Supervised learning
Predicting passenger load in public transport

Heydenrijk-Ottens, Leonie; Degeler, Viktoriya; Luo, Ding; van Oort, Niels; van Lint, Hans

Publication date
2018

Document Version
Accepted author manuscript

Published in
Proceedings of Conference on Advanced Systems in Public Transport (CASPT) 2018

Citation (APA)
Heydenrijk-Ottens, L., Degeler, V., Luo, D., van Oort, N., & van Lint, H. (2018). Supervised learning: Predicting passenger load in public transport. In *Proceedings of Conference on Advanced Systems in Public Transport (CASPT) 2018: 23-25 July, Brisbane, Australia* Article 76

CASPT 2018 Extended Abstract

Supervised learning: Predicting passenger load in public transport

Léonie Heydenrijk-Ottens · Viktoriya Degeler · Ding Luo · Niels van Oort · Hans van Lint

Abstract In this extended abstract, we show the supervised learning approach to predicting passenger load of trams, based on historical passenger load patterns. We look at two different cases: predicting long-term passenger load of any given day and time, and predicting short-term passenger load at a particular public transport vehicle.

Keywords: Public transport · passenger load · supervised learning · prediction

1 Introduction

For many Public Transport (PT) users, overcrowding in PT vehicles has a major decreasing effect on the comfort experience (Li & Hensher, 2013). However, most online routing applications still not take comfort regarding to crowdedness into account, but provide recommendations based on shortest distance, shortest travel-time, or number of interchanges (Campigotto et al. 2017).

Being able to include certain information on crowdedness, requires knowledge about the current and future level of passenger load. Increasing amount and complexity of data describing public transport services allows us to better explore the detection methods and analysis of different phenomena of PT operations. Some countries or operators provide the possibility to use Smart Card (SC) data for occupancy prediction (van Oort et al. 2015). However, SC data is not available in real time (van Oort et al. 2016), which makes it hard to incorporate it into real time recommendation models. In this paper, we show that it is possible to predict the passenger load via supervised learning, eliminating the need for fare collection data beyond the set needed for training.

Léonie Heydenrijk-Ottens · Viktoriya Degeler · Ding Luo · Niels van Oort · Hans van Lint
Department of Transport and Planning, Delft University of Technology
Delft, the Netherlands
Email: {L.J.C.Heydenrijk-Ottens-1, V.Degeler, D.Luo, N.vanOort., J.W.C.vanLint}@tudelft.nl

2 Data sources

Our study concerns three datasets. Static GTFS data provides information about the transportation network geographical structure, stops, routes, and schedules. Furthermore, we employ two dynamic data sources: Automatic Vehicle Location, AVL (Hickman, 2004) and Automated Fare Collection, AFC (van Oort et al. 2016) data. AVL contains actual times of arrival/departure of vehicles, headways, delays, etc. Delays can be represented as a negative value, which implies the arrival is ahead on schedule. AFC includes the tap-in / tap-out times of personalized smart cards and the exact vehicles in which these transactions happened. Since in the Netherlands, the smart card (*OV-chipkaart*) is extremely prevalent over other types of payment, the tap-in and tap-out times of the smart cards can be used to estimate the passenger load of a vehicle. Luo et al. (2018) describe, how the load profiles were computed for this dataset.

3 Classification of passenger load

3.1 Case study: The Hague Public Transport Network

For this study, we used the public transport network of The Hague, the Netherlands, which consists of 12 tram lines and 8 bus lines. The dataset covers the period of the month March of 2015.

3.2 Data preparation

We prepared the dataset by eliminating rows with missing data as well as rows where the AVL departure time was before the AVL arrival time. Outliers in passenger load and AVL arrival / departure delay are filtered, where an outlier is defined to differ thrice the standard deviation or more from the variable mean over the whole month. The day of the week (as an integer, Monday being 0, extracted from the date) is added, as well as the AVL departure delay at the previous stop of the current trip (avlPreStopDepartureDelay), and AVL departure delay at the current stop of the previous trip (preTripAvlDepartureDelay). All features are independently standardized to a standard normal distribution with zero mean and unit variance.
Finally, we retain the following features: The date, day of week, the stop number in the stop-sequence, direction-ID, GTFS arrival time to second precision, GTFS trip-ID, stop-ID, the GTFS trip-ID of the previous trip that addressed this stop, preTripAvlDepartureDelay and avlPreStopDepartureDelay.

Considering tram line 3, cleaning approximately delete 10% of the data, leaving us with around 375 thousand stop records of March 2015.

The dataset contains 31 days, of which 22 working days. We considered working days and weekend days separately due to distinctly defined different passenger load patterns, as shown in Figure 1.

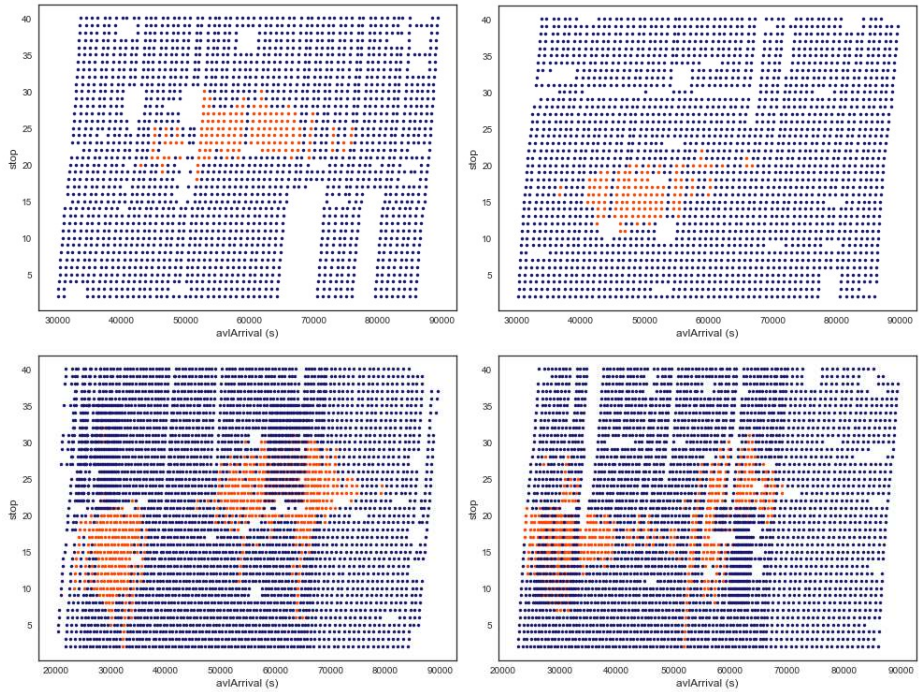

Fig. 1 Passenger load pattern of line 3: seat capacity 86, occupancy threshold 63. Red dots represent load above the threshold, blue dots – below the threshold. (a) March 1st, Sunday, direction 0, (b) March 1st, Sunday, direction 1, (c) March 3, Tuesday, direction 0 (d) March 3, Tuesday, direction 1.

Therefore, the dataset is divided into two sets: data_weekend and data_week. For calibration of the predictors' parameters and the evaluation of the predictors performance, both datasets are split into a training set (data_week: first 17 wee-days, data_weekend: first 7 weekend-days) and a test set (data_week: last 5 week-days, data_weekend: last 2 weekend-days).

3.3 Methodology

Time step. Passenger load patterns are quite stable over time, see, for example, Figure 1. Barring the cases of severe disruptions, we can observe very similar passenger load patterns on same lines and same days of the week. This allows us to use supervised classification for passenger load prediction. We distinguish two prediction target stages, each employing a different set of features:

1. Long term prediction: predicting the load of any given day - with a feature set consisting only of static GFTS data.
   Features = [day, sequence, gtfsArrival, gtfsTripID, preTripID, stopID]

2. Short term, next stop prediction: Predicting the load of the current stop of the next trip – with static features as well as the AVL departure delay for the considered stop of the previous trip and the departure delay of the current trip at the previous stop.
   Features = [day, sequence, gtfsArrival, gtfsTripID, preTripID, stopID, preTripAvlDepartureDelay, avlPreStopDepartureDelay]

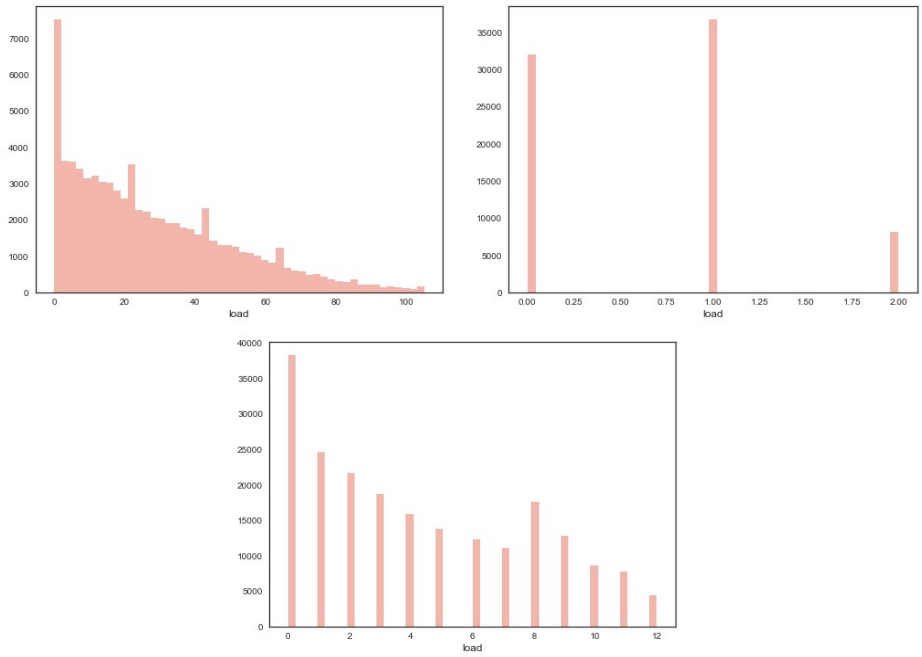

Fig. 2 (a) Passenger load distribution of direction 1 of line 3, seat capacity of 86, March 2015 (weekdays). (b) Number of observations within each passenger load class. Class 0 is low passenger load (max. 20% seat occupancy); 1 is medium load (between 20-70% seat occupancy); 2 is high load (more than 70% seat occupancy). (c) Number of observations within each passenger load class. Up till 50% seat occupancy, load is grouped per 5 passengers, between 50-100% seat occupancy, load is grouped per 10 passengers.

Class definition. The passenger load is manually labelled into classes, which are defined in two different ways. First, the class choice is based on the capacity of the considered vehicle. I.e., the load is defined to be low if a maximum of 20% of the seats is occupied; medium if between 20-70% of the seats is occupied; high if more than 70% of the seats is occupied.

The downside to this definition is that it results in an unbalanced training set, as high passenger loads occur less often than low or medium ones, see, for example, Figure 2 for the distribution of the load of line 3. Moreover, when considering only three classes, small mistakes have a large impact. To confront the balance problem and explore whether the classification can be more robust, we decompose the larger classes (in terms of occurrence) into smaller classes. The second class composition is as follows: Up till 50% seat occupancy, we grouped the load per 5 passengers, between 50-100% seat occupancy, the load is grouped per 10 passengers. For line 3 (86 seat capacity), this results in 13 classes (Figure 2c): [0,5), [5,10),…, [35,40),[40,50),…[70,86), [86,200). In the full article we will present a more detailed strategy to deal with the unbalanced data set, including oversampling the under-represented classes.
We refer to (Yap et al. 2017) for an overview of the seat capacity and standing surface per vehicle type.

Classifiers. For predicting the passenger load, we compare four classifiers: A random forest classification (RFC) model (Breiman et al. 2001), gradient boosting classifier (GBC) (Friedman et al. 1999, Hastie et al. 2001), multi-layer perceptron (MLP) classifier (Chaudhuri et al. 2000, Gardner et al. 1998) and a k-nearest neighbours (KNN) classifier (Aha et al. 1991, Weinberger et al. 2009). In the full article we will cover (the motivation for) the different classifiers, the (tuned) parameter values and time performance.

Metrics. For each dataset and classifier, performance measures will be calculated. The overall average percentage score for each classifier will be calculated by averaging accuracy and F1-measure $F1 = 2 \times (Precision \times Sensitivity)/(Precision + Sensitivity)$, with $Sensitivity = tp/(tp + fn)$, and $Precision = tp/(tp + fp)$.
For the second class partition (per 5-10 passengers), we consider a prediction to be a true positive if the distance to the true label is less than 2. This yields a relaxed interpretation of the average accuracy function:

$$relaxed\ accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} f(y_i, \hat{y}_i),$$

Where $$f(y, \hat{y}) = \begin{cases} 0\ if\ |y - \hat{y}| \geq 2 \\ 1\ if\ |y - \hat{y}| < 2 \end{cases}.$$

4 Discussion/Preliminary results

Tuning of the machine learning parameters and other ways to pre-process the data (e.g., using different imputers to fill the missing data with meaningful values instead of removing a data point) will be further explored in the full paper.

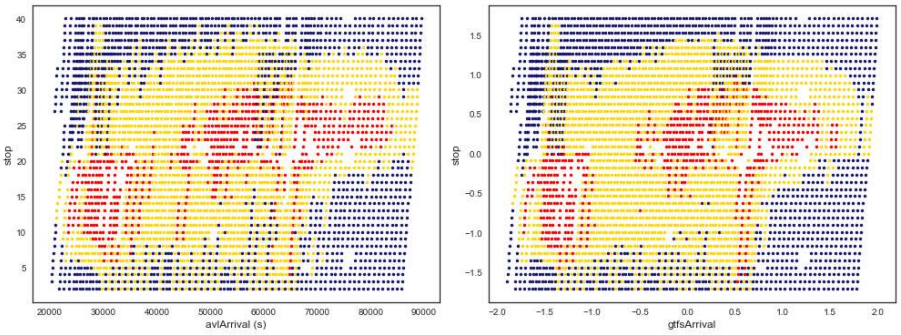

Fig. 3 (a) Visualization of the passenger load pattern of line 3, direction 0, March 26, Thursday. Colours represent the load values: dark blue [0,17), yellow [17,60), red [60,105]. (b) RFC prediction.

Nevertheless, the classifiers achieved results as can be seen in Table 1-4. For line 3, Random Forest Classifier seems to render the most compelling predictions. Furthermore, AVL information about previous stop and trip only seems to have little contribution in the predictive value of the algorithms.

Table 1 Short-term 3-class prediction, line 3, data_week. Acc.: Accuracy, F1: F1-measure, Avg.: Average score in %, MeanE.: Mean error. Measures >80% are marked in bold.

Classifier	Acc.	F1	Avg.	MeanE.
RFC	**83**	**83**	**83**	0.2
GBC	78	77	78	0.2
MLP	**81**	**81**	**81**	0.2
KNN	77	77	77	0.2

Table 2 Long-term 3-class prediction, line 3, data_week. Acc.: Accuracy, F1: F1-measure, Avg.: Average score in %, MeanE.: Mean error. Measures >80% are marked in bold.

Classifier	Acc.	F1	Avg.	MeanE.
RFC	**81**	**81**	**81**	0.2
GBC	78	76	77	0.2
MLP	80	79	80	0.2
KNN	**81**	80	**81**	0.2

Table 3 Short-term 13-class prediction, line 3, data_week. Rel.Acc.: Relaxed accuracy, Pr: Precision, Rec.: Recall, F1: F1-score, Kappa: Cohen's Kappa score, Avg.: Average score in %, MeanE.: Mean error (Non-relaxed). Measures >70% are marked in bold.

Classifier	Rel.Acc.	F1	Avg.	MeanE.
RFC	**74**	**74**	**74**	1.1
GBC	66	64	65	1.4
MLP	**71**	**71**	**71**	1.2
KNN	60	59	60	1.6

Table 4 Long-term 13-class prediction, line 3, data_week. Rel.Acc.: Relaxed accuracy, Pr: Precision, Rec.: Recall, F1: F1-score, Kappa: Cohen's Kappa score, Avg.: Average score in %, MeanE.: Mean error (Non-relaxed). Measures >70% are marked in bold.

Classifier	Rel.Acc.	F1	Avg.	MeanE.
RFC	**72**	**71**	**72**	1.2
GBC	66	64	65	1.4
MLP	67	66	67	1.3
KNN	67	66	67	1.3

It can be hypothesized that more data, with improved pre-processing and tuning of the models' parameters, renders more accurate predictions. In addition, more data can give a better insight in the correlation between passenger load and AVL data, like the exact departure delay at the previous stop or at the same stop measured from the previous trip. With this approach one can also look into the influence of the delay of other (alternative) lines on the load of one line.

Acknowledgements: This research was supported by H2020 project My-TRAC (Grant No. 777640). We also would like to thank HTM and Stichting OpenGeo for providing the AFC and AVL datasets, respectively.

References

Luo D., Bonnetain L., Cats O. and van Lint H. (2018). Constructing Spatiotemporal Load Profiles of Transit Vehicles with Multiple Data Sources. Proceedings of the 97th Transportation Research Board Annual Meeting, Washington DC.

Campigotto, P., Rudloff, C., Leodolter, M., & Bauer, D. (2017). Personalized and situation-aware multimodal route recommendations: the FAVOUR algorithm. IEEE Transactions on Intelligent Transportation Systems, 18(1), 92-102.

Van Oort, N., Brands, T., & de Romph, E. (2015). Short term ridership prediction in public transport by processing smart card data. Transportation Research Record, No. 2535, pp. 105-111.

Breiman L. Random Forests. Machine Learning 2001;45:5–32.

Friedman, J. (1999). Greedy function approximation: the gradient boosting machine, Technical report, Stanford University.

Hastie, Trevor, Robert Tibshirani, and J Jerome H Friedman. The Elements of Statistical Learning. Vol.1. N.p., page 339: Springer New York, 2001.

Chaudhuri BB, Bhattacharya U. Efficient training and improved performance of multilayer perceptron in pattern classification. Neurocomputing. 2000;34:11–27.

Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron), A review of applications in the atmospheric sciences. Atmos Environ. 1998;32:2627–2636.

Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. Mach Learn. 1991;6:37–66.

Weinberger K, Blitzer J, Saul L. Distance metric learning for large margin nearest neighbor classification. J Mach Learn Res. 2009;10:207–244.

Yap, M., Cats, O., Yu, S., & van Arem, B. (2017). Crowding valuation in urban tram and bus transportation based on smart card data.

Li, Z., & Hensher, D.A. (2013). Crowding in Public Transport: A Review of Objective and Subjective Measures. Journal of Public Transportation, Vol. 16, No. 2, pp 107-134.

Van Oort, N., T. Brands, E. de Romph, M. Yap (2016), Ridership Evaluation and Prediction in Public Transport by Processing Smart Card Data:A Dutch Approach and Example, Chapter 11, Public Transport Planning with Smart Card Data, eds. Kurauchi F., Schmöcker, J.D., CRC Press.

Hickman, M. (2004). Evaluating the benefits of bus automatic vehicle location (AVL) systems. In: Levinson, D., D. Gillen (eds). Assessing the benefits and costs of intelligent transportation systems. Kluwer, Boston.