USING AVL DATA TO PREDICT BUNCHING CONSIDERING SENSITIVITY AND SPECIFICITY

Wenzhe SUN[1], Jan-Dirk SCHMÖCKER[2]

[1]Dept. of Urban Management, Kyoto University
(Katsura, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail:wz.sun@trans.kuciv.kyoto-u.ac.jp
[2]Member of JSCE, Dept. of Urban Management, Kyoto University
(Katsura, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail:schmoecker@trans.kuciv.kyoto-u.ac.jp

Public transport is expected to deliver reliable services, while unplanned long dwell time resulting from high boarding/alighting passenger demand and in-vehicle overcrowding may generate significant delay, which can be frequently observed on bus as well as rail lines. The delay accumulated over successive stops may considerably shorten the headway between the delayed service and the following one. A headway being shorter than a threshold is defined as public transport bunching in this paper, and it may force the operator to slow down or hold the following service. This paper proposes a methodology to predict the bunching multiple stops in advance. A logistic regression model taking dwell time and headway at upstream stops as the predictor variables is used to derive the probability of the headway being below a threshold at a distanced downstream stop. Furthermore, how to select the proper cut-off point for the predicted probability is discussed. Bus AVL (Automatic Vehicle Location) data collected in Kyoto City is used to verify the methodology. This paper also discusses similarities and differences to be considered by rail operators for predicting delays and bunching.

Key Words: *public transport bunching, bunching prediction, logistic regression, sensitivity and specificity*

1. INTRODUCTION

Bunching is a frequently occurring undesired event for public transport. Generally it can be defined as the phenomenon of two successive public transport runs (PTRs) of a single line arriving at a stop within significantly shorter headways than the designed one. Bunching involving more than two PTRs is also regularly observed. PT bunching may be initiated by the arrival of one PTR being delayed at an upstream stop. More passengers are likely to accumulate for the delayed run at that stop and it is thus further delayed. Conversely, the subsequent run has fewer passengers to pick up and departs earlier than scheduled. Accumulated delay to the first run and increasingly earlier arrival of the second one result in obvious inequality in dwell times and on-board passenger numbers. As the inequality aggravates over a sequence of stops, the scheduled headway is significantly shortened or eventually offset and the leading run among bunched run is often overcrowded.

Accurate prediction on headway or bunching itself can help to spotlight the coming bunching and further assist the operator to eliminate bunching in real time. A useful prediction tool is expected to a) have a long enough prediction horizon to allow the operator's implementation of countermeasures and b) provide information on the reliability of the prediction. The latter point is important in order to account for different preferences among operators. A bunching-averse operator is willing to frequently control the service to avoid any possible bunching, whereas some other operators may hesitate to take control action that will negatively impact some passengers, they thus only correct the predicted bunching of high confidence level. Therefore, this paper suggests a probabilistic binary prediction method.

This study aims to extend the existing literature in two aspects. Firstly, this study builds a LOGR (Logistic Regression) model to predict the likelihood of bunching to occur using bus GPS data, and tests the prediction performance under a wide range of prediction horizons varying from 1-stop-ahead to 15-stop-ahead, with an emphasis on multi-stop-ahead prediction and understanding the regularity deterioration pattern. Secondly, this study tries to enhance the robustness and flexibility for existing prediction tools. To achieve this ROC (Receiver Operator Characteristic) curves are utilized. This method is widely used in evaluating the performance of binary classification models and in this study it is interpreted as the optimal front of the proposed LOGR. This study explains how to conduct the trade-off between "sensitivity" and "specificity" from an operator's perspective.

The paper is organized as follows. After this introduction, Section 2 contains a literature review concerning the PT bunching problem and headway prediction. The predictive methodology using logistic regression is elaborated in Section 3. Then two headway-predicting algorithms: LR (Linear Regression) and SVM (Support Vector Machine) are taken as the two benchmark approaches in this study and are also briefly introduced in this section. In Section 4, the characteristics of the collected data are described, including data collection period, average stop-to-stop travel time, average scheduled headway, fluctuation patterns for headway, etc. Based on this, a proper prediction horizon and bunching threshold are determined. In Section 5, the prediction performance of the proposed LOGR is evaluated and compared with headway-based methods. The trade-off functionality of LOGR is discussed in Section 5 as well. Conclusions and potential application on railway transit can be found in Section 6.

2. LITERATURE REVIEW

Most of the relevant existing literature can be cast into two categories according to their objective: bunching prediction and corrective strategies. Bus transit system is more vulnerable to bunching, and a large body of literature discussed how to eliminate bus bunching using analytical or simulation methods following the seminal work by Newell and Potts (1964). Osuna and Newell (1972) and Newell (1974) tried to maintain the bus schedule by a single control point. On the other hand, advanced control methods such as dynamic holding control proposed by Eberlein et al (2001), Daganzo (2009), Xuan et al (2011), Bartholdi and Eisenstein (2012), Zhang and Lo (2018) and velocity control developed by Daganzo and Pilachowski (2011) as well as stop skipping discussed by Sun and Hickman (2005) assume frequent and efficient communication between bus drivers and the control center. Berrebi et al (2018) tested the control strategies proposed by Dagazo (2009), Xuan et al (2011), Bartholdi and Eisenstein (2012), Daganzon and Pilachowski (2011) on a bus route in Portland, Oregon. The experiment was based on real bus AVL (Automatic Vehicle Location), APC (Automatic Passenger Counter) and traffic signal data. The effectiveness of each strategy to stabilize bus headways was confirmed. Further, the effect of incorrect future headway prediction on each strategy was discussed. The variance of controlled headway was found rising significantly as the prediction errors increased. Instead of actively adjusting the headway, Schmöcker et al (2016), Wu et al (2017), Sun and Schmöcker (2018) discussed passive strategies such as passenger re-distribution and overtaking which are activated when bunching occurs. These strategies aim to equalize passenger boarding numbers for bunched buses through queue management.

Substantial development in data collection technology recently gives scholars the access to massive public transit data including AVL, APC and AFC (Automatic Fare Collection) data, and has led to a large number of studies concerning real-time prediction on operational aspects. Rather than predicting bunching events, most existing literature focuses on the arrival time and headway. Hans et al (2015) developed a sequential mesoscopic simulation which elaborately considered the stochastics generated during bus dwell time and link travel time. A bundle of possible future trajectories is obtained based on the distribution assumed for associated parameters, delivering robust prediction results to the operator. Distribution or range for future arrival time and headway can also be easily obtained. A shortcoming of this method is that the predicted range of arrival time or headway might be too wide to be conclusive for operators' decision making. Yu et al (2016) conducted a solid literature review on the methods addressing bus arrival time prediction. They reviewed the implemented data source and algorithm of each relevant literature. SVM, KF (Kalman Filter), KNN (K-Nearest Neighbor), ANN (Artificial Neural Network) and regression-based methods are frequently used. Yu et al (2011) used SVM, ANN, KNN, and LR to predict arrival time for a 0.7km common line section where more than 10 bus routes overlapped in Hong Kong. Future headway is the difference between the predicted arrival times of two consecutive buses and can be obtained by arrival time prediction method. There are also some studies directly focusing on the prediction of headway itself. Yu et al (2017) proposed a probabilistic prediction approach using RVM (Relevance Vector Machine) to attach a confidence interval for each predicted headway for

2- and 3-stop-ahead. Outperformance with respect to robustness was concluded by comparing the results with the deterministic single values derived by SVM, KF, KNN and ANN algorithms. Andres and Nair (2017) integrated headway prediction and bus holding control strategies. Regression, ANN and autoregressive models are used in their work to predict future headways with 5min and 10min prediction horizons. The prediction results are applied as input to an analytical model extending Daganzo (2009).

Although headway prediction methods have made great advancement, it remains a challenging work to successfully identify coming bunching events in multiple-stop-ahead prediction. The accuracy of bunching prediction is heavily dependent on the reliability of headway prediction whose results deteriorate gradually as the prediction horizon extends. Yu et al (2016) used several well-developed algorithms to predict headway first then convert the result to binary bunching occurrence. 2min RMSE is obtained for headway and 99% sensitivity is realized for bunching in 2-stop-ahead prediction, but the performance deteriorates to 6min RMSE and 73% sensitivity for 5-stop-ahead prediction. Moreira-Matias et al (2016) built a regression-based model to predict the headway for a downstream stop and calculate the likelihood of bus bunching to occur for all the further downstream stops. The focus of their study was to propose a proactive control framework in which every suspicious event triggers a bunching alarm. The effect of bunching likelihood thresholds was not investigated. It should be noted that Moreira-Matias et al (2016), Andres and Nair (2017), Berrebi et al (2018) combined prediction and correction, and tested the feasibility and benefit of putting corrective strategies into practice. Instead of bunching prediction, Arriagada et al (2019) used bus GPS data and smartcard data to investigate the causes of bus bunching, with an emphasis on the planning side. Scheduled frequency, stop location and configuration (number of the berths), traffic signal and bus lane design are found influential.

3. METHODOLOGY

(1) The identification of bunching event

As a bunching event involves two PTRs we refer to these as front run and back run respectively. Let a binary variable b_m^n denote whether run *m* is caught in bunching as the back run during its dwelling at stop *n*. a_m^n and d_m^n denote the arrival and departure time of run *m* at stop *n* respectively. At stop *n*, for each run m (m≥2) we can obtain $\Delta_{m-1,m}^n$ which is the time interval between the arrival time of run *m* and the departure time of run *m-1* in (1). Run *m* is considered bunched with run m-1 at the stop when $\Delta_{m-1,m}^n$ is below a threshold Δ_0. The threshold can be determined by the operator. Yu et al (2016) and Moreira-Matias et al (2016) used 1/4 of the scheduled headway. $\Delta_{m-1,m}^n$ is defined as the departure-to-arrival headway in this study. Different from arrival-to-arrival or departure-to-departure headway, $\Delta_{m-1,m}^n$ is negative when two runs overlap at the stop. As overtaking is not allowed, for each stop *n*, run *m-1* always arrives and departs earlier than run *m*, and accordingly time interval $\Delta_{m-1,m}^n$ can always be obtained before the departure of run *m*.

$$\Delta_{m-1,m}^n = a_m^n - d_{m-1}^n \tag{1}$$

For each run *m* (*m*≥2), the binary bunching status b_m^n can be derived by (2)

$$b_m^n = \begin{cases} 1, \Delta_{m-1,m}^n \leq \Delta_0 \\ 0, \Delta_{m-1,m}^n > \Delta_0 \end{cases} \tag{2}$$

(2) Variable selection

Following afore reviewed literature the continuous $\Delta_{m-1,m}^n$ can be used as the dependent variable for headway-prediction approaches. For bunching prediction then an additional step is required judging whether the predicted headway is below a prior defined bunching threshold or not. Instead, in this study, b_m^n is used as dependent variable using logistic regression to directly predict the binary bunching status and bunching probabilities.

Gradually accumulated or suddenly significant inequality in dwell time and travel time might lead two successive runs to be bunched. The back run in a bunching event tends to have a shorter forward-looking headway, negative deviation from timetable (ahead of schedule), less on-board passengers and shorter dwell time than those of front runs in a bunching event or of non-bunched runs (analysis based on tram data by Degeler et al, 2018). Yu et al (2016) used boarding and alighting numbers of two successive buses, link travel time and headway at an upstream stop as the input to their headway-based prediction approach. As only AVL data is used in this study, information regarding boarding, alighting as well as on-board passengers are not available. Instead dwell time is included in the variable set in addition to headway. Deviation from the timetable is excluded here, as the dispatching is not based on the timetable in some cities and the data for this variable might not be available. To conclude, dwell time of two successive buses and their headway at an upstream stop *n-k* are used as the main leading indicators of a coming bunching event in the *k*-step-ahead prediction. The detailed notation is as follows:

t_m^{n-k} dwell time of run *m* at stop *n-k*
t_{m-1}^{n-k} dwell time of run *m-1* at stop *n-k*
$\Delta_{m-1,m}^{n-k}$ time interval between the arrival time of run *m* and the departure time of run *m-1* at stop *n-k*
k prediction horizon in terms of number of stops

(3) Logistic regression

LOGR (Logistic Regression) modeling is widely used in classification problems. In binary classification it not only helps to categorize observations into positive or negative classes, but also interprets the causality by producing the significance of each independent variable. Moreover, it computes the probability of each observation to be in the positive or negative class. The binary bunching status from the perspective of the back brun b_m^n ($m \geq 2$) is taken as the dependent variable. t_m^{n-k}, t_{m-1}^{n-k}, and $\Delta_{m-1,m}^{n-k}$ are the independent variables. Let $\boldsymbol{X}_m^n = [t_m^{n-k},\ t_{m-1}^{n-k},\ \Delta_{m-1,m}^{n-k}]$, then probability of run *m* being bunched at stop *n* as the back vehicle can be derived as

$$Pr(b_m^n = 1|X_m^n) = \frac{1}{1+e^{-\beta X_m^n}} \tag{3}$$

with parameters $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3\ ]$ obtained by fitting the model with real data. Then $Pr(b_m^n = 1|\boldsymbol{X}_m^n)$ for each run *m* at any stop *n* for the same or a different data sample can be computed. b_m^n is predicted to be positive (one-event) if $Pr(b_m^n = 1|\boldsymbol{X}_m^n)$ exceeds a probability threshold Pr_x which is also known as the cut-off point, otherwise, negative (zero-event).

$$b_m^n = \begin{cases} 1, Pr(b_m^n = 1|\boldsymbol{X}_m^n) > Pr_x \\ 0, Pr(b_m^n = 1|\boldsymbol{X}_m^n) \leq Pr_x \end{cases} \tag{4}$$

(4) Linear regression and SVM as benchmark solutions

We now turn to two headway prediction methods that we consider as benchmarks compared to the afore introduced direct bunching prediction method. Firstly, we consider LR (Linear Regression) which is a basic tool in addressing prediction problems. To make LR comparable with LOGR, the same set of independent variables $\boldsymbol{X}_m^n = [t_m^{n-k},\ t_{m-1}^{n-k},\ \Delta_{m-1,m}^{n-k}]$ is applied. With $\boldsymbol{\beta}' = [\beta_0', \beta_1', \beta_2', \beta_3']$ the relationship between the headway at stop *n* and the set of the independent variables containing information *k*-stop-ahead is modeled as

$$\Delta_{m-1,m}^n = \boldsymbol{\beta}' \boldsymbol{X}_m^n \tag{5}$$

Secondly, SVM (Support Vector Machine) can map a non-linear relationship for model input and output, and is tested by a number of studies in predicting bus headway or arrival time (Yu et al (2011), Yu et al (2016, 2017)). The same independent variables and dependent variable are applied to the SVM regression, and a RBF (Radial Basis Function) kernel is selected because it is found both efficient for bus arrival time prediction (Yu et al, 2011) and for bus headway prediction (Yu et al, 2016).

4 DATA DESCRIPTION

A circular bus line, Kyoto City Bus No. 205, which connects the city center, railway station and several famous tourist attractions (Fig. 1(middle)) is selected for the case study. There are 53 stops on this bus line in total. To exclude the effect of dispatching at the terminal and factors for which we do not have data (e.g. driver issues, departure time adjustments), the 2nd stop of the line is taken as the initial stop and the 52nd stop as the last one so that each bus run passes 51 bus stops. Data of five weekdays in April 2016 are used as the training dataset and those of another five weekdays in the same month are used for testing the model.

The scheduled headway varies from hour to hour, and the mean scheduled headway at the initial stop is 6.97min from 6 am to 8 pm. The shortest scheduled headway is 3min at 7 am. Based on this, 1min is used for the bunching threshold as larger threshold can include headway variance that does not lead to bunching.

Adequate time is required to project a successful correction, in particular, if the control strategy is based on manual communication between the dispatcher and the bus drivers. In this study, the proposed approach is tested under a long prediction horizon of 10 stops or more which gives the operator more than 15min to react since the mean stop-to-stop travel time is 1.77min.

The headway fluctuation patterns of several bunched buses are demonstrated in Fig. 2. Because of the bunching effect, the forward-looking headway of back buses fluctuate within a small range and keep below one minute once bunching has been occurring, giving further support to our threshold choice of one minute.

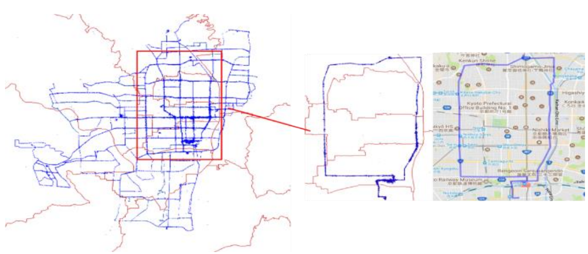


Fig. 1 Data collected (left), data of Kyoto City Bus No. 205 (middle) and its configuration on real map (right).

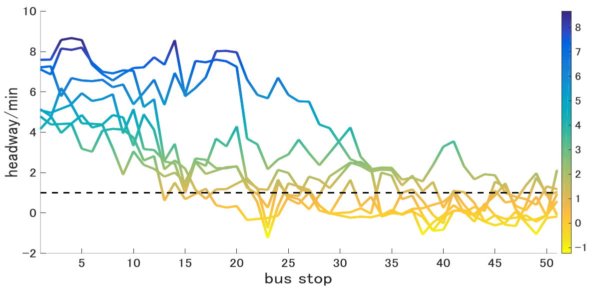


Fig. 2 Headway fluctuation along the line for bunched buses

5 BUNCHING PREDICTION RESULTS

(1) Performance evaluation index

We define an actual bunching as "observed positive" and a predicted bunching as "predicted positive". Similarly, for non-bunching we define "observed negative" and "predicted negative". All the prediction results can be cast into four categories as is shown in Table 3, e.g. it is a true positive if an observed bunching is correctly labeled one in the prediction outcomes. Four indexes can be obtained from (6) to (9). A binary classifier with high true positive rate and high true negative rate is desired. The former is commonly referred to as "sensitivity" and the latter as "specificity". Sensitivity, specificity and accuracy, which is an index computed with (10) to indicate overall prediction performance, are applied to evaluate the binary classification performance of the three algorithms.

Table 1 Four categories for binary classification results

	Observed positive (OP)	Observed negative (ON)
Predicted positive (PP)	True positive (TP)	False positive (FP)
Predicted negative (PN)	False negative (FN)	True negative (TN)

$$\text{True positive rate (TPR, sensitivity, SES)} = \frac{\sum \mathrm{TP}}{\sum \mathrm{OP}} \quad (6)$$

$$\text{True negative rate (TNR, specificity, SPC)} = \frac{\sum \mathrm{TN}}{\sum \mathrm{ON}} \quad (7)$$

$$\text{False positive rate (FPR)} = \frac{\sum \mathrm{FP}}{\sum \mathrm{ON}} \quad (8)$$

$$\text{False negative rate (FNR)} = \frac{\sum \mathrm{FN}}{\sum \mathrm{OP}} \quad (9)$$

$$\text{Accuracy (ACC)} = \frac{\sum \mathrm{TP} + \sum \mathrm{TN}}{\sum \mathrm{OP} + \sum \mathrm{ON}} \quad (10)$$

For headway-based methods, only one combination of sensitivity and specificity is derived, as headway prediction produces an exact value for each headway, resulting in deterministic true positive and negative outcomes. Instead, by using logistic regression different combinations are obtained depending on the cut-off point applied to the predicted probability. The cut-off point is the threshold to determine the predicted positive. The event is judged as positive if its predicted probability exceeds the cut-off point. A high cut-off point tends to only identify events presenting convincingly high probability as positives, and consequently, it thus might misclassify observed positives as negative. Vice versa, a low cut-off point will lead to more false positives. Therefore the cut-off point choice should depend on the operator's attitude towards bunching. Two scenarios are assumed here to represent operators with different weights to false negative errors (missing actual bunching). Moreira-Matias et al (2016) employed a large weight of 10:1 for false negative compared to false positive for aggressive control purposes. We consider more moderate weights of 1:1 and 3:1.

Scenario 1 (LOGR-N): the operator is bunching-neutral, and gives equal weight to false positive and false negative.

Scenario 2 (LOGR-A): the operator is bunching-averse, and gives a 3:1 weight to false negative over false positive predictions.

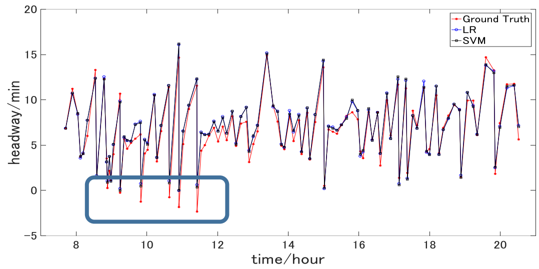
The cost function in (11) computes the total weighted errors given a cut-off point. For LOGR-N, $w_{FP} = w_{FN} = 1$, and for LOGR-A, $w_{FP} = 1$, $w_{FN} = 3$. The cut-off point generating the lowest cost is taken as the optimal one. Based on the scenario-specific predicted positives and negatives, the combination of sensitivity and specificity is determined.

$$c = w_{FP} \sum \mathrm{FP} + w_{FN} \sum \mathrm{FN} \quad (11)$$

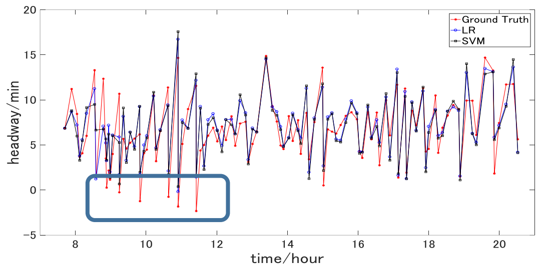
(2) Performance comparison

Headway prediction results at Stop 23 "Kinkaku Temple", one of the most frequented sightseeing spots in Kyoto, is used to illustrate the performance of LR and SVM on the headway. The results of 1-stop-ahead and 10-stop-ahead predictions are illustrated in Fig. 3.

Reliable prediction results (MAPE = 7.42% and RMSE = 0.71min by LR, MAPE = 7.45% and RMSE = 0.71min by SVM) are produced for 1-stop-ahead prediction. For 10-stop-ahead prediction, the results obviously deteriorate (MAPE = 21.64% and RMSE = 1.93min by LR, MAPE = 21.51% and RMSE = 1.92min by SVM). We suggest they can still provide insights into expected fluctuation patterns downstream, but the exact value is not reliable. Furthermore, neither in 1- nor 10-stop-ahead prediction can these two methods perform favorably under the circumstance that the actual headway becomes extremely short and bunching is going to happen, as is highlighted by the blue box in Fig. 4.



(a) 1-stop-ahead headway prediction



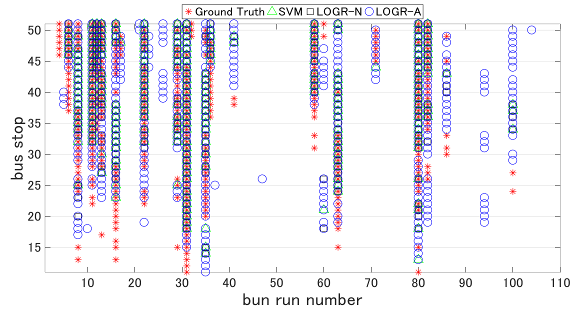
(b) 10-stop-ahead headway prediction

Fig. 3 Performance comparison in terms of exact headway value

Considering that the results derived by LR and SVM are similar, the comparison on bunching prediction is among SVM and two distinguished scenarios based on LOGR. As is presented in Fig 4(a), most bunching events can be detected 1-stop in advance by all three methods, and LOGR-A produces several false positives because it applies a more aggressive strategy to potential bunching events. However, LOGR-A significantly outperforms in 10-stop-ahead prediction, as is illustrated in Fig. 4(b). LOGR-A captures a number of observed positives that are misclassified by SVM and LOGR-N although it generates a few more false positives.

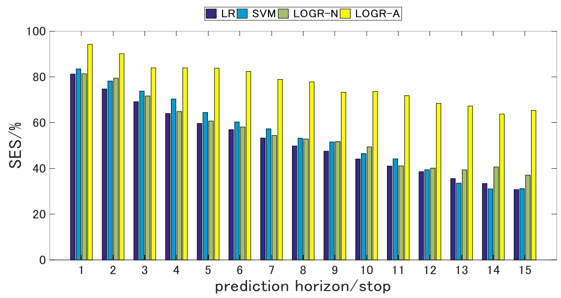
A further comparison among two headway-based approaches and two scenarios of logistic regression is demonstrated in Fig. 5. Sensitivity, specificity and accuracy for the four methods under various prediction horizons are presented. LOGR-A shows remarkable robustness in terms of sensitivity. On the contrary to the obvious deterioration of the other three methods, the sensitivity of LOGR-A keeps above 65% under all the prediction horizons. Besides, it only slightly underperforms the other three methods in terms of specificity, indicating an acceptable trade-off cost.

(a) 1-stop-ahead bunching prediction

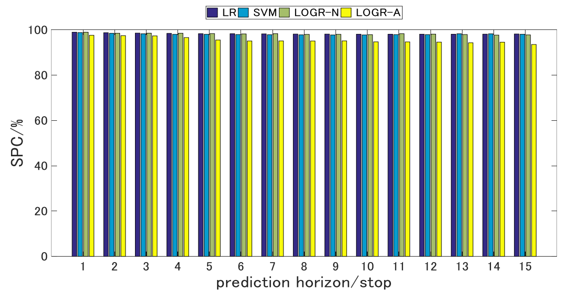


(b) 10-stop-ahead bunching prediction

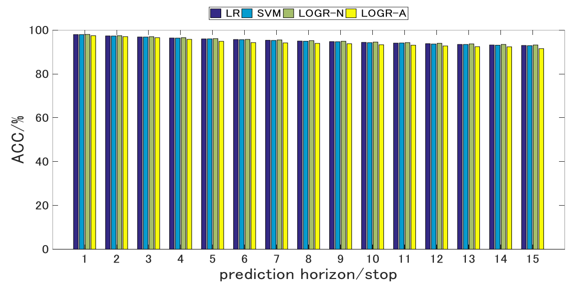
Fig. 4 Performance comparison in terms of binary bunching identification



(a) Deterioration in SES as the prediction horizon extends



(b) Deterioration in SPC as the prediction horizon extends



(c) Deterioration in ACC as the prediction horizon extends

Fig. 5 Performance comparison in terms of SES, SPC and ACC under various prediction horizons

(3) Discussion on the trade-off between sensitivity and specificity

ROC curves created by plotting (1-SPC, SES) for given cut-off points are commonly used to evaluate the classification performance. AUC (Area Under the Curve) being close to one indicates good classification power. ROC curves under various prediction horizons are presented in Fig. 6. Furthermore, the four combinations of sensitivity and specificity derived by the four methods discussed in the previous section are indicated on each curve.

For each horizon, the corresponding curve can be considered the optimal front derived by LOGR. If an algorithm outperforms LOGR, the point it represents should appear above the curve with a higher SES and lower 1-SPC. It can be observed that the two headway-based methods (LR and SVM) generally fall below and sometimes on the LOGR curve, although the downward deviation from the curve is not significant.

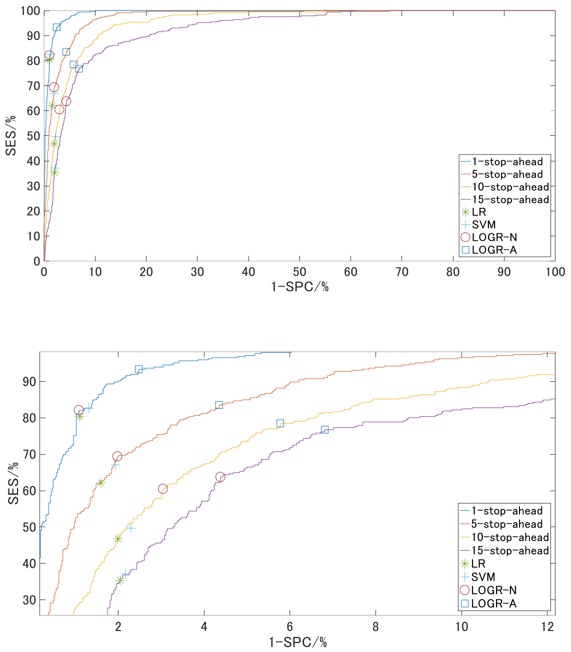


Fig. 6 ROC curves under various prediction horizons (1-stop, 5-stop, 10-stop and 15-stop ahead)

It is easy to conduct the trade-off between sensitivity and specificity on the LOGR curve. The LOGR curve contains all combinations of prediction performance given continuous cut-off points where each cut-off point can be considered as optimal. A bunching-averse operator who is aggressive to eliminate bunching might desire to detect 99% of the positives regardless of the cost to increase false positive rate. This trade-off functionality significantly enhances the flexibility and robustness of existing bunching prediction approaches, especially for putting the predictive methodology into real practice. The curves provide a robust benchmark and insights for future algorithms that address bunching prediction problem. Deterministic methods can only produce one combination of prediction performance which greatly limits its contribution to the real application unless its sensitivity and specificity simultaneously achieve a highly reliable level. Other probabilistic methods generating a curve having higher AUC than LOGR or deterministic methods producing points of substantial upward deviation from the curve under various prediction horizons should be further promising extensions.

6. CONCLUSION AND POTENTIAL APPLICATION ON RAILWAY TRANSIT

In this study, the potential of logistic regression to predict bunching events for public transport is discussed. We compare this method with existing approaches that predict headways and then utilize the headway prediction for bunching prediction. Clearly headway prediction can be used for a larger range of purposes and deeper understanding of the service regularity developments as well as control strategies. However, bunching prediction in itself is important as it can be considered a distinctive state. This paper and other literature illustrate that headways fluctuate, but that, once bunching is reached, this state mostly continues along the line with far less headway fluctuation. We illustrate that when it comes to predicting bunching itself the newly proposed method has the potential to outperform headway-based methods such as LR and SVM in several aspects.

Firstly, LOGR provides superior prediction results under a long prediction horizon. It outperforms LR and SVM by 28% in sensitivity and maintains the same level of specificity in 10-stop-ahead prediction. It also shows improved resistance against deterioration in prediction performance as the prediction horizon extends.

Secondly, robustness and flexibility are significantly enhanced. LOGR provides robust prediction results that contain various sets of bunching outcomes under different cut-off points. This enables the operator to apply weights that are in accordance with their attitude towards bunching and operation budget. Some operators with limited possibility or willingness to apply corrective measures can use SVM or LOGR with neutral cut-off point setting. On the contrary, operators who desire to eliminate any possible bunching might be unwilling to choose headway-based methods which omit a considerable number of bunching in the longterm prediction cases. In this case LOGR-A becomes a much-preferred option. To conclude, LOGR provides operators with a wide range of options that can be tailored by their attitudes towards unexpected system disturbances.

The proposed prediction methodology is validated by bus AVL data in this paper, and the validation by rail transit data is a meaningful extension. This methodology has potential to predict the probability of two trains or metro vehicles being closer than a threshold predetermined by the operator. Whereas for buses we choose a headway that is associated with bunching based on fluctuations (Figure 1), in the rail case a natural selection might be related to the closest distance two trains can have which is largely determined by the signaling system (moving block or fixed block). Headways being equal to this minimum headway will trigger an undesired holding or slowing down. Therefore, the bunching is considered one of the main causes of rail transit delay, and the delay might further affect the following runs. It will benefit the rail transit operator if the causality between dwell time and bunching is investigated, and a tailored tool that can reliably predict rail transit bunching and unexpected longer dwell time is developed.

We suggest that dwell time and headway remain as powerful predictors as tested in the case study of bus transit. Additional variables containing network information should be introduced for rail, however because transfer demand might greatly prolong the dwell times.

Furthermore, the operator can apply different cut-off point to different stations. For the hub station where delay is more undesired, a lower cut-off point can be implemented to achieve high sensitivity and to avoid missing the actual bunching.

ACKNOWLEDGEMENTS

We gratefully acknowledge ASTEM (Advanced Science, Technology and Management Research Institute of Kyoto), notably Hideyuki Yamauchi, for the support in providing us with the bus AVL data.

REFERENCES

1) Andres, M. and Nair, R. (2017) A predictive-control framework to address bus bunching. *Transportation Research Part B*, 104, 123-148.
2) Arriagada J., Gschwender, A., Munizaga, M and Trepanier, M. (2019). Modeling bus bunching using massive location and fare collection data. *Journal of Intelligent Transportation Systems*, 23(4), 332-344.
3) Bartholdi, J. J., and Eisenstein, D.D. (2012). A self-coordinating bus route to resist bus bunching. *Transportation Research, Part B*, 46(4), 481-491.
4) Berrebi, S. J., Hans, E., Chiabaut, N., Laval, J. A., Leclercq, L., and Watkins, K. E. (2018). Comparing bus holding methods with and without real-time predictions. *Transportation Research Part C*, 87, 197-211.
5) Daganzo, C.F. (2009). A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B*, 43(10), 913-921.
6) Daganzo, C.F. and Pilachowski, J. (2011). Reducing bunching with bus-to-bus cooperation. *Transportation Research Part B*, 45(1), 267-277.
7) Degeler, V., Heydenrijk-Ottens, L., Luo, D., Oort, N. and Lint, H. (2018). Unsupervised approach to public transport bunching swings formations phenomenon analysis. *14th International Conference on Advanced Systems in Public Transport (CASPT)*, Brisbane, Australia.
8) Eberlein, X.J., Wilson, M.H.M. and Bernstein, D. (2001). The holding problem with real-time information available. *Transportation Science*, 35(1), 1-18.
9) Hans, E., Chiabaut, N., Leclercq, L., and Bertini, R. L. (2015). Real-time bus route state forecasting using partical filter and mesoscopic modeling. *Transportation Research Part C*, 61, 121-140.
10) Moreira-Matias, L., Cats, O., Gama, J., Mendes-Moreira, J. and De Sousa, J.F. (2016). An online learning approach to eliminate Bus Bunching in real-time. *Applied Soft Computing*, 47, 460-482.
11) Newell, G.F. and Potts, R.B. (1964). Maintaining a bus schedule. *Proceedings of 2nd Australian Road Research Board*, 2, 388-393.
12) Newell, G.F. (1974). Control of pairing of vehicles on a public transportation route, two vehicles, one control point. *Transportation Science*, 8(3), 248-264.
13) Osuna, E.E. and Newell, G.F. (1972). Control strategies for an idealized bus system. *Transportation Science*, 6(1), 52-71.
14) Schmöcker, J. D., Sun, W., Fonzone, A. and Liu, R. (2016). Bus bunching along a corridor served by two lines. *Transportation Research Part B*, 93, 300-317.
15) Sun, A. and Hickman, M. (2005). The real-time stop-skipping problem. *Journal of Intelligent Transportation Systems*, 9(2), 91-109.
16) Sun, W. and Schmöcker, J. D. (2018). Considering passenger choices and overtaking in the bus bunching problem. *Transportmetrica B*, 6(2), 151-168.
17) Wu, W., Liu, R. and Jin, W. (2017). Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour. *Transportation Research Part B*, 104, 175-197.
18) Xuan, Y., Argote, J. and Daganzo, C.F. (2011). Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. *Transportation Research Part B*, 45(10), 1831-1845.
19) Yu, B., Lam, W. H. and Tam, M. L. (2011). Bus arrival time

prediction at bus stop with multiple routes. *Transportation Research Part C*, 19(6), 1157-1170.

20) Yu, H., Chen, D., Wu, Z., Ma, X., and Wang, Y. (2016). Headway-based bus bunching prediction using transit smart card data. *Transportation Research Part C*, 72, 45-59.

21) Yu, H., Wu, Z., Chen, D. and Ma, X. (2017). Probabilistic prediction of bus headway using relevance vector machine regression. *IEEE Transactions on Intelligent Transportation Systems*, 18(7), 1772-1781.

22) Zhang, S. and Lo, H. (2018). Two-way-looking self-equalizing headway control for bus operations. *Transportation Research Part B*, 110, 280-301.