffuse travel hours. The approach is also able to detect subtle trends such as the presence of nightlife activity. The results provided by our method could help the transport operators to have a better knowledge of their passengers' behaviour patterns and adapt their offer to suite the different demands they portray (e.g. specific card type for passengers with regular temporal pattern). Moreover a more precise view of passengers' activity peak than classic aggregate time clustering is provided.

REFERENCES

- Agard, B., Morency, C. and Trépanier M.(2006). public transport user behaviour from smart card data. *The 12th IFAC Symposium on Information Control Problems in Manufacturing (INCOM)*, 17-19.
- Bagchi M. and White P. R. (2004). What role for smart card data from bus systems. *Proceedings of the Institution of Civil Engineers. Municipal Engineer. March Issue ME1*, **157**, 39-46.
- Barry, J. J., Newhouser, R., Rahbee, A. and Sayeda S. (2002). Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record*, 1817, 183-187.
- Briand, A.-S., Côme, E., El Mahrsi, M. K., Oukhellou, L. (2015). A mixture model clustering approach for temporal passenger pattern characterization in public transport. In: *IEEE International Conference on Data Science and Advances Analytics (DSAA), 2015*, , pp. 1-10.
- Chu, K. and Chapleau R. (2008). Enriching Archived Smart Card Transaction Data for Transit Demand Modeling, *Transportation Research Record: Journal of the Transportation Research Board*, 2063, 63-72.
- El Mahrsi, M. K., Côme, E., Baro, J. and Oukhellou, L. (2014). Understanding Passenger Patterns in Public Transit Through Smart Card and Socio-economic Data. *3rd International Workshop on Urban Computing (UrbComp), ACM SIGKDD Conference, New York, USA*, 9.
- Lathia, N. and Capra, L. (2011). How Smart is Your Smartcard?: Measuring Travel Behaviours, Perceptions, and Incentives. *Proceedings of the 13th International Conference on Ubiquitous Computing, UbiComp '11, Beijing, China*, 291-300.
- Lathia, N., Quercia, D. and Crowcroft, J. (2012). The hidden image of the city: sensing community well-being from urban mobility. *Proceedings of the 10th international conference on Pervasive Computing, Newcastle, UK*.
- Lathia, N., Smith, C., Froehlich, J. and Capra, L. (2013). Individuals among commuters: Building personalised transport information services from fare collection systems. *Pervasive and Mobile Computing*, **9**, 5, 643-664.
- McLachlan, G. J. and Krishnan, T. (2008). *The EM algorithm and extensions*, Wiley.
- Morency, C., Trépanier, M. and Agard, B. (2006). Analysing the variability of transit users behaviour with smart card data. *The Ninth International IEEE Conference on Intelligent Transportation Systems, Toronto, Canada, September*.

Trépanier, M., Tranchant, N. and Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Intelligent Transportation Systems*, **11**, 1-14.

Trépanier, M., Khandker M.N. and Morency, C. (2012). Are transit users loyal? Revelations from a hazard model based on smart card data. *Canadian Journal of Civil Engineering*, **39**, **6**, 610-618.

Utsunomiya, M., Attanucci, J. and Wilson, N. (2006). Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning. *Transportation Research* , 1971, 119-126.

Zhang, F. and Yuan, N. J. and Wang, Y. and Xie, X. (2014). Reconstructing individual mobility from smart card transactions: a collaborative space alignment approach. *Knowledge and Information Systems*, 1-25.