TRANSITDATA 2019 Extended Abstract

Visualization tools for spatio-temporal time-series analysis with context awareness: Montreal subway case.

Florian Toqué · Etienne Côme · Latifa Oukhellou · Martin Trépanier

Abstract Forecasting passenger demand is of great interest for public transport operators. Despite the important role that forecasting play in mobility demand understanding, in-depth transport oriented analysis of the forecasting results is often overlooked, since it raised some challenges. In this context we developed two visualization tools with open source frameworks that allow to analyze spatio-temporal time-series forecasting with context awareness. The first visualization tool allows to analyze the forecasting results over large period in all the stations and to zoom in for more precise temporal details. The other tool allows to better understand the passenger demand relations between the different stations of the transport network, and enable a spatial analysis of the results. Analyzed time-series corresponds to the forecast results of the number of passengers entering each station with a fine-grained temporal resolution (15 minutes interval) during one year achieved with a well-known machine learning model, a Random Forest. In order to highlight the spatio-temporal specificity of the passenger demand, we have computed and analyzed the residues of a long-term forecast model that returns normal passenger demand. Here we show that both visualization tools depict the stations and the period hard to predict and allow to have an insight on which contextual element (weather, event on the city and incident on the transport network) could impact the forecasting. Experiment are performed with real data given by the transport authority of Montreal (Société de transport de Montreal, STM).

Keywords Visualization · Spatio-temporal time-series · Smart Card Data · Passenger Demand Forecasting · Public Transport · Event and incident · Weather

Florian Toqué, Etienne Côme, Latifa Oukhellou
Université Paris-Est, IFSTTAR, COSYS-GRETTIA
Champs-sur-Marne, France
E-mail: florian.toque@ifsttar.fr, etienne.come@ifsttar.fr, latifa.oukhellou@ifsttar.fr

Martin Trépanier
Ecole Polytechnique de Montréal, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport (CIRRELT)
Montréal, Canada
E-mail: mtrepanier@polymtl.ca

1 Introduction

In the past few years we could observed an increasing need for travel, in urban and peri-urban area. In the meantime, transport operators have seen a good opportunity to better understand the mobility with the emergence of data science and the availability of more and more digital traces. In this context, transport operators wish to offer the best quality of service in order to convince the population to choose public transport for their daily trips. Today, one of the main challenges for transport operators are to analyze, understand and predict the mobility with these digital traces.

Using smart card data to analyze mobility in public transportation has received substantial amount of attention from researchers as we can see in the work of Pelletier et al. (5). Interactive visualization is of great interest for transport operators because it can help them to analyze passenger behavior. In (6) the authors have focused their work on visualization of passenger flows (origin-destination) and social media data (tweets) in order to easily analyze the variation of passenger behavior in case of complaints of the passengers about transport services. In the same line of work the authors of (4) have created a visualization framework for large scale train trips record analysis. In both previous work (6; 4), visualizations consist on; one heatmap that provides a temporal overview of unusual phenomena and where the temporal aggregation could vary (daily or monthly) and one map that shows spatial overview with passenger flows depicted by oriented ribbon. In (3) the authors propose a web interface to analyze bus passenger use and the load profiles of the routes taken by users in specific areas of the network.

In the same line of work of (6; 4; 3), we propose interactive spatio-temporal visualization tools with the goal to analyze forecasted passenger demand at each node of a transport network. In order to let this work available to the public transport community we developed the visualization tools with frameworks that can be used in open source (Data-driven Documents, D3 (1) and Mapbox). Visualization tools are described in Section 2, then we depict data case study and the forecasting results used in the visualization in Section 3.

2 Spatio-temporal visualization tools

Here we developed two visualization tools to simplify the transport-oriented analyses of spatio-temporal time-series forecasting located in nodes of the transport network that do not overlap.

2.1 Analysis with temporal focus

In order to analyze temporally the forecasting results over all the stations we developed a heatmap allowing to highlight the period and stations hard to predict. Interactive selection allows to choose the forecasting results to analyze, the station to show and the type of normalization to apply. Furthermore, it is possible to select the time aggregation between year mode and day mode and the type of time aggregation such as sum, average and max. Web interface allows "mouse over" functionality in this type of visualization. We use it to show some information in a dialogue box about the period and station selected such as the presence of event or incident. An example of this visualization tool is depicted in Figure 1.

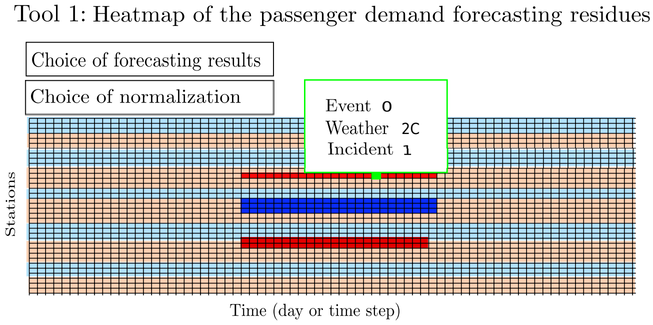

Fig. 1 Heatmap of the passenger demand forecasting residues with event, incident and weather information. High intensity corresponds to high residues (blue less passenger than normal, red more passenger than normal situation). This visualization allows to see duration of bad forecasting, to know which station is impacted and to have a great insight on the cause (event, incident or special weather).

2.2 Analysis with spatial focus

This spatio-temporal map allows to analyze the residuals at each time-step and at each station or node of the transport network. Incident could be localized spatially with circles and event are depicted by heatmap point around the impacted station. Selection of the day is possible with a calendar. Mouse over functionality shows in-depth information about event and incident on the selected station. It is possible to select the time-step with a slider in order to temporally analyze the passenger demand forecasting. Forecasting results could be selected as well as the normalization of forecasting results. This visualization focused on spatial aspect is detailed in Figure 2.

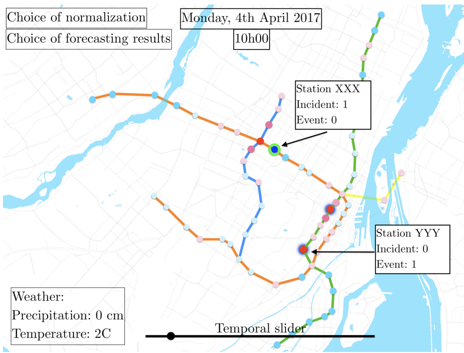

Fig. 2 Forecasting residues visualization with spatial focus. Event, incident are depicted per station. Weather information is given for the city.

3 Experiments

We used a real data case and a state of the art forecasting algorithm that predict the normal passenger demand depending on the type of day.

3.1 Case study

We use four real data sources to analyze the forecasting of passenger demand at each station of a transport network. Smart card data logs (tap-in) with a temporal resolution of 15 minutes, event database (e.g., concert, show, hockey game, etc.) and incident of the transport network database were provided by the transport authority of Montreal, QC, Canada. We also used weather database downloaded from the Canadian government website.

3.2 Forecasting

We analyze the results of two regression forecasting model that consider information about the day predicted. Both models are Random Forest (2) which gives the best forecast accuracy on this data-set compared with other regression models. Such forecasting corresponds to passenger demand in normal situation and allows to highlight the abnormal phenomena of passenger demand.

Acknowledgements The authors wish to thank the transport organization authority of Montreal (Société de Transport de Montréal, STM) for graciously providing the smart card data, the event and the incident database.

References

1. Bostock, M., Ogievetsky, V., Heer, J.: D data-driven documents. IEEE Transactions on Visualization and Computer Graphics **17**(12), 2301–2309 (2011). DOI 10.1109/TVCG.2011.185
2. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
3. Giraud, A., Trépanier, M., Morency, C., Légaré, F.: Data fusion of apc, smart card and gtfs to visualize public transit use. Tech. rep., CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport (2016)
4. Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., Kitsuregawa, M.: Visual fusion of mega-city big data: an application to traffic and tweets data analysis of metro passengers. In: 2014 IEEE International Conference on Big Data (Big Data), pp. 431–440. IEEE (2014)
5. Pelletier, M.P., Trépanier, M., Morency, C.: Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies **19**(4), 557–568 (2011)
6. Yokoyama, D., Itoh, M., Toyoda, M., Tomita, Y., Kawamura, S., Kitsuregawa, M.: A framework for large-scale train trip record analysis and its application to passengers flow prediction after train accidents. In: Pacific-Asia conference on knowledge discovery and data mining, pp. 533–544. Springer (2014)