Prediction of bus passenger flow using Deep Learning

Walid Kheriji, Sami Kraiem, Guilhem Sanmarty,
Ghazaleh Khodabandelou and Fouad Hadj Selem

Vedecom, 23 bis Allée des Marronniers, 78000 Versailles
first name.last name@vedecom.fr

Abstract. Public transport is important for urban sustainability. Buses are the most commonly used mode of public transport in cities, and are particularly crucial in large areas such as Île-de-France, where millions of people travel by bus every day. Predicting bus passenger flow is very important to urban transportation development, which facilitates the decision-making processes for local authorities and transport operators and provide citizens with an efficient and safe travel experience. In this work, we propose a deep learning model to predict bus passenger flow between each potential couple of origin and destination stops of a given bus network based on land use, point-of-interest (POI) and smart card data.

In recent years, many researchers have developed effective models for predicting bus passenger flows. Liu et al. [1] proposed an unsupervised training model based on a stacked autoencoder (SAE) combined with a supervised training model based on deep neural network (DNN) to predict hourly passenger flow. The input features used for their model are the temporal features including the day of a week, the hour of a day, and holidays, the scenario features including direction (inbound and outbound) and type (tickets and cards), and the passenger flow features including the previous average passenger flow and real-time passenger flow. The proposed SAE-DNN can predict the hourly passenger flow. However, it cannot capture spatial information of traffic flow well and requires the estimation of the previous flow. To overcame this limit, other researchers [2] propose a new method to capture both spatio-temporal correlation and specific scenario patterns of bus traffic flow. They used transportation smart card data which contain about 177 million records of passengers bus transaction information. Each record contains the following key attributes: card id, bus route id, bus vehicle code, boarding and alighting time, latitude and longitude coordinates of boarding and alighting stations. However, those models do not include any land use data into the learning process (such as population size, demography of the different socio-professional categories, number of hospitals, malls, companies, etc.). In this way, the optimized models, may not be used in different cities and especially for a new stops project. In fact, it is widely expected that there are strong regularities and similarities in human mobility patterns, and it is possible to learn the dependence between the bus passenger trip flow and the urban context data. Moreover, researchers have pointed out that urban mobility results from the interaction of socio-economic indicators are similar in all countries [3, 4]. These results suggest that it is scientifically possible to predict different mobility flows using demographic and socio-economic indicators around origin and destination stations as independent variables in the learning process. To the best of our knowledge, the only paper that includes this kind of features is [5]. The proposed model is based on artificial neural network including, as input variables, each traffic zone proportion of residential, commercial, and industrial traffic and the area in the bus passenger trip flow forecast.

In this paper, we are motivated by the wish to extend the approach developed by Yu et al. [5] to more general POI data including more than 500 features. In addition, we add key features as theoretical travel time using different transport mode (particular vehicle, motorcycle, public transport and walk). Furthermore, unlike [5], we propose to include the temporal structure of the passenger flow in the learning model using smart card data. Although the smart card data do not have the alighting locations, we use a trip chaining model [6] for estimating 65% of passengers' destinations. To infer the missing destinations we use a probability method (we can alternatively use deep learning model [7] [8]). Once this information is completed, we obtain the time series target variable of length n (number of slots during a day) for each pair of origin and destination stops. It is worth noticing that we are concerned with accurate estimation at a given day and a general trend at a given time slot. Thus, at each time slot, the Y variables are estimated by meaning the flow values at the same time slot during several weeks. On the other hand, the input matrix X contains three groups of variables. The first two are descriptive urban variables for the origin and destination stops. This includes points of interest, populations estimates, demographic estimates of the different socio-professional categories, number of hospitals. It should be noted that these contextual data are collected at the IRIS level (Grouped Areas for Statistical Information), which is the most detailed geographical statistics available in France. The last group consists of a theoretical estimate of the travel time between the two stations by public transit or personal vehicle. Obviously, they are useful predictors and they are collected using the online API service "Navitia".

Finally the feasibility and reliability of our model will be tested with the data of 26 towns located in southern Île-de-France and which served by the Sénartbus network. The main contributions expected from this work are first of all to improve the scores obtained in the reference by using the urban variables added to the models. Indeed, we use 500 variables instead of 5 used in the reference. Secondly, to test more appropriate learning models that may include, among other things, spatial or temporal structures. Thirdly, identify the key variables describing the internal mechanism of mobility via public transport. For example, the percentages of particular socio-professional categories may be directly related to the use of this mode of transport. Differences between the theoretical travel times by different modes of transport should also be analyzed. The results of this work may lead to future research. Indeed, by being able to estimate the flows between each pair of bus stops, then it would be possible to optimize the combinations of stops forming a bus line in order to respect the different constraints and then optimize the performance of a particular line or the global network. Thus, the estimates from our model can be used as input data for numerical optimization algorithms as a decision support tool.

Keywords: Deep-learning, Passenger Flow, Prediction, Origin - Destination matrix, public transport

References

1. Lijuan Liu and Rung-Ching Chen. A novel passenger flow prediction model using deep learning methods. *Transportation Research Part C: Emerging Technologies*, 84:74 – 91, 2017.
2. Panbiao Liu, Yong Zhang, Dehui Kong, and Baocai Yin. Improved spatio-temporal residual networks for bus traffic flow prediction. *Applied Sciences*, 9(4), 2019.
3. Rémi Louf, Camille Roth, and Marc Barthelemy. Scaling in transportation networks. *PLoS One*, 9(7):e102007, 2014.
4. Rémi Louf and Marc Barthelemy. How congestion shapes cities: from mobility patterns to scaling. *Scientific reports*, 4:5561, 2014.
5. Y. Shaoqiang, S. Caiyun, Y. Yang, Z. Shuyuan, and Y. Wenlong. Prediction of bus passenger trip flow based on artificial neural network. *Advances in Mechanical Engineering*, 8(10):1687814016675999, 2016.
6. R. Chapleau M. Trépanier and N. Tranchant. Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems*, 11:1–14, 2007.
7. 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. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1071–1076, Nov 2016.
8. J. Jung and K. Sohn. Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. *IET Intelligent Transport Systems*, 11:334–339, 2017.