Representation Learning of public transport data. Application to event detection

Kevin Pasini, Allou Same, Mostepha Khouadjia, Fabrice Ganansia, Latifa Oukhellou

Representation Learning of public transport data. Application to event detection.

Kevin Pasini*[1,2], Allou Samé[1], Mostepha Khouadjia[2], Fabrice Ganansia[3], Latifa Oukhellou[1]

[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] IRT SystemX, Site du Moulon. Bât. 660 - Rue Noetzlin. 91191 Gif-Sur-Yvette, Saclay, France.
[3] SNCF- Innovation & Recherche, Immeuble Le Jade, Bureau B2-18, 1-3 avenue François Mitterrand. 93210 La Plaine St Denis, France.

* *Corresponding author: kevin.pasini@irt-systemx.fr*

Extended Abstract

A great deal of current research is focusing on mining large scale mobility data with exploratory, clustering or prediction purposes. Regarding public transport, numerous studies focus on prediction of travel demand [1] [2] [3] [4] [5]. This forecasting can be used for transportation planning when long-term time horizon is considered. It can also be used by public transport authorities and transport operators in the case of short/medium-term time horizon either to enrich public transport route planner or to help the co-regulation of the transport network.

On the basis of data collected by counting sensors deployed on trains, this paper deals with a forecasting of passenger load in public transport taking into account train operation. Providing passengers with train load forecasting, in addition to the expected arrival time of the next train, can indeed be useful for a better planning of their journeys, which can prevent over-crowding situations in the trains [6] [7]. The proposed approach is built on both a hierarchy of recurrent neural networks [8] and representation learning [9] with the aim to explore the ability of such mobility data processing to simultaneously perform a forecasting task and highlight the impact of events on the public transport operation and demand. An event refers here to an unexpected passenger transport activity or to a modification in transport operation compared to those corresponding to normal conditions. Two kind of historical data are used, namely train load data and automatic vehicle location (AVL) data. This latter source contains all information related to the train operation (delay, time of arrival/departure of vehicles ...). The proposed methodology is applied on a railway transit network line operated by the French railway company SNCF in the suburban of Paris. The historical dataset used in the experiments covers the period from 2015 to 2016.

Two contributions will be highlighted in this paper:

- The first contribution is related to the passenger load forecasting. Most of the researches carried out in this domain deals with the issue of prediction at an aggregated level, where the time step is considered to be fixed. We focus here on the forecasting at non-aggregated level taking into account real time train schedules. This induces an irregular time step on the train load series to be predicted. The prediction method we propose is able to take into account contextual factors such as calendar information (day, time, holiday ...) and train operation. Combining iteratively recurrent neural networks with representation learning on contextual factors allows us to perform prediction of the train load.
- The second contribution is to show how to explore the latent space provided by the neural network in order to highlight the impact of events on both the train operation and the train load. This new descriptive analysis is conducted by extracting from the initial latent space additional features according to which the data can easily be grouped into homogeneous clusters [10,11].

Following this line of research, an embedding of the input data combined with a clustering approach are used to learn efficient representation of mobility data, which allows to reflect the impact of event. Such approach is able to learn representation for mobility data while clustering them. Indicators are then defined to qualify the impact of the event.

The experiments are carried out on real data collected on a railway line operated by the French railway company SNCF. This line serves the north area of suburban of Paris and carries around 250.000 daily passengers. As it can be seen in Figure 1, this line can be divided into four branches serving different terminals. This transport network structure induces a complex railway operation of this line. The dataset includes both counting data of passengers boarding and alighting at each station and the timetable information. These heterogeneous sources of data lead us to reconstruct the passenger load of each train per inter-stations as shown on Figure 1. An in-depth analysis of the obtained results will be detailed in the paper to show the ability of the proposed approach to outperform classical machine learning models in terms of multi-step prediction performances. Moreover, indicators defined on the latent space will be exhibited and analyzed to quantify their capacity to highlight the impact of the events.

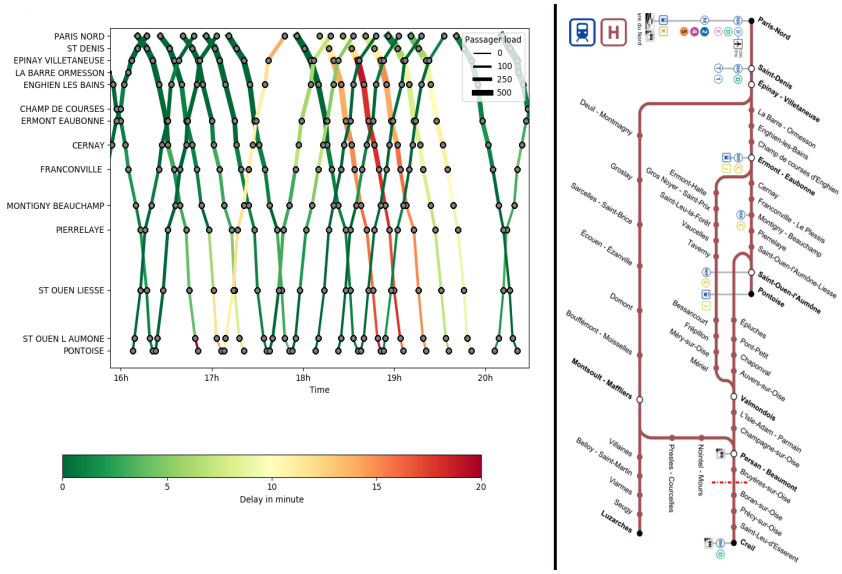

Fig. 1: Train table of the Pontoise branch and map of the line H (SNCF)

REFERENCES

[1] P. Wang, C. Wu, and X. Gao, "Research on subway passenger flow combination prediction model based on rbf neural networks and lssvm," in Control and Decision Conference (CCDC), 2016 Chinese. IEEE, 2016, pp. 6064–6068.

[2] F. Toqué, E. Côme, M. K. El Mahrsi, and L. Oukhellou, "Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks," IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 1071–1076, 2016.

[3] M. Ni, Q. He, and J. Gao, "Forecasting the subway passenger flow under event occurrences with social media," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 6, pp. 1623–1632, 2017.

[4] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," Transportation Research Part C: Emerging Technologies, vol. 77, pp. 306–328, 2017.

[5] L. Heydenrijk-Ottens, V. Degeler, D. Luo, N. van Oort, and J. van Lint, "Supervised learning: Predicting passenger load in public transport," 2018.

[6] I. Ceapa, C. Smith, and L. Capra, "Avoiding the crowds: Understanding tube station congestion patterns from trip data," ACM SIGKDD International Workshop on Urban Computing, pp. 134–141, 2012.

[7] N. Lathia and L. Capra, “Mining mobility data to minimise travellers’ spending on public transport,” 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1181–1189, 2011.

[8] K. Cho, B. Van Merri ̈enboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” EMNLP Empirical Methods in Natural Language Processing, p. 17241734, 2017

[9] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[10] Jiang, Z., Zheng, Y., Tan, H., Tang, B., and Zhou, H. (2017). Variational deep embedding: an unsupervised and generative approach to clustering, Proceedings of the 26th International Joint Conference on Artificial Intelligence.

[11] Yang, B., Fu, X., Sidiropoulos, N. D., and Hong, M. (2016). Towards k-means-friendly spaces: Simultaneous deep learning and clustering. ICML 2016.