IJITCE

ISSN 2347- 3657

# International Journal of Information Technology & Computer Engineering


Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

# HOURLY BUS PASSENGER DEMAND PREDICTION THROUGH MACHINE LEARNING ALGORITHMS

S.Venkata Ramana[1], K.Mounika2,K.Kavya3,K.Sruthilaya4
[1]Assistant Professor, Department of CSE, Malla Reddy Engineering College for Women, Hyderabad,ramanasoftster@gmail.com
[2,3,4]UG Students, Department of CSE, Malla Reddy Engineering College for Women, Hyderabad.

## ABSTRACT

The tap-on smart-card data provides a valuable source to learn passengers' boarding behaviour and predict future travel demand. However, when examining the smart-card records (or instances) by the time of day and by boarding stops, the positive instances (i.e. boarding at a specific bus stop at a specific time) are rare compared to negative instances (not boarding at that bus stop at that time). Imbalanced data has been demonstrated to significantly reduce the accuracy of machine learning models deployed for predicting hourly boarding numbers from a particular location. This paper addresses this data imbalance issue in the smart-card data before applying it to predict bus boarding demand. We propose the deep generative adversarial nets (Deep-GAN) to generate dummy travelling instances to add to a synthetic training dataset with more balanced travelling and non-travelling instances. The synthetic dataset is then used to train a deep neural network (DNN) for predicting the travelling and non-travelling instances from a particular stop in a given time window. The results show that addressing the data imbalance issue can significantly improve the predictive model's performance and better fit ridership's actual profile. Comparing the performance of the Deep-GAN with other traditional resampling methods shows that the proposed method can produce a synthetic training dataset with a higher similarity and diversity and, thus, a stronger prediction power. The paper highlights the significance and provides practical guidance in improving the data quality and model performance on travel behaviour prediction and individual travel behaviour analysis.

## I. INTRODUCTION

THE rapid progress of urbanization leads to expansion of population in the urban area, increased demand for travel and associated adverse effects in traffic congestion and

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

air pollution [1]–[3]. Public transport has been widely recognized as a green and sustainable mode of transportation to relieve such transport problems. As a conventional public transport mode, buses have always played a dominant role in passenger transportation [4], [5]. However, unreliable travel time, bus-bunching and crowding have led to low level-of services for buses [6]–[8]. This has decreased the bus ridership in many cities, particularly with the advent of ride-hailing services in recent years [9]–[11]. To sustain and increase bus patronage, bus operators must find a way to improve its performance and enhance its image and attraction. Advanced operation and management for bus systems can significantly improve the level-of-service and service reliability, which in turn helps increase the bus ridership [12]–[14]. This requires understanding the spatial and temporal variations in passenger demand and making necessary changes on the supply side [15]–[18]. The smart-card system is initially designed for automatic fare collection. As the system also records the boarding information, for example, who gets on buses, where and when, smart-card data has become a ready-made and valuable data source for spatio-temporal demand analysis [19], public transport planning [20]–[23], and further analysis of emission reduction for the sustainable transport [24], [25]. From the smart-card data, we can easily observe the passenger flow at bus stops and on bus lines, and from which to derive the spatial and temporal characteristics of bus trips [26], [27]. However, extracting useful information from big data automatically still poses a significant challenge. In recent years, machine learning techniques have emerged as an efficient and effective approach to analyzing large smart-card datasets. For instance, Liu et al. [28] captured key features in public transport passenger flow prediction via a decision tree model. Zuo et al. [29] built a three-stage framework with a neural network model to forecast the individual accessibility in bus systems.

In our own recent research [30], we demonstrate that smartcard data combined with machine learning techniques can be a powerful approach for predicting the spatial and temporal patterns of bus boarding. The predictions were found to be highly accurate at an aggregated level, averaged over all travelers. However, our research has also thrown light on the data imbalance issues, when trying to predict travel behavior at the level of individual travelers and fine spatial-temporal details. For instance, the boarding of an individual smart-card holder at a

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

specific stop during a particular time window (e.g. an hour) is a rare event: most of the records would denote negative (non-travelling, or not boarding at this bus stop during this time window) instances, and only a few are positive (travelling, boarding at this stop at this time) instances. Such data imbalance issues can significantly reduce the efficiency and accuracy of machine learning models deployed for predicting travel behavior at the level of individual travelers and fine spatial-temporal details. This motivates this current study where we propose an over-sampling method, deep generative adversarial nets (Deep-GAN) model (initially developed in the context of image generation) to address the data imbalance issue in predicting disaggregate boarding demand (i.e. individual passengers boarding behavior during each hour of the day). We show that, with the synthesized and more balanced database, the prediction accuracy improves significantly. The performance of the proposed approach, based on the Deep- GAN method, is further benchmarked against other resampling methods (including Synthetic Minority Oversampling Technique and Random Under-Sampling) and is shown to have superior performance.

The rest of the paper is organized as follows. Section II reviews the key resembling methods and their applications in transport studies. Section III describes the specific data imbalance issue in predicting the hourly boarding demand. Section IV uses a Deep-GAN to provide a synthesized, more balanced training data sample and a deep neural network (DNN) to predict the individual smart-card holders' boarding actions (boarding or not boarding) in any hour of a day. Section V applies the proposed method to a real-world case study, and the results are discussed in Section VI. Finally, Section VII summarizes the main findings and contributions of this paper and suggests future investigations.

## II.EXISTING SYSTEM

Smart card data has emerged in recent years and provide a comprehensive, and cheap source of information for planning and managing public transport systems. This paper presents a multi-stage machine learning framework to predict passengers' boarding stops using smart card data.

The framework addresses the challenges arising from the imbalanced nature of the data (e.g. many non-travelling data) and the 'many-class' issues (e.g. many possible

boarding stops) by decomposing the prediction of hourly ridership into three stages: whether to travel or not in that one-hour time slot, which bus line to use, and at which stop to board. A simple neural network architecture, fully connected networks (FCN), and two deep learning architectures, recurrent neural networks (RNN) and long short-term memory networks (LSTM) are implemented. The proposed approach is applied to a real-life bus network.

We show that the data imbalance has a profound impact on the accuracy of prediction at individual level. At aggregated level, FCN is able to accurately predict the rideship at individual stops, it is poor at capturing the temporal distribution of ridership. RNN and LSTM are able to measure the temporal distribution but lack the ability to capture the spatial distribution through bus lines.

Disadvantages

• The data generated by SMOTE and ADASYN are susceptible to outliers. They may generate some data in the majority data space due to minority outlier instances (usually noisy data), causing blurred classification borderlines and making the learning difficulties of the classification model.

• The under-sampling methods usually have to pay the price of losing parts of the information of the majority of data because they have to remove a part of the data. Although the Easy Ensemble and Balance Cascade tried to solve the problem of lost information, they increased the number of models tens of times, significantly increasing the computational burden.

• Little study has noticed the loss caused by the data imbalance issue in the public transport system. There is also no research to validate the efficiency of the existing resampling methods on imbalanced data in the boarding prediction task.

## III.PROPOSED SYSTEM

• The data imbalance issue in the public transport system has received little attention, and this study is the first to focus on this issue and propose a deep learning approach, Deep-GAN, to solve it.

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

• This study compared the differences in similarity and diversity between the real and synthetic travelling instanced generated from Deep-GAN and other over-sampling methods. It also compared different resampling methods for the improvement of data quality by evaluating the performance of the next travel behaviour prediction model. This is the first validation and evaluation of the performance of different data resampling methods based on real data in the public transport system.

• This paper innovatively modelled individual boarding behaviour, which is uncommon in other travel demand prediction tasks. Compared to the popular aggregated prediction, this individual-based model is able to provide more details on the passengers' behaviour, and the results will benefit the analysis of the similarities and heterogeneities.

Advantages:

- The system proposes an over-sampling method, deep generative adversarial nets (Deep-GAN) model (initially developed in the context of image generation) to address the data imbalance issue in predicting disaggregate boarding demand (i.e. individual passengers boarding behavior during each hour of the day).
- The system shows that, with the synthesized and more balanced database, the prediction accuracy improves significantly. The performance of the proposed approach, based on the Deep- GAN method, is further benchmarked against other resampling methods (including Synthetic Minority Oversampling Technique and Random Under-Sampling) and is shown to have superior performance.

## IV.MODULES

1.Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Hourly Boarding Demand Type, View Hourly Boarding Demand Type Ratio, Download Trained Data Sets, View Hourly Boarding Demand Type Ratio Results, View All Remote Users,

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

2.View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

3.Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, Predicting Hourly Boarding Demand Type, VIEW YOUR PROFILE. Decision tree classifiers

## V.ALGORITHMS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

### Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

**K-Nearest Neighbors (KNN)**

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

**Logistic regression Classifiers**

*Logistic regression analysis* studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

**Naïve Bayes**

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .
Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

**Random Forest**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

**SVM**

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

**VI.CONCLUSION**

The motivation of this study was because we have faced the challenge of imbalanced data when we used the real world bus smart-card data to prediction the boarding behavior of passengers at a time window. In this research, we proposed a Deep-GAN

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

to over-sample the travelling instances and to re-balance the rate of travelling and non-travelling instances in the smart-card dataset in order to improve a DNN based prediction model of individual boarding behavior. The performance of Deep-GAN was evaluated by applying the models on real-world smart-card data collected from seven bus lines in the city of Changsha, China. Comparing the different imbalance ratios in the training dataset, we found out that in general, the performance of the model improves with more imbalanced data and the most significant improvement comes at a 1:5 ratio between positive and negative instances. From the perspective of prediction accuracy of the hourly distribution of bus ridership, the high rate of imbalance will cause misleading load profiles and the absolutely balanced data may over predict the ridership during peak hours. Comparison of different resembling methods reveals that both over-sampling and under-sampling benefits the performance of the model. Deep- GAN has the best recall score and its precision scores best among the over-sampling methods. Although the performance of the predictive model trained by the Deep-GAN-data is not significantly beyond other resembling methods, the Deep- GAN also presented a powerful ability to improve the quality of training dataset and the performance of predictive models, especially when the under-sampling is not suitable for the data.

The contributions of this study are:

• The data imbalance issue in the public transport system has received little attention, and this study is the first to focus on this issue and propose a deep learning approach, Deep-GAN, to solve it.

• This study compared the differences in similarity and diversity between the real and synthetic travelling instanced generated from Deep-GAN and other over-sampling methods. It also compared different resembling methods for the improvement of data quality by evaluating the performance of the next travel behavior prediction model. This is the first validation and evaluation of the performance of different data resembling methods based on real data in the public transport system.

• This paper innovatively modeled individual boarding behavior, which is uncommon in other travel demand prediction tasks. Compared to the popular aggregated prediction, this individual-based model is able to provide more details on the passengers' behavior, and the results will benefit the analysis of the similarities and heterogeneities.

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

As technology and computing power develop, predicting models will become more and more refined. In the field of demand prediction of the public transport systems, the target will gradually evolve from the bus network and bus lines to individual travel behavior. This advancement can greatly benefit public transport planning and management, such as the digital twin of the public transport system. It is foreseeable that future prediction work in public transport systems will also encounter the challenge of imbalanced data. Our research proposes a Deep-GAN model to address the data imbalance issue in travel behavior prediction. The validation via real world data illustrated that the Deep-GAN showed a better ability to deal with the data imbalance issue and benefits the predictive models compared to other resembling methods. This research provides valuable experience for more researchers and managers in dealing with similar data imbalance issues, especially in public transport.

It may be noted that despite the great performance of Deep- GAN and DNN models, there are still some limitations. First, in this research, Deep-GAN is solely applied for the oversampling. However, there is also a hybrid variant of Deep-GAN where positive instances are over-sampled and negative instances are under-sampled. The promising results of the Deep-GAN oversampling serve as a motivation to test the performance of the hybrid Deep-GAN in future research. Second, this study makes the prediction at the individual level, which creates an explosion of information and makes the computation more difficult. Classifying the passengers (using clustering methods for instance) may be useful in terms of reducing the size of the dataset. Third, the current Deep- GAN does not consider the spatio-temporal characteristics of boarding behavior. Customizing the networks of generator and discriminator in GAN based on the characteristics of the boarding behavior will further improve the quality of generated dummy travelling instances and the performance of the following predictive models. Finally, the proposed Deep- GAN selected the features and variants of the data augmentation independently. So, the improvements are likely to be sub-optimal. Jointly selecting the features and the optimum imbalance ratio is likely to result in further improvements but at the cost of computational complexity. This can be tested in future. Similarly, the optimum rate of imbalance for Deep- GAN has been assumed to be the optimum rate for other

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

resembling methods. This assumption needs to be tested in future research. Even in its current form, this research demonstrates the extent of improvement offered by the Deep-GAN method in addressing the data imbalance issue in modeling boarding behavior. By better predicting the boarding behavior, the findings can help the public transport authorities to improve the level-of-service and efficiency of the public transport system. It can also be extended to other components of the public transport usage behavior – better prediction of the alighting or transfer behavior, for instance.

## VII.REFERENCES

[1] X. Guo, J. Wu, H. Sun, R. Liu, and Z. Gao, "Timetable coordination of first trains in urban railway network: A case study of beijing," Applied Mathematical Modelling, vol. 40, no. 17, pp. 8048–8066, 2016.

[2] W. Wu, P. Li, R. Liu, W. Jin, B. Yao, Y. Xie, and C. Ma, "Predicting peak load of bus routes with supply optimization and scaled shepard interpolation: A newsvendor model," Transportation Research Part E: Logistics and Transportation Review, vol. 142, p. 102041, 2020.

[3] N. Be˘sinovi´c, L. De Donato, F. Flammini, R. M. Goverde, Z. Lin, R. Liu, S. Marrone, R. Nardone, T. Tang, and V. Vittorini, "Artificial intelligence in railway transport: Taxonomy, regulations and applications," IEEE Transactions on Intelligent Transportation Systems, 2021.

[4] S. C. Kwan and J. H. Hashim, "A review on co-benefits of mass public transportation in climate change mitigation," Sustainable Cities and Society, vol. 22, pp. 11–18, 2016.

[5] Y. Wang, W. Zhang, T. Tang, D. Wang, and Z. Liu, "Bus od matrix reconstruction based on clustering wi-fi probe data," Transportmetrica B: Transport Dynamics, pp. 1–16, 2021, doi: 10.1080/21680566.2021.1956388.

[6] S. J. Berrebi, K. E. Watkins, and J. A. Laval, "A real-time bus dispatching policy to minimize passenger wait on a high frequency route," Transportation Research Part B: Methodological, vol. 81, pp. 377–389, 2015.

[7] A. Fonzone, J.-D. Schm¨ocker, and R. Liu, "A model of bus bunching under reliability-based passenger arrival patterns," Transportation Research Part C: Emerging Technologies, vol. 59, pp. 164–182, 2015.

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

[8] J. D. Schm¨ocker, W. Sun, A. Fonzone, and R. Liu, “Bus bunching along a corridor served by two lines,” Transportation Research Part B: Methodological, vol. 93, pp. 300–317, 2016.

[9] D. Chen, Q. Shao, Z. Liu, W. Yu, and C. L. P. Chen, “Ridesourcing behavior analysis and prediction: A network perspective,” IEEE Transactions on Intelligent Transportation Systems, pp. 1–10, 2020.

[10] E. Nelson and N. Sadowsky, “Estimating the impact of ride-hailing app company entry on public transportation use in major us urban areas,” The B.E. Journal of Economic Analysis & Policy, vol. 19, no. 1, p. 20180151, 2019.

[11] Z. Chen, K. Liu, J. Wang, and T. Yamamoto, “H-convlstm-based bagging learning approach for ride-hailing demand prediction considering imbalance problems and sparse uncertainty,” Transportation Research Part C: Emerging Technologies, vol. 140, p. 103709, 2022.

[12] R. Liu and S. Sinha, “Modelling urban bus service and passenger reliability,” 2007.

[13] J. A. Sorratini, R. Liu, and S. Sinha, “Assessing bus transport teliability using micro-simulation,” Transportation Planning and Technology, vol. 31, no. 3, pp. 303–324, 2008.

[14] Y. Wang, W. Zhang, T. Tang, D. Wang, and Z. Liu, “Bus od matrix reconstruction based on clustering wi-fi probe data,” Transportmetrica B: Transport Dynamics, pp. 1–16, 2021.

[15] Y. Hollander and R. Liu, “Estimation of the distribution of travel times by repeated simulation,” Transportation Research Part C: Emerging Technologies, vol. 16, no. 2, pp. 212–231, 2008.

[16] W. Wu, R. Liu, and W. Jin, “Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour,” Transportation Research Part B: Methodological, vol. 104, pp. 175–197, 2017.

[17] W. Wu, R. Liu, W. Jin, and C. Ma, “Stochastic bus schedule coordination considering demand assignment and rerouting of passengers,” Transportation Research Part B: Methodological, vol. 121, pp. 275–303, 2019.

[18] W. Wu, R. Liu, and W. Jin, “Designing robust schedule coordination scheme for transit networks with safety control margins,” Transportation Research Part B: Methodological, vol. 93, pp. 495–519, 2016.

ISSN 2347–3657
Volume 12 , Issue 1, Jan 2024

[19] S. Zhong and D. J. Sun, A Spatio-temporal Distribution Model for Determining Origin–Destination Demand from Multisource Data. Springer, Singapore, 2022, pp. 33–52.

[20] M. Bordagaray, L. dell'Olio, A. Fonzone, and Ibeas, "Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques," Transportation Research Part C: Emerging Technologies, vol. 71, pp. 231–248, 2016.

[21] B. Chidlovskii, "Mining smart card data for travellers' mini activities," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 11, pp. 3676–3685, 2018.

[22] T. Tang, R. Liu, and C. Choudhury, "Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data," Sustainable Cities and Society, vol. 53, p. 101927, 2020.

[23] X. Zhang, Q. Zhang, T. Sun, Y. Zou, and H. Chen, "Evaluation of urban public transport priority performance based on the improved topsis method: A case study of wuhan," Sustainable Cities and Society, vol. 43, pp. 357–365, 2018.

[24] F. Chen, Z. Yin, Y. Ye, and D. Sun, "Taxi hailing choice behavior and economic benefit analysis of emission reduction based on multi-mode travel big data," Transport Policy, vol. 97, pp. 73–84, 2020.

[25] D. J. Sun, Y. Zheng, and R. Duan, "Energy consumption simulation and economic benefit analysis for urban electric commercial-vehicles," Transportation Research Part D: Transport and Environment, vol. 101, p. 103083, 2021.