Article:

Wu, W, Li, P, Liu, R orcid.org/0000-0003-0627-3184 et al. (4 more authors) (2020) Predicting peak load of bus routes with supply optimization and scaled Shepard interpolation: A newsvendor model. Transportation Research Part E: Logistics and Transportation Review, 142. 102041. ISSN 1366-5545

https://doi.org/10.1016/j.tre.2020.102041

Predicting peak load of bus routes with supply optimization and scaled Shepard interpolation: A newsvendor model

Weitiao Wu[a], Peng Li[a], Ronghui Liu[b], Wenzhou Jin[a], Yuanqi Xie[a], Baozhen Yao[c] and Changxi Ma[d]

a. Department of Civil and Transportation Engineering, South China University of Technology, Guangzhou 510641, China

b. Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, United Kingdom

c. School of Automotive Engineering, Dalian University of Technology, Dalian 116024, China

d. School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

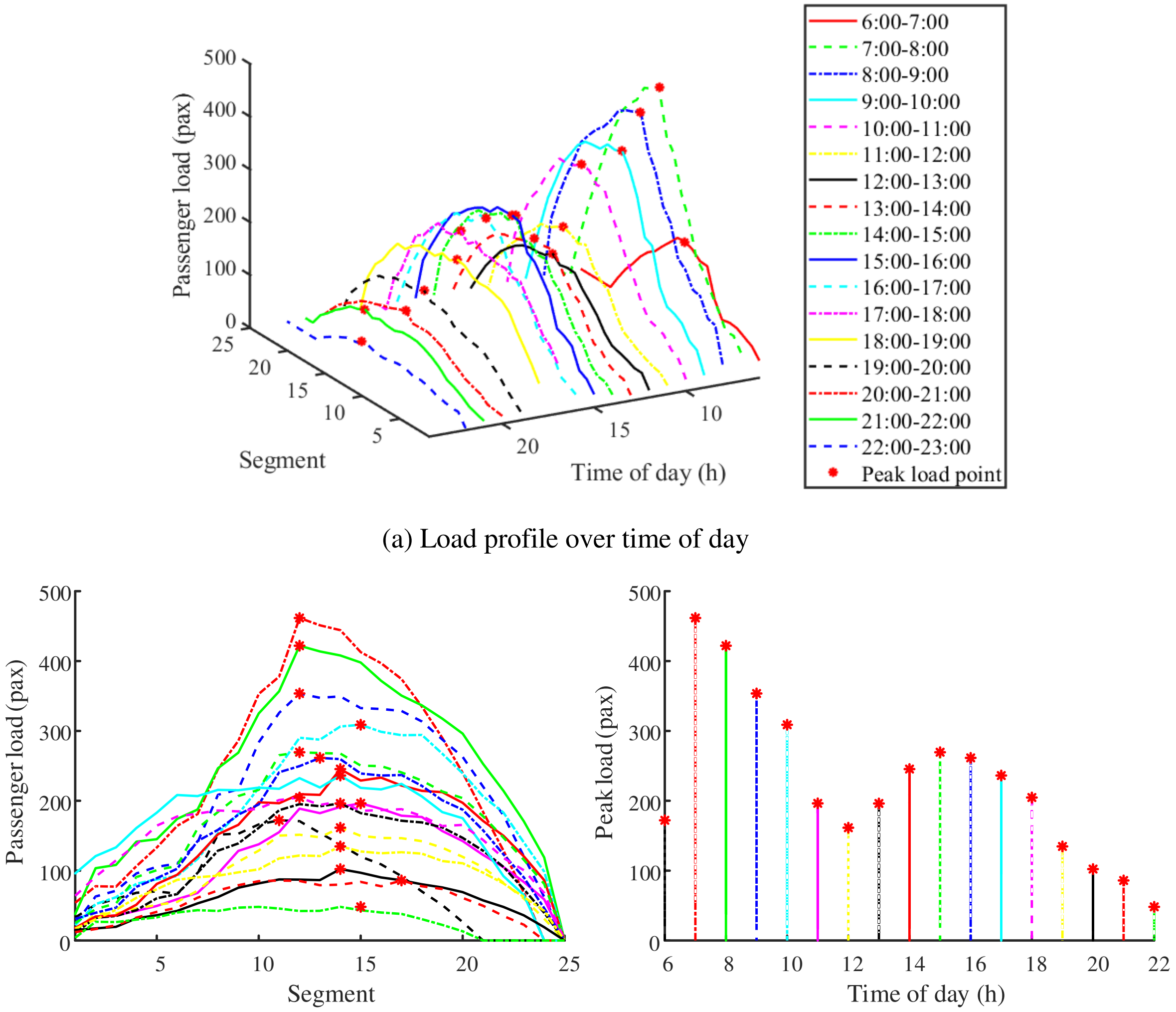
Abstract:

The peak load of a bus route is essential to service frequency determination. From the supply side, there exist ineffective predicted errors of peak load for the optimal number of trips. Whilst many studies were undertaken to model demand prediction and supply optimization separately, little evidence is provided about how the predicted results of peak load affect supply optimization. We propose a prediction model for the peak load of bus routes built upon the idea of newsvendor model, which explicitly combines demand prediction with supply optimization. A new cost-based indicator is devised built upon the practical implication of peak load on bus schedule. We further devise a scaled Shepard interpolation algorithm to resolve discontinuities in the probability distribution of prediction errors arising from the new indicator, while leveraging the potential efficacy of multi-source data by adding a novel quasi-attention mechanism (i.e., scaling feature space and parameter optimization). The real-world application showed that our method can achieve high stability and accuracy, and is more robust to predicted errors with higher capacity. Our method can also produce a larger number of better trip supply plans as compared to traditional methods, while presenting stronger explanatory power in prioritizing the relative contribution of influential factors to peak load prediction.

Keywords: Public transport; Peak load forecast; Supply optimization; Interpolation; Influential factors

1. Introduction

Operation of public transport features spatio-temporally uneven distributed passenger demand and resource shortage. Supply optimization is of great importance to public transport systems as it should realize simultaneously high-quality service to passengers and a cost-effective operation (Liu and Sinha, 2007; Yu et al., 2012; Yao et al., 2014). Accurate forecast of passenger demand is essential therefore to plan a cost-effective public transport operation and improve the level of service via proper allocation of scarce supply. Meanwhile, passenger demand presents both regularity and complexity, while a variety of influential factors can be collected through multi-source information. This brings both opportunities and challenges for leveraging latent knowledge hidden in big data to accurate prediction.



(a) Load profile over time of day

(b) Side view of the load profile

(c) Peak load over time of day

Fig. 1 Illustration of load profile and peak load

The information of passenger loads is essential to crowding management in public transit. From the perspective of users, the provision of in-vehicle crowding information allows passengers to make better informed decisions about whether to board a vehicle or not, or which vehicle to board. From the perspective of operators, the passenger load information can help the design of fleet allocation. The load profile of a period refers to the number of passengers passing through a route in a given direction (during a time period), as shown in Fig. 1(a). The peak load corresponds to the maximum value among all segments. Fig. 1(b) and (c) are two side views of Fig. 1(a) where the horizontal ordinates represent the segment between consecutive stops and the time of day, respectively. As we can see, while the peak load may arise in different positions over different time of day (Fig. 1(b)), the value of peak load is unique for each time of day (Fig. 1(c)). Generally, the peak load of a period determines the in-vehicle crowding level (during a time period), which may exert an influence on the supply optimization. Given its practical significance, this paper attempts to address the prediction of peak load at different time of day.

A common practice in bus supply optimization is to determine the departure frequencies and resultant number of trips during the planning horizon according to a series of factors, such as the desired in-vehicle occupancy, vehicle capacity, and operating costs. The main objective is to achieve the matching between vehicle resources and passenger travel needs. According to the well-known maximum loading point, the frequency should be set such that the bus load at the most heavily-used point along the route does not exceed the desired in-vehicle occupancy (Ceder, 2007). Therefore, the prediction of peak load is essential to the frequency settings, which in turn affects the required number of trips and fleet size. Unlike general demand forecasts in other field, in the context of peak load prediction, the predicted errors would not always change the supply, unless exceeding the vehicle capacity. In other words, when the predicted errors do not lead to a change in the number of trips, the forecast result is still feasible for bus scheduling. However, when the predicted value is smaller than or exceeds the vehicle capacity, insufficient or wasted fleet capacity would be induced, and the positive and negative errors will make a difference. The main challenge is how to incorporate supply optimization into the prediction framework to achieve the cost-effective forecast, taking into consideration the interest of different stakeholders.

The insufficient capacity represents the degradation of service levels, while the surplus capacity denotes a waste of supply and imposed extra cost. The newsvendor model is a promising approach to optimize the supply and handle the trade-off between the service levels and costs under uncertain demand. In the newsvendor problem, a decision maker facing random demand for a perishable product decides how much to stock for a selling period, with the objective to maximize the expected profit given the retail price, purchase price and refund of a newspaper. Due to the uncertain demand, stocking excessive newspapers may result in potential loss due to unsold copies, while inadequate stocking may lead to insufficient sales and reduced revenue. The newsvendor models have been successfully applied in the supply chain inventory management, aviation area and hotel services reservation (Khouja; 1999; Hadas and Herbon, 2015; Bai et al., 2019). Essentially, bus scheduling optimization requires operators to make a trade-off between economic viability of the system and maintaining good service for passengers. Such a cost trade-off between users and operators resembles the newspaper sale. In other words, the vehicle departure or supply determination directly leads to possible loss for either users or operators. Given the similarity between the newspaper sale and bus schedule, we propose a new performance indicator and develop a prediction model for the peak load of bus routes built upon the idea of newsvendor model, which explicitly combines demand prediction with supply optimization. This indicator is able to explicitly capture the effects of vehicle under-utilization or denied boarding due to the surplus or insufficient bus trips, which facilitates the minimization of predicted error costs.

The introduction of the new cost-based indicator, however, has further increased the challenges of developing a stable forecast model. As it turns out (Section 3 and 4), such an indicator creates a discontinuity in the probability distribution of prediction outputs, which in turn increases the instability of prediction outputs. In practice, the fluctuation of predicted error costs could lead to frequent reschedule and operational instability. As machine learning models are generally sensitive to the outliers, such instability may even worsen the prediction performance when using the traditional machine learning models. To address this issue, we propose a scaled Shepard interpolation algorithm to resolve discontinuities in the probability distribution of prediction errors. The mechanism is that the samples with higher similarity in influential factors are closer, and the weighted regression of historical data is used to obtain the predicted value based on the similarity in influential factors. Since the impact of outliers (such as the surge demand) will be weakened in the process of weighted regression of historical data, higher stability of prediction results can be achieved. Commendably, to leverage the potential efficacy of multi-source data, we enhance the standard Shepard interpolation algorithm by scaling the feature space and parameter optimization. The real-world application showed that our method can achieve outstanding prediction performance. Our method can also produce a larger number of better trip supply plans as compared to traditional methods, while presenting stronger explanatory power in prioritizing the relative contribution of influential factors to peak load prediction.

2. Literature review and main contributions

In this section, we review state-of-the-art solution methods, while comparing them to our solution methodology. We begin by reviewing supply optimization in public transport. It proceeds to review prediction models and point out the objectives and contributions.

2.1 Supply optimization in public transport

Public transport planning involves several hierarchically-related procedures including the network design, frequency setting, timetabling, vehicle and crew scheduling. Among them, frequency setting is the prerequisite of subsequent procedures after line alignment. Generally, the methods to determine service frequencies can be classified into three types, namely, maximum loading point methods, load-profile methods (Ceder, 1984; Ceder, 2007) and optimization-based methods (Furth and Wilson, 1982; Hadas and Shnaiderman, 2012; Gkiotsalitis and Cats, 2018).

There also exist other studies on bus operational strategies, such as limited-stop service (Liu et al., 2013; Chen et al., 2015; Wu et al., 2019), interlining service (Gkiotsalitis et al., 2019), real-time bus control (Li et al., 2019) and schedule synchronization (Wu et al., 2016). The main objective of deploying bus operational strategies is to correspond and scale to passenger demand by providing different fleet supplies along the route.

2.2 Transit demand prediction models

According to the prediction time span, passenger flow forecast can be divided into long-term and short-term levels. The former contributes to the design of transit system infrastructure and route alignment, while the latter caters to operational design (Noursalehi et al., 2018). Methodologically, prediction methods can be grouped into two types: parametric approach and non-parametric approach. The parametric models, however, have certain characteristic defects. For example, time series models predict the future from a historic trend, such that they substantially rely on the similarity between the historical data and the predicted data. Given the assumption of linear relationships among time lagged variables, ARIMA cannot well reflect the nonlinear relationships between dependent variables and predictors. In comparison, non-parametric approaches can better tackle the issues of nonlinearity and high dimension. Existing studies mostly used machine learning models to predict the transit demand. To name a few, Liu and Chen (2017) developed a multi-stage deep learning architecture to forecast the passenger flow for Bus Rapid Transit stations. Liu et al. (2019) proposed a multilayer deep learning architecture for short-term passenger flow forecasting in urban rail transit. Tang et al (2019) developed a gradient boosting decision tree algorithm to estimate bus passengers alighting stops. Zhang et al. (2019) presented a deep learning based multitask model for network-wide traffic speed prediction. Wei and Chen (2012) predicted the passenger flow for Taipei rail transit by integrating empirical mode decomposition and back-propagation neural networks. Ma et al. (2014) constructed various demand relevant pattern time series, and developed an interactive multiple model-based hybrid method to forecast passenger demand. Jiang et al. (2014) combined the ensemble empirical mode decomposition and grey support vector machine models in the forecasting of high-speed rail demand. Lin et al. (2018) quantified the uncertainty in traffic volume prediction by combining particle swarm optimization and extreme learning machine.

Although extensive machine learning models have been employed to forecast traffic demand, little evidence is provided on the relative importance of explanatory variables due to their ‘black-box’ procedures and weak interpretation power. In addition, machine learning models substantially rely on the quality and quantity of data and are sensitive to the outliers.

2.3 Evaluation of prediction performance

In the aspect of the evaluation for forecast results, existing performance measures are usually based on the level of accuracy, such as absolute error, absolute percentage error and root square error (Yu et al., 2019). Although these indicators are appropriate for traffic volume prediction, they may be not suitable for peak load prediction due to the peculiarity of practical implication on bus scheduling. One of the basic objectives for public transport service is to adapt the allocated capacities to the maximum number of on-board passengers along the entire route, or peak load, over a given time period. On the supply side, the departure frequency and resulting number of trips in a planning horizon is usually determined by the peak load (Ceder, 2007), while being not directly related to the total passenger demand. Therefore, when the predicted errors do not lead to a change in the number of trips, the forecast result is still feasible for bus scheduling. However, when the predicted error is smaller than or exceeds the vehicle capacity, insufficient or wasted fleet capacity would be induced. In this study, we term such resultant additional costs as "predicted error cost".

2.4 Objectives and contributions

Previous transportation research is rich with approaches for demand prediction and supply optimization, while being short of fully treating both these aspects. More specifically, existing studies usually model demand prediction and supply optimization separately, and optimize the supply given the output of an independent demand model. However, this may result in suboptimal solution on the supply side in that the predicted errors will not always affect the optimal supply. To address this issue, in this paper we devise a framework which explicitly incorporates the supply optimization into the peak load prediction. Our objective is to predictively optimize the service frequencies and the number of trips during the planning horizon. This is achieved by devising a new cost-based loss function, drawing an analogy to the newsvendor problem.

The main contributions of this work can be summarized as follows: (1) we develop a model to achieve a cost-effective prediction of peak load of bus routes, which explicitly combines demand prediction with supply optimization via the philosophy of the newsvendor model. Our model is scalable and can be extended to other frequency setting method (load-profile method) that takes into account load variations; (2) we propose a new indicator based on predicted error cost, which can comprehensively reflect the practical implication of predicted peak load on bus schedule, and the cost associated with vehicle under-utilization or denied boarding as a result of surplus or insufficient trips; (3) We devise a novel interpolation method called scaled Shepard interpolation algorithm to resolve discontinuities in the probability distribution of prediction errors arising from the new indicator, while leveraging the potential efficacy of multi-source data by scaling feature space and parameter optimization. The scaled Shepard interpolation algorithm processes stronger explanatory power in prioritizing the relative contribution of influential factors to peak load prediction compared to the standard version.

The exposition of this paper is organized as follows. In Section 3, modeling frameworks are provided. In Section 4, prediction approach is devised. In Section 5, model extensions are presented. In Section 6, numerical examples are conducted to validate the models. Finally, the conclusions and future research are given in Section 7.

3. Model development

3.1 Problem description

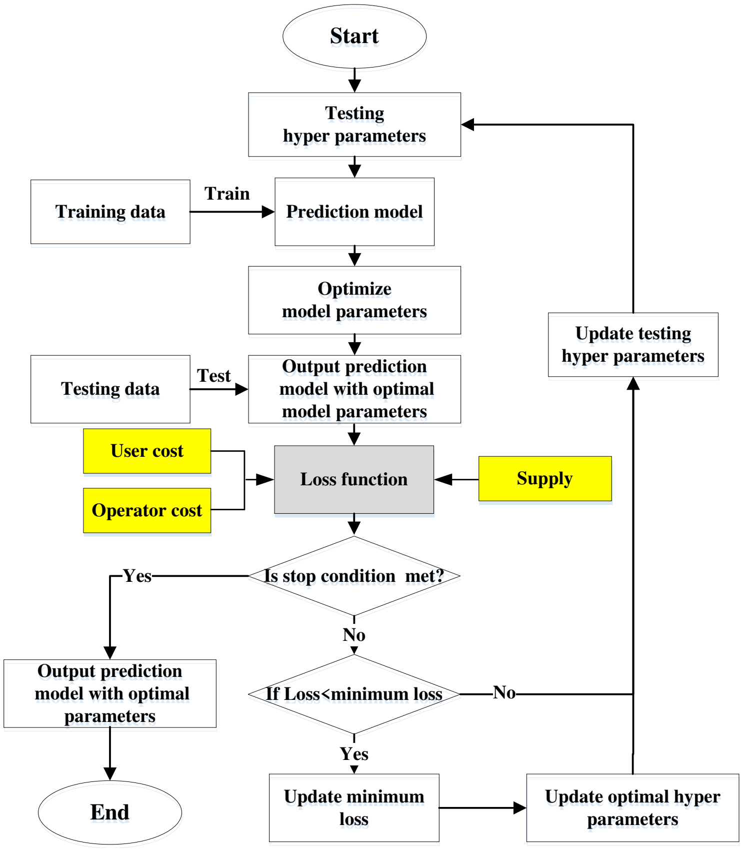


Fig. 2 A modified procedure of prediction models

Generally, the objective of traditional machine learning models is to pursue 'loss' as less as possible by optimizing the hyper parameters and model parameters. The hyper parameters are those values set prior to the commencement of the learning process, while the model parameters refer to the internal model variables that could be estimated by dataset. The loss function measures the quality of the predicted output relative to the actual output, which is generally represented by a specific error function, such as mean square error, the mean absolute percent error, and the mean relative percent error. However, when it comes to the peak load prediction problem, the predicted value has its special implications since it directly determines the bus dispatching plan (the number of trips in a period) and the level of service.

A common practice in public transport planning is to determine the service frequency and resultant number of trips in a planning horizon according to the predicted value of peak load (Ceder, 1984). Therefore, the deviation between the actual value and the predicted value would result in surplus or insufficient number of dispatching trips. To illustrate, we consider a scenario whereby the actual value of peak load is 550 pax and the in-vehicle occupancy is 50 pax. If the predicted peak load is 500 pax, then the planned number of trips is 10. In this case, 50 passengers will be unable to get aboard and have to wait for the subsequent bus, which results in additional waiting time cost. Meanwhile, part of these passengers might even leave the stop and switch to other travel mode or bus routes, which results in lost revenue.

On the other hand, if the predicted value is 600 pax, then the planned number of trips becomes 12. In this case, one redundant trip with unoccupied seats will be induced. This indicates, from the operator's perspective, the non-utilization of resources that imposes additional operating cost. Although the absolute error of the passenger flow forecast is identical (50 pax) for both cases, the resulting costs may be quite distinct, which is closely related to operational parameters such as vehicle capacity, value of waiting time, departure headway, and unit operating cost. Therefore, from the standpoint of bus scheduling, a critical issue of peak load forecast is how to achieve a cost-effective bus scheduling scheme instead of purely pursuing less absolute errors. In this sense, the loss function of the prediction model should be modified in the context of peak load prediction. To this end, we propose a modified procedure of prediction models, where the evaluation of loss function is associated with the transit supply and the interests of users and operators (Fig. 2). The stop condition can be associated with the maximum number of iterations, or the gap of predicted error cost between successive iterations. In this way, the peak load prediction with supply optimization can be achieved.

The main assumptions of this work are:

(1) The departure frequencies are determined such that bus load at the most heavily-used point along the route does not exceed the desired in-vehicle occupancy. This is a reasonable assumption since the overarching objective of bus transit service is to ensure adequate space to accommodate the maximum number of on-board passengers along the entire route over a given time period (Ceder, 1984), particularly for the urban areas with high level of service requirement. Nevertheless, our model can be extended to load-profile method that takes into account load variations with the additional information provided by the load profile.

(2) In case of boarding failure, a proportion of denied boarding passengers may leave the stop for other traffic modes or common routes. The number of swapping passengers depends on the departure frequency. This assumption is to make sure that the networking effect of overlapping and crossing bus lines can be captured.

(3) The model aims to predictively optimize the transit supply (the number of trips) during the planning horizon. In other words, the peak load of a specific period can be predicted at least one day ahead of time (rather than one hour ahead of time within day). This assumption is reasonable since frequency settings and fleet sizing are usually undertaken in the planning stage.

3.2 A new cost-based loss function

In this section, we introduce a new indicator for evaluating the peak load prediction from the standpoint of bus scheduling. By assumption (A1), the departure frequency in period t is the ratio of the peak load demand to the desired in-vehicle occupancy, while not exceeding the minimum departure frequency, that is,

$$F_t = max\left(\frac{y(t)}{T_t \cdot d_{0t}}, F_{mt}\right) \tag{1}$$

where $y(t)$ is the peak load in period t. d_{0t} is the desired in-vehicle occupancy, which equals to the vehicle capacity c discounted by a load factor γ_t, i.e., $d_{0t} = \gamma_t \cdot c$, $0 < \gamma_t \leq 1$. This load factor in bus capacity is used to accommodate the randomness in passenger demand within a period. F_{mt} is the required minimum departure frequency in period t. T_t is the length of period t, which is usually measured by 60 min in practice. However, the time interval can vary to deal with different granularity and demand variation, which would not affect the generalization of the framework. Generally, a shorter interval reduces the corresponding demand variability; however, it will complicate the subsequent timetabling task that involves setting departure times in the transition segments between adjacent intervals (Ceder, 2007).

Rounding up the product of the departure frequency and the period duration yields the number of trips required in the given planning period. The formulation takes the following form:

$$N(t) = \lceil F_t \cdot T_t \rceil \tag{2}$$

where $N(t)$ is the number of trips required in period t.

When the optimal number of trips resulting from the "actual" peak load is identical to the planned number of trips resulting from the "predicted" peak load, the predicted error of peak load is "ineffective" from the bus scheduling standpoint in that there will be no change of the number of trips. However, if the predicted error of peak load for a planning period leads to the change in the optimal number of trips, surplus or insufficient capacity will arise. There are two options in this case: When the fleet capacity is inadequate, a number of passengers will be left behind, which affects the level of transit service. By contrast, when the fleet capacity is excessive, it will lead to empty seats on buses and a waste of supply. Taken together, there exist both effective and ineffective predicted errors in the prediction of peak load. However, the traditional indicators such as absolute error or relative error cannot capture such effect.

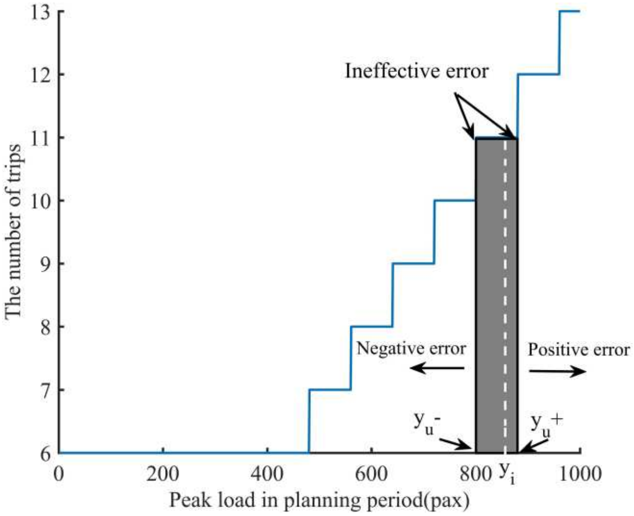


Fig. 3 Diagram of the concept of predicted error costs

With this in mind, we propose a new cost-based indicator, which is defined as the predicted error cost as a result of excessive or inadequate trips. Fig. 3 illustrates the concept of predicted error cost. The horizontal axis represents the volume of hourly peak load, and the vertical axis represents the corresponding number of trips. The blue curve shows the number of trips as a function of the peak load given $d_{0t} = 80$ and $F_m = 6$. As we can see, the number of trips increases in a stepwise manner with the increase of peak load. When the predicted value lies within the shaded area $[y_{u-}, y_{u+}]$, the number of trips required remains unchanged. In other words, the predicted error within this range will not change the number of trips, and thus we term it as "ineffective error". y_{u-} and y_{u+} denote the minimum and maximum peak load between which the number of trips maintains fixed, respectively. The formulations are given by Eqs. (3) and (4). Evidently, the desired in-vehicle occupancy is related to the range of $[y_{u-}, y_{u+}]$ (Eq. (5)).

$$y_{u-} = \left\lfloor \frac{y(t)}{d_{0t}} \right\rfloor \cdot d_{0t} + 1$$

(3)

$$y_{u+} = \left\lceil \frac{y(t)}{d_{0t}} \right\rceil \cdot d_{0t}$$

(4)

$$y_{u+} - y_{u-} = d_{0t} - 1$$

(5)

On the other hand, when the predicted value lies outside the shaded area $[y_{u-}, y_{u+}]$, predicted error cost will arise. The total cost is made up of a shortage component (when demand is higher than capacity) and a surplus component (when demand is lower than capacity). We now model each component as follows.

3.2.1 Shortage component

Let y_i be the actual value of peak load (for the i-th experiment point), which can be obtained through field observations or certain approaches (see detailed discussion in Section 4.2). When the actual value y_i falls beyond the right boundary of the shaded area, i.e., $y_i > y_{u+}$, the resultant error is positive. Under such circumstance, the passenger demand is higher than the fleet capacity, which results in boarding failure and additional waiting time cost. Then, the total error cost is dependent on the number of denied boarding passengers, the departure headway and value of waiting time. By definition, the predicted error $y_i - y_{u+}$ is equal to the number of denied boarding passengers. As such, the error cost increases with the predicted error proportionally, while presenting differential growth rates with different values of passenger waiting time.

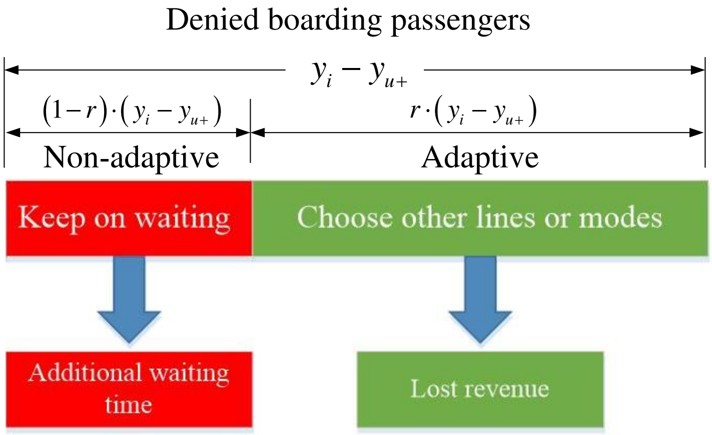


Fig. 4 Illustration of passenger choice behaviour

To better reflect the reality, we assume that the passengers react to boarding failure via two travel strategies: non-adaptive and adaptive trips. In the first strategy, the denied boarding passengers will stick to the original bus route and wait for the subsequent bus; while in the second strategy, a proportion of denied boarding passengers would choose other traffic modes or common routes. The passenger choice behaviour is illustrated in Fig. 4. Similar to Hadas and Herbon (2015), the swapping proportion takes the following form:

$$r = r_0 + r_1 \left(\frac{T_t}{N(t)}\right)^{\beta}$$

(6)

where r_0 stands for the minimal proportion of passengers leaving the stop among the denied boarding passengers. The second component represents the additional proportion of passengers leaving the stop due to the waiting time (or headway), $\frac{T_t}{N(t)}$. The selection of coefficient $r_1 (r_1 > 0)$ should not cause the swapping proportion r to exceed 1. β is a parameter representing the power. As we can see, the swapping proportion becomes lower in a high-frequency route where $N(t)$ is relatively large.

Correspondingly, the error cost consists of two components, namely the additional waiting time for non-adaptive denied boarding passengers C_m, and the lost revenue by adaptive denied boarding passengers C_t.

$$CE_i = C_m + C_t \tag{7}$$

The additional waiting time for non-adaptive denied boarding passengers is related to the product of expected headway $\frac{T_t}{N(t)}$ and the value of waiting time C_p.

$$C_m = (1 - r)\left(y_i - \left\lceil \frac{\hat{y}_i}{d_{0t}} \right\rceil \cdot d_{0t}\right)\frac{T_t}{N(t)} \cdot C_p \tag{8}$$

where $\hat{y}_i$ is the predicted value of peak load. Note that the predicted value of peak load is rounded up since it is associated with the number of passengers.

The lost revenue can be simply calculated by the number of adaptive denied boarding passengers multiplied by the ticket price P.

$$C_t = r\left(y_i - \left\lceil \frac{\hat{y}_i}{d_{0t}} \right\rceil \cdot d_{0t}\right) \cdot P \tag{9}$$

3.2.2 Surplus component

When the real value y_i falls beyond the left boundary of the shaded area, i.e., $y_i < y_{u-}$, the resultant error is negative. In this case, the passenger demand is lower than the fleet capacity. This results in the costs of empty seats and redundant trips. From the viewpoint of operator, the non-utilization of resources imposes additional monetary costs. Unlike the shortage cost, since the surplus cost is related to the cost of operating an extra bus, the error cost presents stepwise increase with the increment of the predicted error $y_{u-} - y_i$, that is,

$$CE_i = \left\lceil \frac{y_{u-} - y_i}{d_{0t}} \right\rceil \cdot C_b \tag{10}$$

where C_b represents the direct cost of operating a standard vehicle, which can be estimated as the product of the unit operating cost of a bus per kilometer and the route length. Note that y_{u+} and y_{u-} can be calculated by Eq. (3) and Eq. (4) given the predicted value of peak load $\hat{y}_i$, respectively.

3.2.3 Consolidated formulation

By combing Eqs. (7) and (10) and the "ineffective error" region as shown in Fig. 2, the predicted error cost can be formulated as the following piecewise function.

$$CE_i = \begin{cases} \left\lceil \frac{y_{u-} - y_i}{d_{0t}} \right\rceil \cdot C_b & y_i < y_{u-} \\ (1-r)\left(y_i - \left\lceil \frac{\hat{y}_i}{d_{0t}} \right\rceil \cdot d_{0t}\right) \frac{T_t}{N(t)} \cdot C_p & \\ +r\left(y_i - \left\lceil \frac{\hat{y}_i}{d_{0t}} \right\rceil \cdot d_{0t}\right) \cdot P & y_i > y_{u+} \\ 0 & y_{u-} \le y_i \le y_{u+} \end{cases} \tag{11}$$

where CE_i represents the predicted error cost for the i-th experiment point, which includes three components. The first component refers to the total surplus costs. The second component refers to the total shortage costs. The third component is related to the “ineffective error” as discussed in Fig. 3.

In the following sections, we develop a newsvendor model and formalize the prediction problem where the transit operator wishes to estimate the peak load of a bus route, while minimizing the total predicted error costs composed of the shortage and surplus components. In Eq. (11), C_b and C_p can be understood as the penalties for lack of capacity and excess capacity, respectively. These two parameters are equivalent to the loss of potential revenue caused by insufficient and excess stock of newspapers in the newsvendor model. In practice, the parameters can be adjusted to trade off the interests between passengers and operators.

3.3 Validation of the cost-based loss function

As mentioned above, we propose a new cost-based loss function to judge on the performance in the context of peak load prediction. In order to demonstrate the necessity of using such an indicator, in this section we make a comparative analysis of traditional error-based loss function and the new cost-based loss function using the experimental data (Section 4).

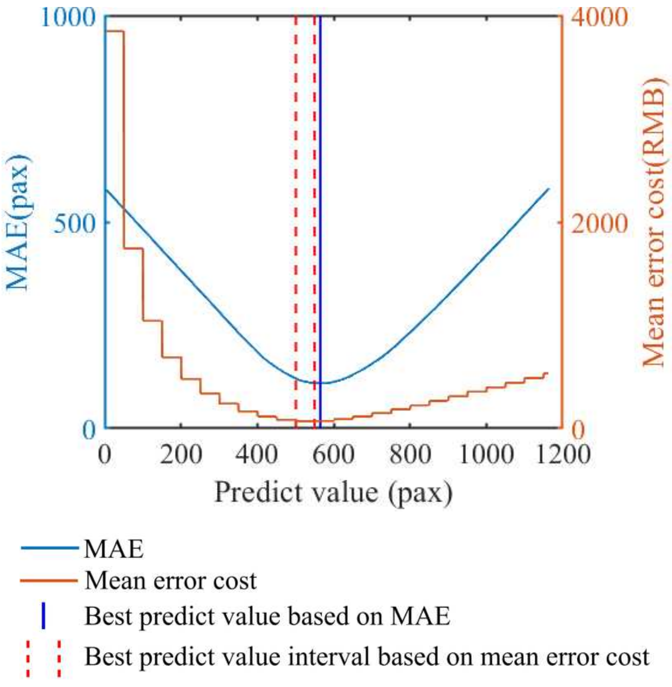


Fig. 5 Comparison of the error-based and cost-based loss of assumed predict value

The actual passenger flow of working days in a typical period (7:00 am-8:00 am) was extracted from the dataset. Fig. 5 shows the mean absolute error (MAE) and mean error cost corresponding to different predicted values. The blue curve represents the result of MAE, and the red curve represents the result of average error cost. The interval between the red vertical dotted lines stands for the interval with minimum error cost, while the blue vertical line corresponds to the minimum MAE. As we can see, there exist a sequence of scales and "step phenomenon" for average error cost, which is due to the presence of ineffective errors within a certain range of peak load. Distinctly from the symmetrical distribution of MAE, the error cost presents an asymmetrical distribution in that it increases at a decreasing rate when the predicted value shifts right. One can also see that the minimal value of MAE does not necessarily fall into the interval with minimum error cost. In this example, the 'optimal' predicted value based on MAE is larger than that based on error cost by approximately 40 pax. When the desired crowdedness level d_{0t} does not exceed 40 pax, such a difference will result in extra trip cost of a standard vehicle. The accumulated error costs will be even more prominent when the operation period is longer (e.g., several months), which cannot be neglected in the bus schedule planning. This suggests that the bus schedule developed according to the minimal value of MAE does not lead to the minimum costs. Therefore, there is an imminent need to adopt the cost-based loss function to predict peak load to capture the real implication and reduce the system costs.

4. Prediction approach

The purpose of this study is to predict the hourly peak load of a bus route, whereas the introduction of the new cost-based indicator has further increased the challenges of developing a stable forecast model. As we can see from Fig. 5, the curve of MAE is quite smooth, whereas the introduction of cost-based indicator creates a discontinuity in the probability distribution. Such a discontinuity indicates the instability of prediction outputs. In practice, the fluctuation of predicted error costs could lead to frequent reschedule and operational instability. Such instability might even worsen the prediction performance of the traditional machine learning models. To address this issue, we propose a new interpolation forecast method to predict the peak load, combining the merits of multi-source data.

4.1 Interpolation method

The principle of interpolation forecast algorithm is that the target values with similar influential factors are closer to each other. Based on the similarity of influential factors, the predicted value is obtained by the historical data with weighted regression. Through simple approximated functions, the correlation between the passenger flow and its influential factors can be predicted by interpolation with known data points. The effect of the outliers (such as surge demand) will be weakened in the process of weighted regression of historical data. Due to its inherent mechanism, the interpolation method could achieve high stability in the prediction. Since the relative importance of each influential factor may vary over time, the passenger flow can be effectively predicted through effective matching between historical data and the influential factors. In this algorithm, the passenger flow in each period of the target day is predicted independently by using the influential factors. In this sense, the passenger flow can be predicted at least one day ahead of time (rather than one hour ahead of time within day). By assumption (A3), this algorithm is particularly applicable to the transit supply optimization at the planning level, which is the center of our interest.

The Shepard algorithm is a popular interpolation method based on similarity, which is usually used to interpolate the scattered experimental data and produce a continuous surface. One advantage of scattered data interpolation lies in the fact that the points are not required to be structured with regard to their relative locations, which contributes to the prediction stability. This method is based on a distance-weighted, least-square approximation technique, with the weights varying by the distance of the data points. The closer to the forecasting point, the greater weight is assigned. Shepard method has been successfully applied in the design of composite structures (Shi and Xia, 2018), but has not been ever applied in the field of transportation.

4.1.1 Scaled Shepard interpolation algorithm

In what follows, we propose a novel interpolation method, called scaled Shepard interpolation algorithm, to predict the peak load of bus routes. This method is a modification of the standard Shepard interpolation algorithm combining the merits of multi-source data. It contains three stages. The first stage extracts the influential factors and constructs the feature space. The second stage scales the feature space, and the third stage optimizes the weights and power parameter. According to the influential factor vectors of the predicted values, the predicted value is interpolated using the inverse distance weight of historical global sample points. The main computational steps are described as follows:

1) Compute the weighted Euclidean distances between the influential factors of the target value and the influential factors of each historical value.

2) The inverse distance weights between target values and historical values are set as the b-th power of reciprocals of the weighted Euclidean distance.

3) All historical values are inversely weighted and accumulated to compute the predicted value of the target value.

The framework of the scaled Shepard interpolation algorithm is shown in Fig. 6.

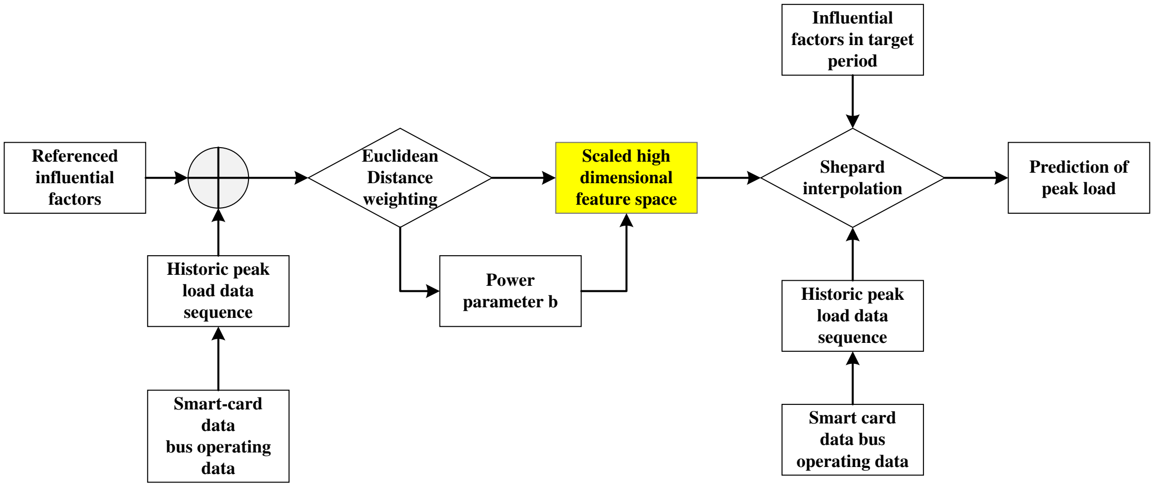

Influential factors in target period
Referenced influential factors
Euclidean Distance weighting
Scaled high dimensional feature space
Shepard interpolation
Prediction of peak load
Historic peak load data sequence
Power parameter b
Historic peak load data sequence
Smart-card data bus operating data
Smart card data bus operating data

Fig. 6 The framework of the scaled Shepard interpolation algorithm

4.1.2 Innovation in the standard Shepard interpolation algorithm

In a standard Shepard algorithm, the weights of different influential factors are identical, underlying that the difference between the target values is proportional to the Euclidean distance between these two vectors. This process can cause some issues to the prediction performance:

- This will undermine the prediction accuracy of the interpolation algorithm because the contribution of each dimension to the distance may be distinct. In other words, the influential factors at different times are likely to have a different influence on the passenger flow. To resolve this issue, we design a ‘quasi-attention’ mechanism that scales the feature space for multi-source data to automatically exploit different levels of importance of an influential factor sequence at different times. The scaling is then optimized to minimize the total predicted error cost.
- Since the weights of different influential factors in the standard Shepard algorithm are identical, the relative influences of predictor variables on passenger flow could not be evaluated. The modifications applied to standard Shepard algorithm improve the interpretability of the prediction results: it can not only identify, but also rank, the relative importance of influential factors on peak load prediction.

4.2 Detail of each stage in the scaled Shepard interpolation algorithm

In the previous sections, the tailored Shepard interpolation algorithm was described succinctly. Some stages need more elaboration, which are explained in the following sections.

4.2.1 Construction of feature space for influential factors

Bus passenger flow at a given time period is affected by a series of factors, such as date, working days, weather, and temperature. The impact of each factor may be quite distinct. In the era of internet-of-thing, it is possible to identify the passenger characteristics under a specific environment using multi-source big data. In addition to the smart-card data, other attributes such as working days/holidays and school days/holidays can be readily collected from the announcements of relevant departments. For example, the historical weather information can be obtained through local historical weather records, and the future weather information can be obtained via weather forecasts. This opens up more opportunities to build the linkage between the passenger flow and influential factors.

In order to utilize the interpolation algorithm, each influential factor should be quantified as an effective parameter. To this end, the influential factors are extracted and abstracted into multi-dimensional vectors, and the dimensional effect is eliminated by normalization. The sequence of sample influential factors is defined as $\{x(i,j,t)|i = 1, \ldots, n; j = 1, \ldots, m; t = 1, \ldots, T\}$. The sequence of historical peak load is defined as $\{y(i,t)|i = 1, \ldots, n; t = 1, \ldots, T\}$, where n is the number of samples, m is the number of influential factors, T is the number of periods, $x(i,j,t)$ is the quantized value of j-th influential factor of i-th samples in period t, and $y(i,t)$ is the historical passenger flow data of i-th sample point in period t.

Generally, the passenger demand presents seasonal variation pattern. For instance, during the summer, educational trips decrease while recreational trips increase (Amiripour et al., 2014). For this reason, we select the day of year (X_1) as the first influential factor. The working day attribute is a major influential factor on demand generation and distribution. The workday attribute, denoted by X_2, is set as 0 for a weekday and 1 otherwise. The temperature may have an impact on the temporal distribution of passenger flows and the elastic demand. According to the period and actual conditions in the survey area (Guangzhou, China), the temperature range, which is represented by X_3, is set in the range between 0 and 40℃. The weather may also exert an influence on the distribution of passenger flows and elastic demand. In this paper, the rainy day, denoted by X_4, is classified into three categories according to the precipitation. Value 1 indicates the weather with no rain and little rain, including sunny, cloudy and clear to light rain. Value 2 indicates the weather with moderate rainfall, including overcast to light rain, thunder shower and light to moderate rain. Value 3 indicates the weather with high rainfall, including heavy rain and extreme weather. In practice, passenger flow usually exhibits a recurrent fluctuation within one week. Meanwhile, the passenger flows from Monday to Friday in the working days may be quite distinct, particularly for large cities. For example, in Guangzhou and Beijing, the ridership on Monday and Friday is the highest among the working day due to the surging demand, such as students and commuters. For this reason, the day of week, denoted by X_5, is defined as a value ranging between 1 and 7. In addition, the air quality may have a certain impact on the ridership. The air quality index of the day is represented as X_6, which indicates to what extent the air is polluted currently is or is anticipated to become. As the air quality index increases, a larger percentage of the population is likely to experience increasingly severe adverse health effects. The selection of various influential factors and their corresponding values are summarized in Table 1.

Table 1 Selection and quantification of influential factors

Influential factors	Time of day t	Day of year X_1	Working day X_2	Temperature X_3	Rainy day X_4	Day of week X_5	Air quality index X_6
Ranges	0~23	1~365	1/0	0~40	1~3	1~7	0~500

To eliminate the dimensional effect of the influential factors, the influential factors are first normalized by the following formula:

$$x'(i,j,t)=\frac{x(i,j,t)-E(j,t)}{S(j,t)} \tag{12}$$

where $E(j,t)$ and $S(j,t)$ represent the mean and standard deviation of the sample series for the j-th influential factor in period t, respectively.

4.2.2 Applying standard Shepard interpolation algorithm and its shortcomings

As a prerequisite, the standard Shepard interpolation algorithm is first introduced in this section. In the present study, let $\hat{y}(n+1,t)$ denote the predicted value in period t under the impact of the $(n+1)$-th influential factor vector. According to the interpolation method proposed by Shepard (1968), the regression forecast value of the target value $\hat{y}(n+1,t)$ is the weighted sum of the inverse distances of historical values.

$$\hat{y}(n+1,t)=\sum_{i=1}^{n}\frac{h_i y(i,t)}{\sum_{j=1}^{n}h_j}$$

(13)

where $\frac{h_i}{\sum_{j=1}^{n}h_j}$ is the normalized interpolation function with h_i being the weight of the i-th (historical) experiment point. h_i is calculated as the inverse power of distance between an influential factor and the forecast target, that is,

$$h_i=d_i^{-b} \tag{14}$$

where b denotes the power parameter greater than 1, which may be interpreted as the contribution of the dissimilarity in influential factors to the target value. d_i is the Euclidean distance between the influential factor $x'(i,j,t)$ and the forecast target $x'(n+1,j,t)$, which is expressed as:

$$d_i=\sqrt{\sum_{j=1}^{m}(x'(i,j,t)-x'(n+1,j,t))^2} \tag{15}$$

Nevertheless, the standard Shepard interpolation processes certain shortcomings in predicting passenger flow with multi-source data. In Eq. (15), the weights of different influential factors are identical, underlying that the difference between the target values is proportional to the Euclidean distance between these two vectors. This will undermine the prediction accuracy of the interpolation algorithm since contribution of each dimension to the distance may be distinct. To address this issue, in the following we propose a modified Shepard interpolation algorithm to scale the feature space for multi-source data, followed by parameter optimization to prioritize the influential factors.

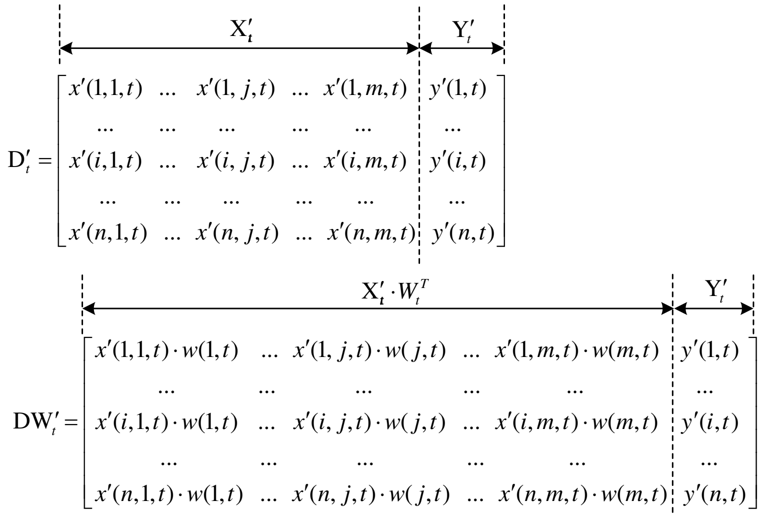
4.2.3 Modification to standard Shepard interpolation: Scaling feature space for multi-source data

The prediction accuracy largely depends on the selection of features and the correlation between the features and the predicted targets. In the context of multi-source data, it is likely that some of the side information is helpful while others are not helpful. As such, a critical challenge to leveraging multi-source data is to build the linkages between the features and the prediction target. The relationships between some features and predicted labels are clear, whereas some others are implicit or context-dependent. For instance, the commuting demand during peak hours will be more sensitive to the working day (X_2) than the rainy day (X_4). If the weights of X_2 and X_4 are identical, then the difference between the target values may be not proportional to the Euclidean distance between these two vectors. This affects the prediction accuracy of the interpolation algorithm. The complexity of passenger demand composition and distinctive behavior further complicates the linkages and bus passenger flow forecasting.

As mentioned previously, the mechanism of the interpolation is that samples of higher similarity in influential factors are closer. A shorter distance between two vectors x indicates the closer corresponding target values y. However, the value of each dimension in x contributes differently to the distance. Therefore, to improve accuracy, it is necessary to assign corresponding weights (i.e., the weighted Euclidean distance) to each dimension according to their contribution to the distance. For this reason, we introduce the weighting process for multi-source data.

Let the weight of the j-th influential factor on the target value $y(t)$ denote as $w(j,t)$, which can be interpreted as the probability that the influential factor of period t cause future passenger flow, or as the relative importance of the influential factor in this period for future forecasting. A larger value of $w(j,t)$ indicates the greater influence of the j-th influential factor in period t on the target value. *This process can be regarded as the scaling of the feature space, by which the dimension of the feature space is reduced or enlarged.* The use of Euclidean distance weighting ensures that the samples of higher similarity are closer to each other in the new feature space. Afterwards, the weights for each dimension are then optimized using historical data.

Fig. 7 illustrates the Euclidean distance weighting, where X'_t denotes the matrix of influential factors; Y'_t denotes the matrix of the peak load; W_t denotes the matrix of Euclidean distance weights, i.e., $W_t = [w(1,t), \dots, w(j,t), \dots, w(m,t)]$; D'_t denotes the matrix of original influential factors and the corresponding peak load; DW_t is a matrix of weighted influential factors and their corresponding peak load. The weighting process is to multiply the influential factor vector of X'_t by the corresponding values of W_t. As a result, Euclidean distance weighting for each dimension can be achieved, such that the value of each dimension in the feature space is reduced or enlarged. By incorporating W_t, the target values y of the data points with shorter Euclidean distance between x will be closer in the feature space. In this way, the multi-dimensional passenger flow data sequences are mapped into the new feature space.



$$\mathrm{D}_t' = \left[\begin{array}{ccccc:c} \multicolumn{5}{c:}{\mathrm{X}_t'} & \mathrm{Y}_t' \\ x'(1,1,t) & \dots & x'(1,j,t) & \dots & x'(1,m,t) & y'(1,t) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'(i,1,t) & \dots & x'(i,j,t) & \dots & x'(i,m,t) & y'(i,t) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'(n,1,t) & \dots & x'(n,j,t) & \dots & x'(n,m,t) & y'(n,t) \end{array}\right]$$

$$\mathrm{DW}_t' = \left[\begin{array}{ccccc:c} \multicolumn{5}{c:}{\mathrm{X}_t' \cdot W_t^T} & \mathrm{Y}_t' \\ x'(1,1,t)\cdot w(1,t) & \dots & x'(1,j,t)\cdot w(j,t) & \dots & x'(1,m,t)\cdot w(m,t) & y'(1,t) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'(i,1,t)\cdot w(1,t) & \dots & x'(i,j,t)\cdot w(j,t) & \dots & x'(i,m,t)\cdot w(m,t) & y'(i,t) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'(n,1,t)\cdot w(1,t) & \dots & x'(n,j,t)\cdot w(j,t) & \dots & x'(n,m,t)\cdot w(m,t) & y'(n,t) \end{array}\right]$$

Fig. 7 Diagram of influential factors of passenger flow and Euclidean distance weighting process (scaling feature space) of each dimension

As a result, the Euclidean distance Eq. (15) in the standard Shepard algorithm is modified as the *weighted Euclidean distance*, and the formulation takes the following form:

$$d_i = \sqrt{\sum_{j=1}^{m} w(j,t)(x'(i,j,t) - x'(n+1,j,t))^2} \tag{16}$$

The optimization process of W_t will be given in the following section. It is worth mentioning that, the optimized weights can be regarded as the relative influence of each influential factor on the passenger flow (see the analysis in Section 6.3). A greater weight indicates higher importance of the influential factor in predicting passenger flow. Therefore, through scaling the feature space, the modified Shepard interpolation algorithm processes strong explanatory power in prioritizing the relative contribution of influential factors to peak load prediction, whereas the standard version cannot. In essence, the scaling process resembles the attention mechanism in deep learning, with the objective to determine which part of information is more valuable to the prediction task.

4.2.4 Modification to standard Shepard interpolation: Parameter optimization

There are two conditions for the application of Shepard interpolation algorithm: (a) the correlation between the predictor and the target value should be statistically significant. (b) the historical sample dataset between the predictor and the target value should be sufficiently representative. To satisfy condition (a), in this study we develop an optimization model to explicitly evaluate the relative influence of each predictor on the target value. The formulations are shown in Eqs. (17)-(19). For condition (b), the penetration of smart cards should be sufficiently high. Taking the studied bus route as an example, the proportion of passengers swiping smart cards accounts for 90%, where the information of the passenger flow can be fully reflected. In addition, the smart-card data could record the historical long-term passenger flow with fine time granularity.

Another key issue in the Shepard interpolation algorithm is to determine the optimal power parameter b based on the historical data. If the value of b is too small, the weight of distant historical values will be too large, such that the fitting surface will be flat and the interpolation accuracy will be insufficient. On the other hand, if the value of b is too large, the predicted value tends to be equal to the nearest historical value, such that the fitting surface will be rough and over-fitting will occur. The value of b can be set as a real number.

As mentioned previously, the shortage cost due to a failure to board, or surplus cost due to excessive available capacity are undesirable from the viewpoint of transit planners and should be eliminated as much as possible. The classic newsvendor model seeks to find out the optimal stock level to minimize the expected surplus and shortage costs. We model the peak load prediction problem herein by drawing an analogy to the newsvendor problem, where both surplus and shortage costs are formulated as functions of the predicted peak load. To this end, by integrating the proposed indicator based on predicted error cost with the modified Shepard interpolation algorithm, the following optimization model is developed to optimally find the model parameters (weights and power parameters) that minimize the total predicted error cost.

$$min \sum_{i=1}^{n} CE_i \tag{17}$$

$$s.t. \quad b_{min} \le b \le b_{max}$$

(18)

$$0 \le w(i,j) \le 1$$

(19)

The objective function Eq. (17) minimizes the total predicted error cost. Eq. (18) represents the power parameter should fall within a domain. Eq. (19) denotes the weight constraint of each influential factor. The parameter b determines the granularity of fitting surface, while $w(i,j)$ affects the predicted value and resulting error cost through Eqs. (13), (14) and (16).

Therefore, the problem is formulated as a nonlinear programming model. The optimization model is non-convex and contains many non-integer decision variables, so it is difficult to find an exact method to solve this model in polynomial time, which is known as a NP-hard (non-deterministic polynomial-time hard) problem. In addition, the solution space will be increased exponentially as the number of influential factors increases. In view of the NP-hardness of the model, a variety of random search algorithms, such as genetic algorithm, can be adopted to obtain the optimal parameters. Specifically, different populations are generated given various values of b. After carrying out the evolutionary process for each population, the optimal individuals of each population can be obtained. As a result, the optimal power parameter b and the corresponding weight values of Euclidean distance W_t are obtained from the optimal individuals.

5 Model extensions

In the above model, the frequency settings are assumed to be determined based on maximum loading point rule (Ceder, 1984). However, the resulting frequencies based only on maximum loading may be conservative when the temporal demand is highly heterogeneous. Our model can be retrofitted to take into account load variations as outlined in Ceder (2007), that is, load-profile method.

The additional information provided by the load profile can be used to overcome the problem of maximum loading point rule. As opposed to the maximum loading point method, the maximum load $y(t)$ is viewed as the ratio A_t/L in the load-profile method. This method can handle demand changes without increasing the available fleet size. The load-profile method is expressed as follows:

$$F_t = max\left(\frac{A_t}{T_t \cdot d_{ot} \cdot L}, \frac{y(t)}{T_t \cdot c}, F_{mt}\right) \tag{20}$$

where A_t represents the area in passenger-km under the load profile in period t and L is the route length. The other notations were previously defined in Eq. (1). The area in passenger-km of each segment can be calculated as the load and the corresponding length. Summing up the area in passenger-km of each segment yields the total passenger-km A_t.

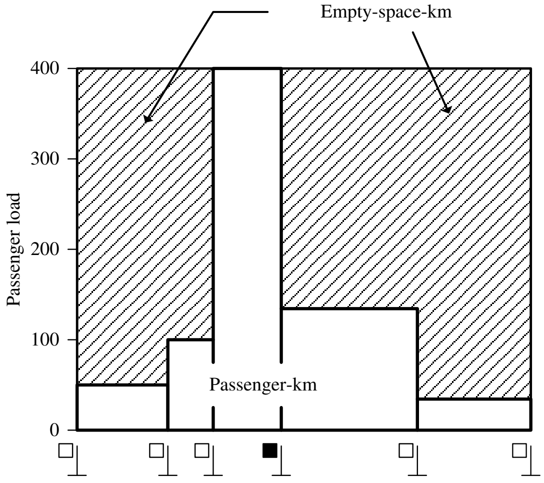


Fig. 8 Illustration of load profile and empty-space-km

Let ρ_t be the density of load profile in period t, which can be calculated from the total passenger-km A_t, divided by the product of the length of the route L and its maximum load $y(t)$. The load-profile density can be used to evaluate the profile characteristics. A large value indicates a relatively flat profile, whereas a small value indicates low variability among the route stops. In practice, the load profile density can be estimated using the historical passenger flow data. If a straight line is drawn across the load profile where the number of passengers is equal to the observed average hourly max load, then the area below this line but above the load profile is the empty-space-km. Therefore, the area of empty-space-km is equal to $(1-\rho_t)\cdot y(t)\cdot L$, and the total passenger-km in period t is dependent on the peak load, which can be estimated as follows:

$$A_t = \rho_t \cdot y(t) \cdot L \tag{21}$$

As a result, the load-profile method can be rewritten as follows:

$$F_t = max\left(\frac{\rho_t \cdot y(t)}{T_t \cdot d_{0t}}, \frac{y(t)}{T_t \cdot c}, F_{mt}\right) \tag{22}$$

When the load-profile method is used to estimate the departure frequency, Eq. (1) can be simply replaced by Eq. (22).

6. Case study

6.1 Case description

The model is tested using the datasets of a route on bus number 60 in Guangzhou, the capital of Guangdong Province in China. The reason of choosing this bus route is due to the heavy demand and optimal schedule requirement. The map of bus route 60 is shown in Fig. 9. There are 21 stops along the route and the total distance extends 16.3 kilometers. The terminals are the Airport Terminal Station and the Olympic Sports Center Station. The bus route passes through Tianhe, Yuexiu and Baiyun District of Guangzhou City. Along this route, urban functional areas such as residential areas, commercial areas, important transportation hubs, hospitals and schools are passed, and the passenger flow composition and influential factors are complex. The buses operating in this route are equipped with GPS devices and the operational data have been completely recorded. The proportion of passengers using smart cards reaches 90%. The direction of the bus route discussed in this article is from the Airport Terminal Station to the Olympic Sports Center Station.

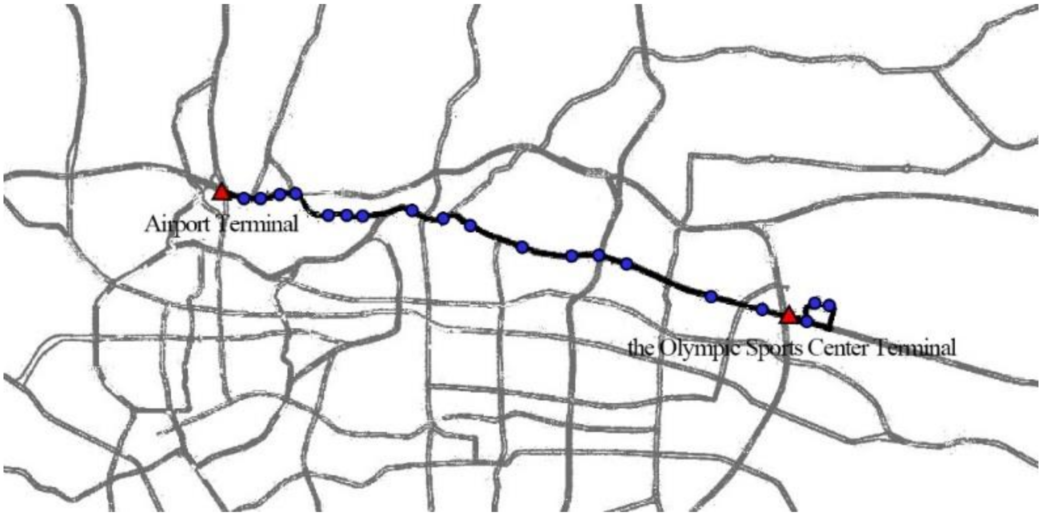


Fig. 9 A map of Bus Route 60 in Guangzhou

The GPS data and smart-card data are collected from the local bus company. The other multi-source data (e.g., weather, air quality, etc) is acquired from the Internet, such as the government information and the China Meteorological Administration website. The dataset is collected on 1/10/2017-12/31/2017 (06:00 am-22:00 pm) from the smart cards and the automotive vehicle location (AVL) system. The dataset is divided into two subsets: the data during 1/10/2017-11/31/2017 for training and the remainder for testing. In the base case, the length of the period T_t is set as 60 min. The value of waiting time C_p is set as 10 RMB/h, and the direct cost of operating a standard vehicle C_b is taken as 60 RMB per trip. The vehicle capacity c is taken as 100 pax. The load factor γ_t is taken as 0.5. The ticket price P is taken as 2 RMB. The minimum power parameter b_{min} and maximum power parameter b_{max} are set as 0.1 and 10, respectively. The other default settings are: $r_0 = 0.2$; $r_1 = 0.1$; $\beta = 0.1$. These default parameters remain the same expect where they are the subject of sensitivity analysis.

In what follows, we first present how data is processed in this paper. As the Euclidean distances and power parameter are the decision variables of our model, we proceed to present the optimized results and their practical implications, followed by the sensitivity analysis to model parameters. Subsequently, we compare our model with other traditional machine learning models. Finally, we highlight the key findings and managerial insights.

6.2 Data processing

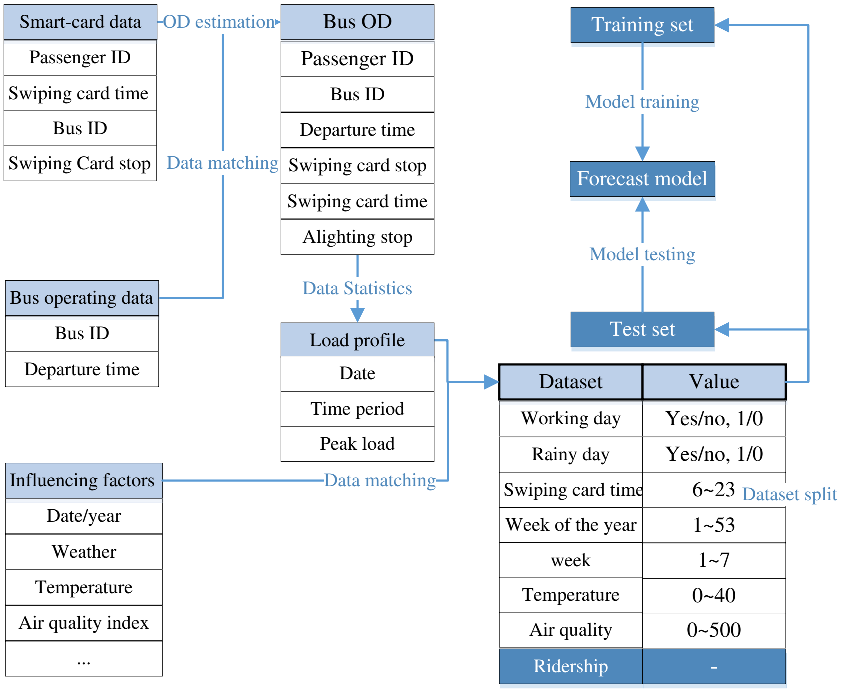


Fig. 10 Overview of the data processing

The predicted results are the peak load in the target period of a target date. Fig. 10 shows how data is processed in this paper. The main steps are described as follows:

1) Extract the information about the swiping card times using smart-card data, and obtain the vehicle ID and the departure time using GPS data.

2) Infer the load profile and corresponding peak load by matching the smart-card data and GPS data. The following steps describe the detailed process:

a. Extract the information on all trips within the time windows and OD information of each passenger on a designated bus trip.

b. Compute the load profile along the bus route. The throughput at each stop in a direction is the number of passengers boarding before this stop (including this stop) minus the number of passengers alighting after this stop (including this stop).

c. Obtain the dataset of peak load in each hour. The maximum value of the stop-specific load along the route in period t is the peak load $y(i,t)$, where i represents the i-th historical experiment point.

3) Split the overall dataset. Match the statistics dataset of the peak load and the corresponding influential factors in a chronological order to obtain the training dataset $D_t{}'$. The influential factors of the training dataset are quantified to construct the feature space in the interpolation forecast algorithm.

4) Select one time point and set the dataset before this time point as the training dataset, while setting the dataset after this time point as the testing dataset. Note that the time point is a day of the year.

5) Use the training dataset to train the forecast model, and test the forecast model with the testing dataset. It proceeds to obtain the indicator results. Typically, for a bidirectional bus route, the peak load of upstream and downstream directions should be predicted separately.

It is worth noting that, to obtain the historic peak load and the total passenger-km, the bus OD matrix and historical load profile should be acquired in advance. The load profile can be estimated given the bus OD matrix. Bus OD matrix can be directly obtained by automatic fare collection (AFC) system that can provide user information for both boarding and alighting locations. For the entry-only smart card data, a variety of methods in the literature could be used to forecast the destinations of smart-card users (Jung and Sohn, 2017). Since the smart-card data is entry-only in Guangzhou, in this study we estimate the bus OD matrix using the alighting probabilities estimation method following Liu et al (2013) and Chen et al (2015), that is, passengers boarding at a designated stop are assumed to evenly alight at the remaining stops. However, in principle, it can be substituted by any other methods or empirical rules, which would not affect the generalization of the framework.

6.3 Euclidean distance weights optimization and interpretation

Due to the diversity of passenger flow composition and travel behavior, each influential factor contributes differently to the passenger demand in different time periods. For this reason, we use the genetic algorithm to optimize the weights of the Euclidean distance to the target value of each attribute in each hour. The weight can be regarded as the relative influence of each influential factor on the passenger flow. A greater weight indicates higher importance of the influential factor in predicting passenger flow. The optimal solutions are shown in Fig. 11. Based on the optimized weights, the relative importance of each influential factor can be further identified and ranked, and the results are presented in Table 2.

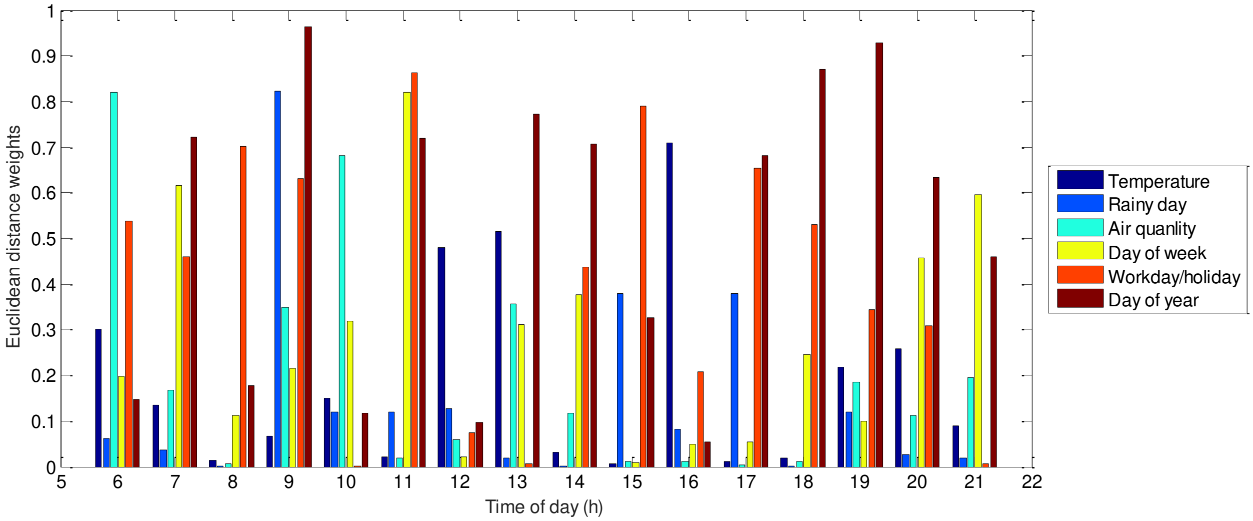


Fig. 11 Euclidean distance weights between attributes and the passenger flow

As we can see, each influential factor contributes differently to passenger demand. During the typical peak hours (e.g., 7:00 am-9:00 am; 17:00 pm-20:00 pm), workday/holiday attribute is a significant contributing factor. This is because in the peak hours the majority of passengers are commuters who are quite sensitive to the working day factor. Moreover, the day of year factor plays an important role relative to other factors, which indicates that the peak-hour demand presents seasonal variation pattern in the long term. In addition, the weekday factor is also a great contributor, which suggests that the volume of peak hour commuters presents recurrent fluctuation within a week.

Table 2 Relative contribution of influential factors to passenger follow in typical period

	8:00 am-9:00 am		12:00 pm-13:00 pm		17:00 pm-18:00 pm	
Influential factors	Rank	Relative importance (%)	Rank	Relative importance (%)	Rank	Relative importance (%)
Temperature	4	1.36	1	55.78	5	0.65
Rainy day	6	0.23	2	14.86	3	21.24
Air quality	5	0.68	5	6.96	6	0.16
Day of week	3	11.02	6	2.46	4	3.04
Workday	1	69.12	4	8.58	2	36.67
Day of year	2	17.61	3	11.35	1	38.23

During the off-peak period, many factors affect the passenger flow. This is because the demand composition during this period is rather complex and the travel patterns are diverse. For example, during 6:00 am-7:00 am, the temperature and air quality considerably contribute to the passenger demand apart from the workday/holiday attribute. This implies that in addition to the morning commuters, the passengers who are sensitive to temperature and air quality will make their trips collectively in the early morning. Temperature contributes most to the passenger demand during some off-peak hours at noon and afternoon (e.g., 12:00 am-13:00 pm; 16:00 pm-17:00 pm). The reason is that the high temperature during the mid-afternoon off-peak hours largely affects the elastic demand. The impact of influential factors on passenger flow during 19:00 pm-22:00 pm is rather complex. This is because the complexity of demand composition and travel behavior during this period are the highest throughout the day. Each influential factor fairly contributes to the generation of the passenger flow.

6.4 Optimization of the power parameter

As the power parameter represents the relative effect of the dissimilarity in influential factors on the target value, it is interesting to explore how the value of b changes with time periods and how the value of b impacts the error cost. To optimally find out the values, different populations are developed under various values of power parameter *b* for given time period. That is, each population evolves separately, and the solution of each population is taken as a local optimal solution. Finally, the optimal solution is selected from the local optimal solutions. The population size for the genetic algorithm is taken as 200. The crossover and mutation probability in genetic algorithm are taken as 0.6 and 0.2, respectively. The maximum evolutionary iteration is set as 200. Since the departure time of first bus and last bus of this route is 6:00 am and 22:00 pm, we only provide the results during 6:00 am-22:00 pm.

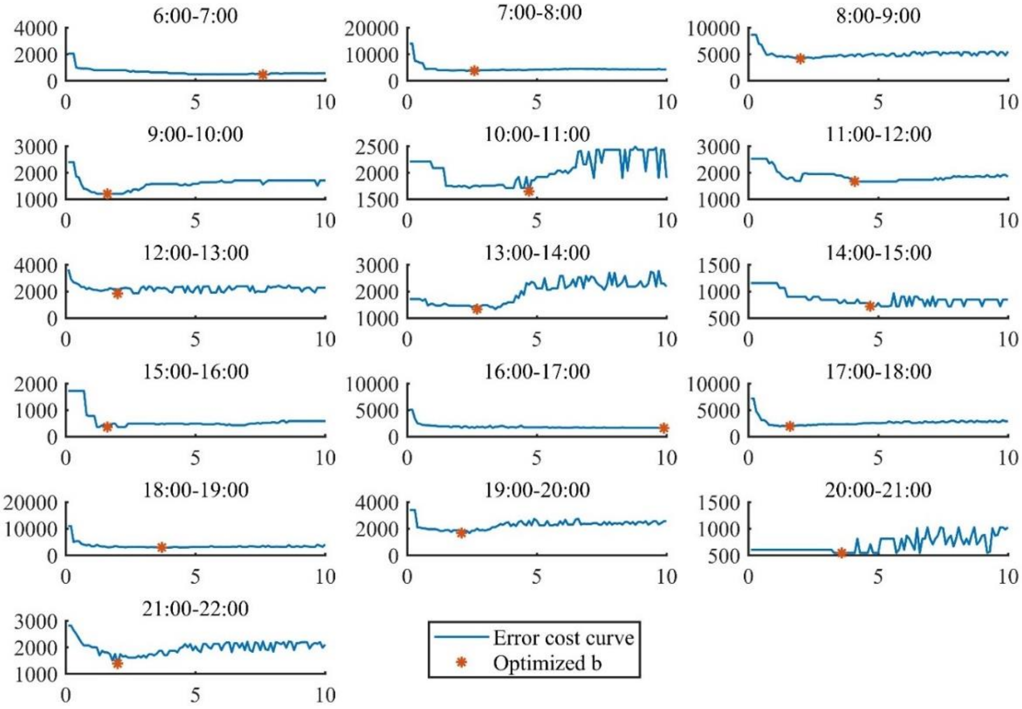


Fig. 12 Cost profile and the minimum cost under different values of *b*

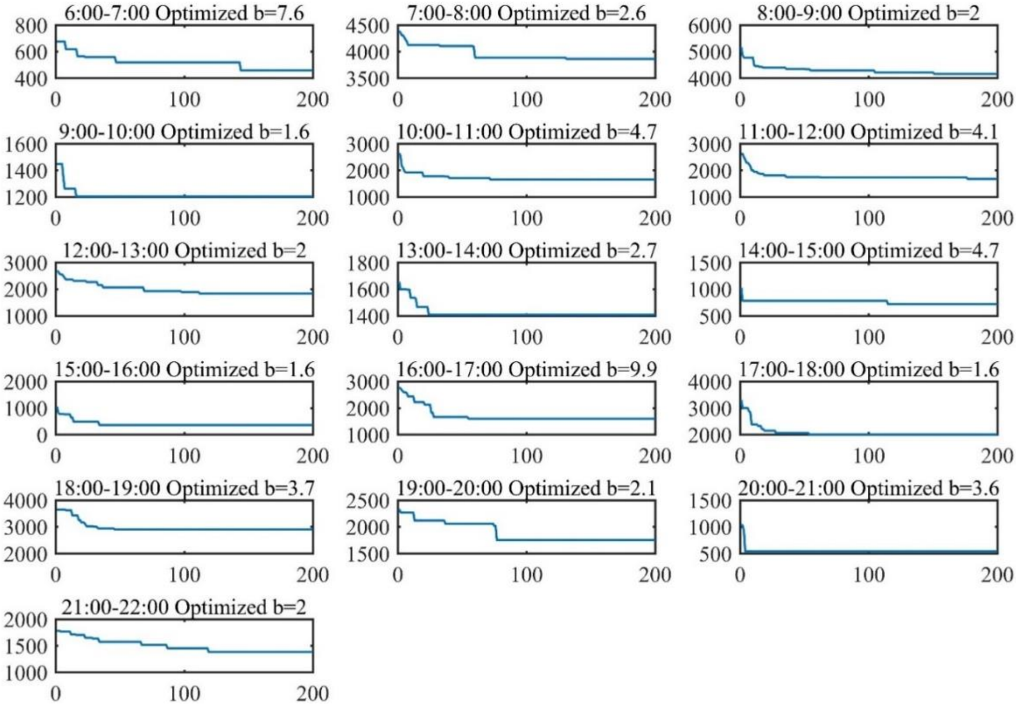


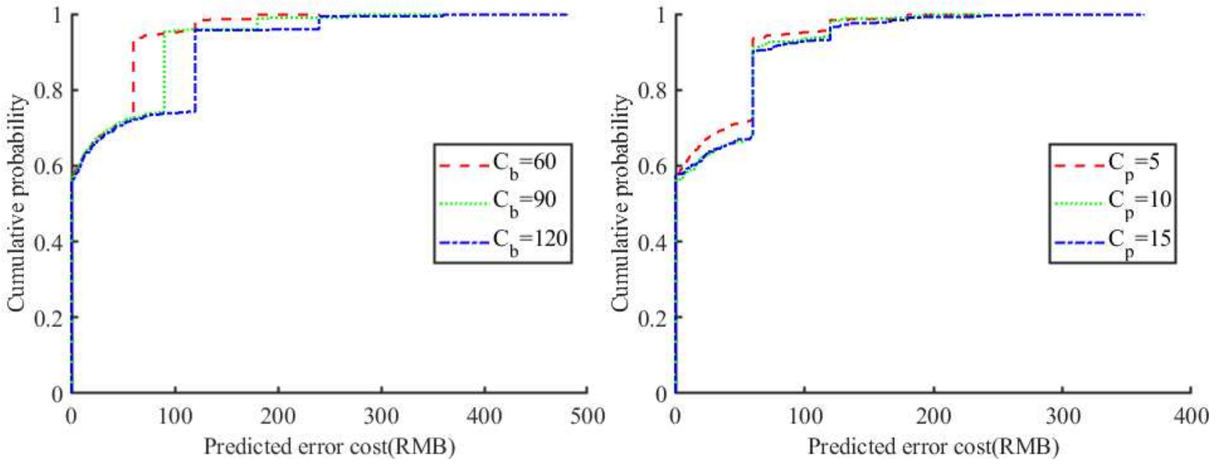
Fig.13 Cost convergence curves with optimal b

Fig. 12 presents the minimum error cost under various values of b for different time periods and the corresponding optimized values. One can see that the optimized value of b differs in each time period. Generally, the minimum error cost first decreases then increases as the value of b grows. This is because, as mentioned in Section 3.4.2, this parameter determines the granularity of fitting surface. A value too large or too small may cause under-fitting or over-fitting problems, which results in poor performance of the prediction model. There exists an optimal value in the range between 1 and 10, and the majority of the optimized values are smaller than 5. The error cost profiles of the off-peak hours (e.g., 10:00 am-15:00 pm; 19:00 pm-22:00 pm) appear to fluctuate more remarkably than those of peak hours. This is due to the complex interactions between the power parameter and Euclidean distance weights in the optimization of loss function. More specifically, as discussed in Section 6.3, the working day attribute is dominant among the influential factors in the peak hours, whereas in the off-peak hours the influential factors vary as a result of complex passenger flow composition.

Fig.13 shows the curves of cost convergence in each period under optimized value of b. The result shows that the evolution converges rapidly (before 100 generations) for typical periods of morning and evening peak hours (e.g., 7:00 am-8:00 am; 9:00 am-10:00 am; 13:00 pm-14:00 pm; 17:00 pm-18:00 pm) where the passenger flow is greatly affected by the working day factor. On the other hand, the evolution converges slowly (after 100 generations) in other time periods (e.g., 6:00 am-7:00 am; 12:00 pm-13:00 pm, 21:00 pm-22:00 pm). This is due to the interactions between many influential factors in the presence of complex passenger flow composition.

6.5 Sensitivity analysis

In this section, we conduct a sensitivity analysis of three key parameters (C_b, C_p and d_{0t}) in our model. Fig. 14 presents the sensitivity to the trip cost C_b and the value of waiting time C_p. The parameters C_b and C_p can be interpreted as the penalties for insufficient and excessive capacity. Fig. 14(a) shows the cumulative probability of predicted error cost under different values of C_b. A steeper slope in the cumulative probability distribution indicates a narrower error distribution and thus better worst-case forecast. The first observation is that as the trip cost C_b increases, the cumulative probability converges to one more slowly (and thus the predicted error cost becomes larger). This is expected since the error cost is positively related to the passenger waiting time costs (Eq. (8)). Moreover, there exists "step phenomenon" for the cumulative probability. The curve jumps vertically at the integer multiples of C_b, and the jump amplitude is considerably reduced as the error cost increases. This is because the error cost of operating an extra bus is C_b, and most of the error costs of this model are concentrated in the range with low predicted error costs. However, the vertical jump when the multiplier of C_b is greater than 1 is not significant because the model accuracy is sufficiently high, such that a very limited number of predicted error costs reach an integer multiple of C_b. The increase in error cost between the integer multiples of C_b is mainly due to the increase of additional waiting time cost.



(a) C_b

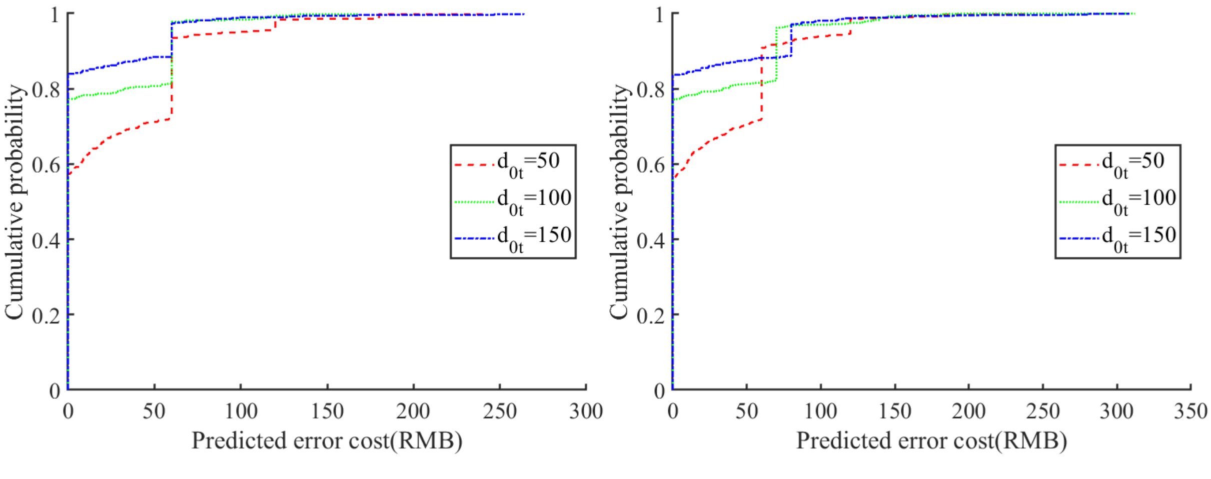
(b) C_p

Fig.14 Sensitivity analysis to C_b and C_p

Fig. 14(b) shows the cumulative probability of error costs for different values of waiting time C_p. It can be seen that a lower value of C_p leads to better performance as the curve of cumulative probability converges faster between consecutive jumps. This is because the error cost is proportional to the passenger waiting time costs (Eq. (8)). Interestingly, the jump of the cumulative probability curve still occurs at the integer multiples of a standard trip cost C_b, and the jump amplitude is reduced with the increase of error costs.

Next we investigate the effect of vehicle capacity on prediction performance. Generally, the trip cost C_b increases with the bus size or vehicle capacity. To evaluate the independent effect of desired in-vehicle occupancy, we first conduct sensitivity analysis via keeping the trip cost constant and varying only the desired in-vehicle occupancy. Subsequently, the trip cost is assumed to vary proportionally with the bus size so as to analyze the trade-off between desired in-vehicle occupancy and trip cost.

Fig. 15(a) shows the sensitivity to the in-vehicle occupancy d_{0t} with the base trip cost. Similar to Fig. 14(a), the curve jumps vertically at the integer multiples of C_b, and the jump amplitude is reduced with the increase of the error cost. One can see that the predicted error cost decreases as the desired in-vehicle occupancy increases. When the in-vehicle occupancy increases from 50 to 150 pax, the percentage of forecast results without the error costs increases from 57% to 83% (26% improvement), while that with the error cost of C_b increases from 92% to 96% (4% improvement). This suggests that with higher capacity, the predicted error costs can be reduced, and the model is more robust to predicted errors. The reason is that, as indicated in Fig. 3, larger vehicle capacity means a wider range of ineffective errors and greater tolerance for the predicted errors. In other words, the predicted errors in the peak load are less likely to affect the optimal number of trips when the vehicle capacity is larger, thereby reducing the cost loss by the predicted errors.



(a) Independent with trip cost

(b) Dependent with trip cost

Fig.15 Sensitivity analysis to the desired in-vehicle occupancy d_{0t}

The independent effect of in-vehicle occupancy shows that the predicted error cost decreases as the vehicle capacity increases (given the load factor), whereas a larger vehicle capacity also indicates higher trip cost. As concluded from Fig. 15(a), as the trip cost C_b increases, the predicted error cost becomes larger since the cumulative probability converges to one more slowly. Therefore, a trade-off may exist between prediction quality and additional operation cost associated with the vehicle size (such as additional fuel consumption due to greater weight and additional parking areas).

To further verify such a cost trade-off, let us assume the trip cost vary proportionally with the bus size, that is, the trip cost C_b equals to 60, 70, and 80 RMB when the desired in-vehicle occupancy is 50, 100, and 150 pax, and the result is shown in Fig. 15(b). As we can see, the percentage of forecast results without the error costs increases when the desired in-vehicle occupancy increases, which is consistent with Fig. 15(a). However, the curve jumps vertically at the integer multiples of C_b, which is in line with Fig. 14(a). This suggests that a higher capacity could also make the cumulative probability converge to one more slowly (and thus the predicted error cost may become larger). As a result, the increase in desired in-vehicle occupancy (or vehicle capacity) can either increase or decrease the predicted error cost, depending on the relationship between the bus size and direct trip cost.

6.6 Model comparisons

To evaluate the forecasting performance of the proposed model, a set of traditional error-based machine learning models are adopted for comparisons using the same dataset, i.e., Decision Tree (DT), Neural Network (NN), Random Forest (RF), K-Nearest Neighbors (KNN), and Linear interpolation (LI), where KNN and LI belong to interpolation methods. The proposed model, KNN, and LI are deterministic models in that the prediction output is unique, whereas the others are stochastic models where prediction output may differ in each simulation run.

Table 3 Model characteristics

Model	Prediction output	Hyper parameters	Model parameters
Proposed	Deterministic	-	*b, w*
DT	Stochastic	Minimum number of leaves	-
RF	Stochastic	Number of trees Minimum number of leaves	-
NN	Stochastic	Number of hidden layers Number of hidden layer nodes Number of input layer nodes Number of output layer nodes	Weights
KNN	Deterministic	Number of nearby points	-
LI	Deterministic	-	-

To begin with, the hyper parameters and (if any) model parameters are optimized by grid search, which is to manually specify the subset of parameter space and explore the optimal combination of parameters

(Lin et al., 2018). The specified hyper parameters and model parameters of each model are presented in Table 3.

Since our model is cost-based, to conduct a comprehensive and fair comparison, the comparative experiment is performed in two dimensions: a) comparison between the proposed model and traditional error-based models; and b) comparison between cost-based models. In the second dimension, the loss functions of traditional error-based models are replaced by cost-based indicators in the optimization of hyper parameters and model parameters. Each experiment is repeated by 20 times, and the results are shown in Table 4, including MAE and sum of error cost. As we can see, our model outperforms other models in terms of the sum of error cost. Meanwhile, the MAE of our model is close to those of DT and RF, while lower than those of other models.

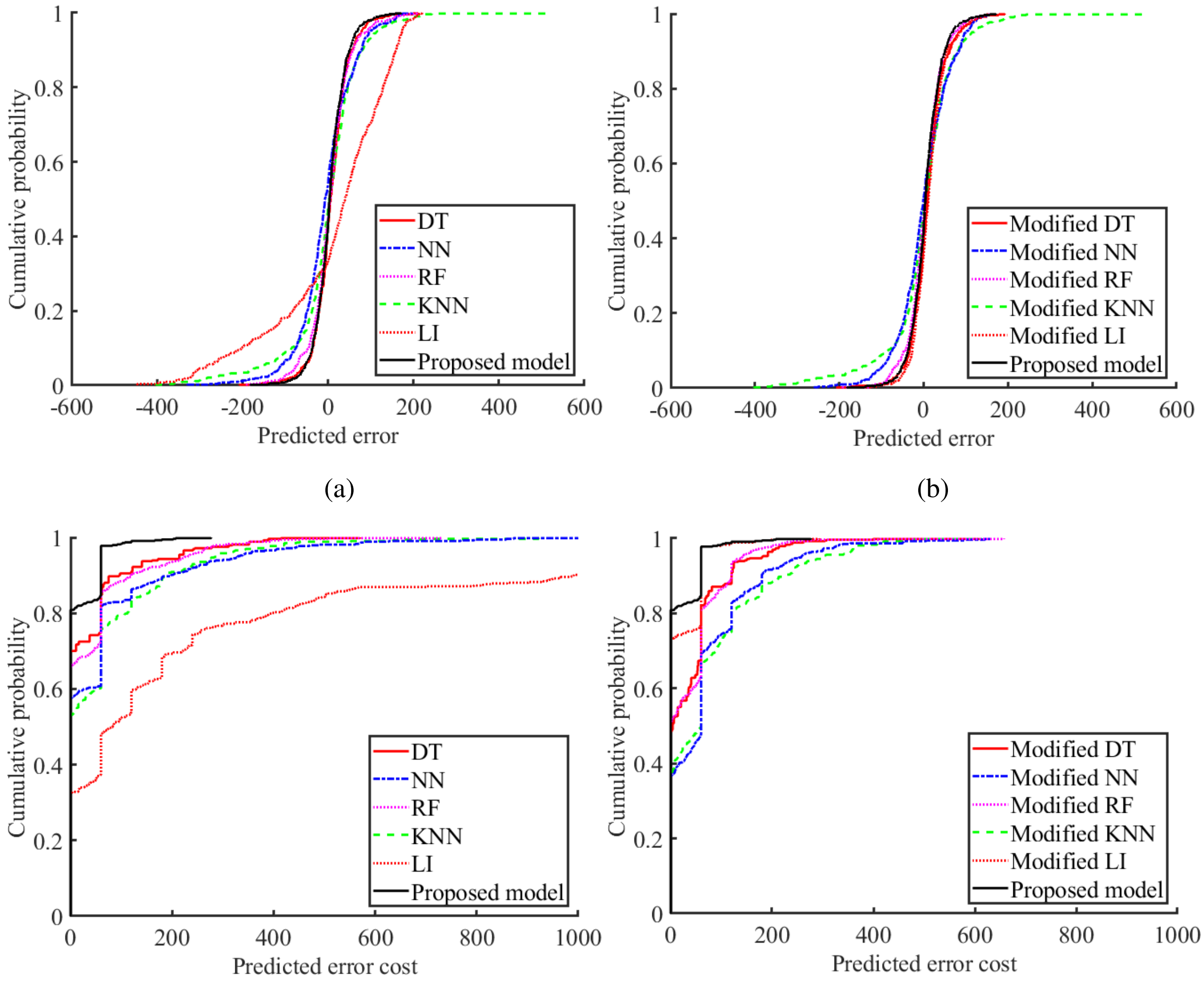
Table 4 Comparative results of different models

	Error-based		Cost-based	
Model	MAE	Sum of error cost (RMB)	MAE	Sum of error cost (RMB)
Proposed	-	-	30.01	6741.39
DT	29.54	16891.58	29.44	16651.16
RF	29.56	16542.81	32.91	19657.40
NN	63.26	54981.92	79.63	82921.87
KNN	52.71	31918.99	52.67	31882.55
LI	108.08	129237.57	29.98	9365.75

After replacing the loss function, the changes of MAE are trivial for DT and KNN, while the sum of error cost is reduced. However, the sum of error cost for RF and NN increase in turn. The possible reason is that the standard RF and NN are stochastic models which are easy to be trapped to local minima (thus result in the instability), while the “step phenomenon” of error cost worsens the instability of prediction outputs. This reinforces the message that our model presents superiority over traditional machine learning models for predicting peak load of bus routes.

Fig. 16 presents the distributions of accumulated absolute errors and accumulated predicted error costs of each model. Note again that a steeper slope in the cumulative probability distribution is an indication of narrower error distribution and better worst-case forecast. The modified models in the figures represent those with cost-based indicators. As shown in Figs. 16(a) and (b), the error distribution of our model appears to be close to DT and RF, while presenting narrower error distribution as compared to KNN, LI and NN. With respect to error cost (Figs. 16(c) and (d)), our model has the fastest slope change rate and the narrowest error distribution, and the worst-case forecast of our model is considerably better than those of other models.

The error costs of our model are more concentrated in the range of lower costs. More specifically, about 81% of the predicted error costs of our model are lower than 60 RMB. There are two reasons for this. First, our model can make a cost trade-off between the shortage component and the surplus component to reduce the total error cost. Second, given the desired in-vehicle occupancy $d_{0t} = 100$, most of prediction errors are distributed in the non-effective region. In other words, predicted errors exist in most of the forecasting results but without additional error costs. These results further verify that our model can provide relatively more accurate and robust prediction than its alternatives.



(a)

(b)

(c)

(d)

Fig. 16 Distributions of cumulative predicted error (cost) for different models: (a) predicted errors of error-based models; (b) predicted errors of cost-based models; (c) predicted error cost of error-based models; (d) predicted error cost of cost-based models

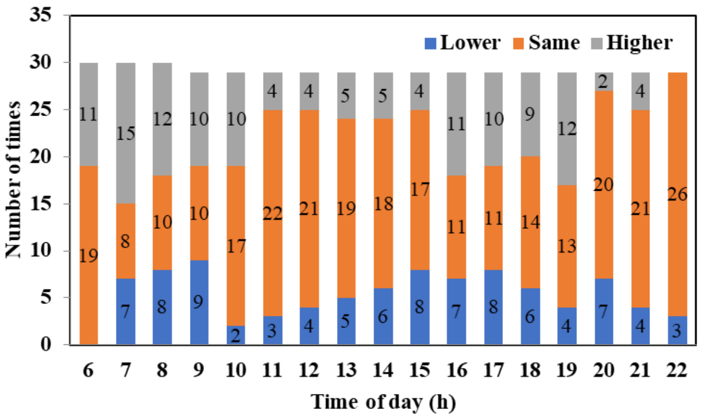
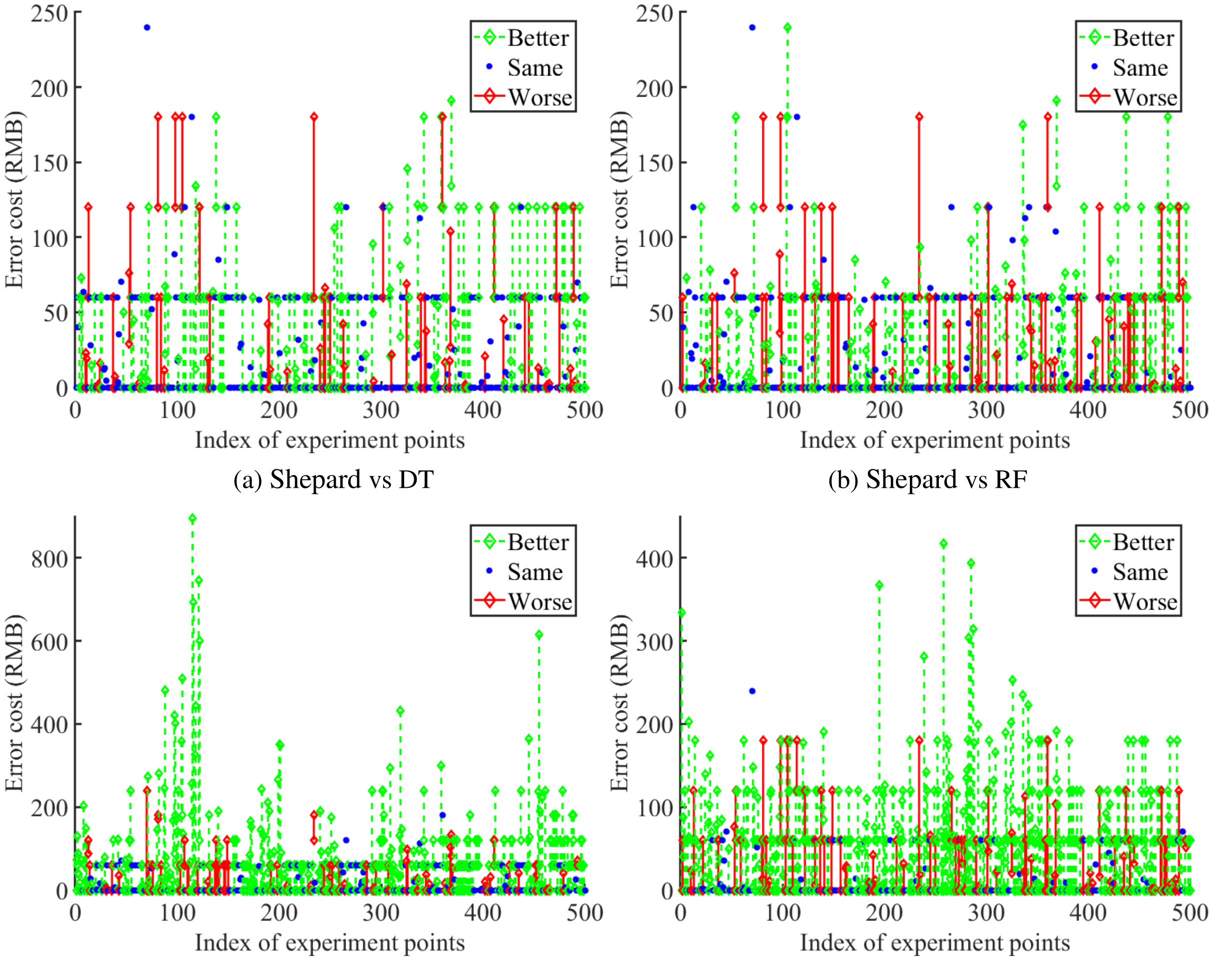


Fig. 17 Relative size relationship of transit supply over different time of day for the testing set

As the predicted peak load is used to predictively optimize the transit supply (number of trips) during the planning horizon, the relationship between the 'real' supply and 'predicted' supply deserves some discussion. To this end, we compare the hourly number of trips for the testing set. According to Eq. (2), the values of real (the predicted) number of trips per hour are calculated using the real (the predicted) peak load data in the testing set. Fig. 17 presents the aggregated relative size relationship between the predicted values and real values over time of day for the testing set using the scaled Shepard interpolation algorithm. The height of the orange bar chart indicates the number of times that the predicted number of trips equals to the real number of trips. The height of the grey bar chart and blue bar chart indicate the number of times that the predicted number of trips is higher and lower than the real number of trips, respectively. As we can see, during the peak hours (e.g., 7:00 am-11:00 am; 16:00 pm-20:00 pm), the predicted number of trips tend to be higher than the real values. During the off-peak hours (e.g., 12:00 pm-16:00 pm; 20:00 pm-23:00 pm), the predicted number of trips is likely to be either lower than or equal to the real values. These results are associated with the cost trade-off between the surplus component and shortage component. This suggests that, during the peak hours, to achieve the minimum total cost, a larger number of departure trips are expected to reduce the passenger waiting time and the resulting shortage cost due to the heavy demand. On the other hand, the number of trips lower than the real values indicates that the surplus cost is dominated in the system cost.

Given the relative size relationship between predicted supply and real supply, their discrepancy of quality also deserves some discussion. Fig. 18 shows the comparison between Shepard interpolation and traditional models in the quality of trip supply determination for the testing set. The quality is measured by the resultant predicted error cost. The green dashed line represents an instance where the error cost of the proposed model is lower than that of the traditional model. The upper end corresponds to the error cost of the traditional model, while the bottom corresponds to that of the traditional model. The length of the dashed green line represents the gap between the proposed model and traditional model. Conversely, the red solid line represents an instance where the error cost of the proposed model is higher than that of the traditional model. The upper end corresponds to the error cost of the proposed model, while the bottom corresponds to that of the proposed model. The blue dot represents an instance where the error cost of the proposed model is equal to that of the traditional model. Counting the respective instances yields the quality distribution of trip supply determination shown in Table 5. The results show that the proposed model outperforms its counterparts in producing a larger number of better trip supply plans in that the proportion of better cases exceeds those of worse cases. The advantages are more obvious when compared to the KNN and NN, which is consistent with the gaps of predicted error costs as shown in Table 4.

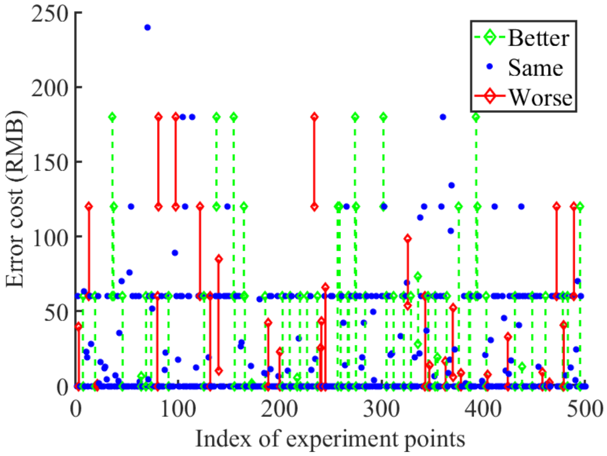


(a) Shepard vs DT

(b) Shepard vs RF

(c) Shepard vs KNN

(d) Shepard vs NN



(e) Shepard vs LI

Fig. 18 Comparison between Shepard interpolation and traditional models in the quality of trip supply determination for the testing set

Table 5 Quality distribution of trip supply determination for the testing set

Counterpart	Better (%)	Worse (%)	Same (%)
Shepard vs DT	23.71	10.81	65.46
Shepard vs RF	23.14	13.85	62.99
Shepard vs NN	53.32	14.61	32.06
Shepard vs KNN	46.47	11.38	41.93
Shepard vs LI	11.38	5.50	83.11

6.7 Frequency settings considering load variations

As discussed in Section 5, our model can be also applicable for the load-profile method that takes into account load variations. In this case, the values of departure frequencies are calculated by Eq. (22). Using the historic load data and the length of each route segment, the average load profile density at different time of day can be calculated, and the results are shown in Fig. 19. Fig. 20 shows the number of trips per hour for different methods for the testing set, where the frequencies of max-load method are calculated by Eq. (1). As we can see, the frequencies of load-profile method are generally lower than those of max-load method. This results from the trade-off between the load variations and max load in the load-profile method.

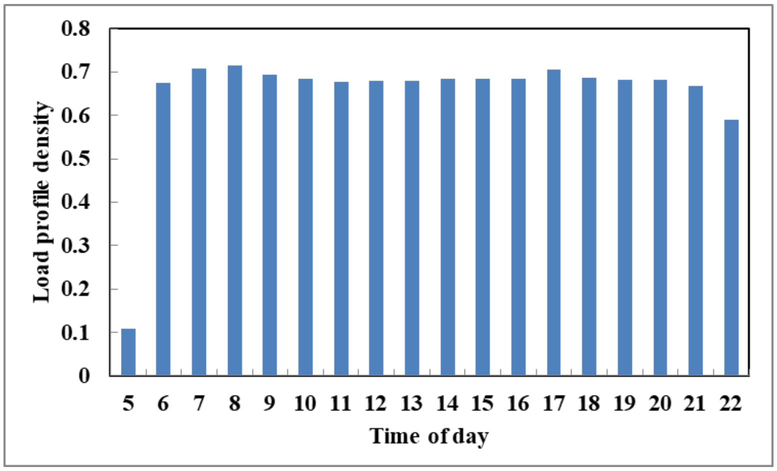


Fig. 19 Load profile densities at different time of day

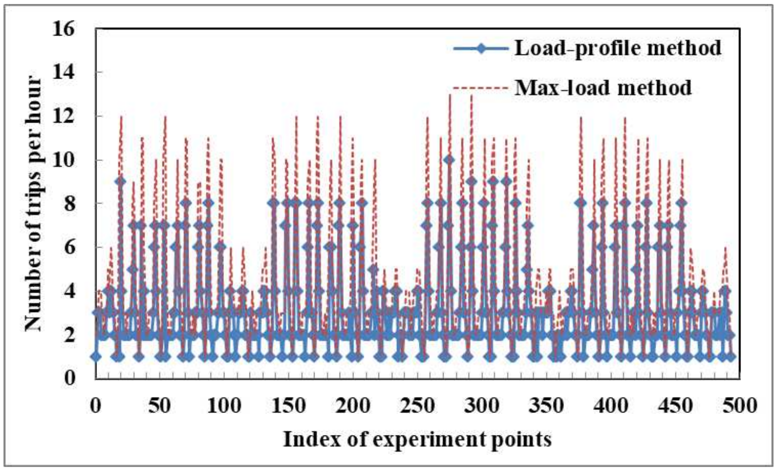


Fig. 20 Number of trips per hour for different methods for the testing set

7. Concluding remarks

The level of service on public transit routes is directly affected by the frequency and resultant number of trips in the planning horizon. Unlike general demand forecasts in other field, in the context of peak load prediction, there exists non-effect region for predicted errors within which the optimal number of trips remains unchanged, while the effective errors (positive and negative errors) contribute differently to the costs associated with the inadequate and excessive available capacity. The main contribution of this paper is a new framework for predicting the peak load of bus routes that explicitly combines demand prediction with supply optimization. We introduce a new cost-based indicator in the context of peak load prediction, which is able to comprehensively capture both shortage and surplus costs due to the effect of insufficient or excessive trips. Another key contribution is a scaled Shepard interpolation algorithm that can resolve discontinuities in the probability distribution of prediction errors arising from the new indicator, while leveraging the potential efficacy of multi-source data through a novel quasi-attention mechanism (scaling feature space and parameter optimization). The cost-based indicator is coupled with the modified Shepard interpolation algorithm. Our model is scalable and can be extended to other frequency setting method (load-profile method) that takes into account load variations.

The proposed model was tested using the data of a real-world bus route in Guangzhou, China. We analyzed the sensitivity of model performance under different operational settings. Interestingly, the error cost profile exhibits step phenomenon. The performance of the proposed method showed a clear improvement in the accuracy and stability of predicted error costs as compared to the state-of-the-art methods. Our method can also produce a larger number of better trip supply plans as compared to traditional methods. More importantly, unlike traditional machine learning approaches usually applying "black-box" procedures, our method has superior performance in model interpretation power, where the relative influences of influential factors on peak load prediction can be identified and ranked.

Based on the key findings described above, the following practical insights and recommendations can be drawn.

(a) *Bus scheduling (re)design*. Although the increase in vehicle capacity can enhance the model robustness, it can also either increase or decrease the predicted error cost, depending on the relationship between the bus size and direct trip cost. Therefore, in the design or redesign of bus scheduling, transit planners should make a trade-off between the prediction quality and additional operation cost associated with the vehicle size (such as additional fuel consumption due to greater weight and additional parking areas) and possibly introduce the vehicle with higher capacity.

(b) *Prediction philosophy and parameters determination.* As there will be both ineffective and effective errors in the prediction of peak load, from the standpoint of transit authorities, it is not necessary to pursue only the accuracy in terms of the volume. Instead, it makes sense to pursue the minimization of predicted error costs, which involves a trade-off between economic viability of the system and maintaining good service for passengers. This can be done by defining different ratios of C_b/C_p for different time periods according to the scheduling philosophy of bus operators. For example, during the peak hours, the additional waiting time cost is more detrimental to the system than the extra operation cost due to the heavy demand. As poor service quality indicates a drop in demand and possible loss of route operating franchise, it would be helpful to decrease the ratio of C_b/C_p to reduce the shortage cost. On the other hand, during the off-peak hours, the extra operation cost may be more harmful than the additional waiting time cost. Then it is rational to increase the ratio of C_b/C_p to reduce the surplus cost.

It is generally expected that the emergence of "big data" can help build better prediction models. Due to limited resources, the selected influential factors in this article are mostly external factors, such as temperature and weather. However, in practice, internal factors such as passenger attributes and external influences may interact with each other. In the future, we will collect and fuse more datasets to improve the performance of prediction model.

Acknowledgements

This work is partly supported by the National Science Foundation of China (Project No. 61703165; 71890972; 71890970), the UK Department for Transport (Project: “Future Streets”), and the Royal Academy of Engineering (Grant: UK-CIAPP\286).

REFERENCES

Amiripour S.M., Ceder A., Mohaymany A.S., 2014. Designing large-scale bus network with seasonal variations of demand, Transportation Research Part C, 48, 322-338.

Bai T., Wu M., Zhu S., 2019. Pricing and ordering by a loss averse newsvendor with reference dependence, Transportation Research Part E, 131, 343-365.

Ceder A., 1984. Bus frequency determination using passenger count data, Transportation Research Part A, 18(5), 439-453.

Ceder A., 2007. “Public transit planning and operation: theory, modelling and practice,” Elsevier.

Chen J., Liu Z., Zhu S., Wang W., 2015. Design of limited-stop bus service with capacity constraint and stochastic travel time, Transportation Research Part E, 83, 1-15.

Furth P.G., Wilson N.H.M., 1982. Setting frequencies on bus routes: Theory and practice, Transportation Research Record: Journal of the Transportation Research Board, 818, 1-7.

Gordon J., Koutsopoulos H.N., Wilson N.H.M., 2018. Estimation of population origin-interchange-destination flows on multimodal transit networks. Transportation Research Part C, 90, 350-365.

Gkiotsalitis K., Wu Z., Cats O., 2019. A cost-minimization model for bus fleet allocation featuring the tactical generation of short-turning and interlining options, Transportation Research Part C, 98, 14-36.

Gkiotsalitis K., Cats O., 2018. Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways, Transportation Research Part C, 88, 187-207.

Hadas Y., Shnaiderman M., 2012. Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time, Transportation Research Part B, 46(8), 1068-1084.

Herbon A., Hadas Y., 2015. Determining optimal frequency and vehicle capacity for public transit routes: A generalized newsvendor model, Transportation Research Part B, 71, 85-99.

Jiang X., Zhang L., Chen X., 2014. Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China, Transportation Research Part C, 44, 110-127.

Jung J., Sohn K., 2017. Deep-learning architecture to forecast destinations of bus passengers from entry only smart-card data, IET Intelligent Transport Systems, 11(6), 334-339.

Khouja M., 1999. The single-period (news-vendor) problem: Literature review and suggestions for future research, Omega, 27(5), 537-553.

Li, S., Liu, R., Yang, L., Gao, Z., 2019. Robust bus controls considering delay disturbances and passenger demand uncertainty. Transportation Research Part B, 123, 88-109.

Liu R., Sinha S., 2007. Modelling urban bus service and passenger reliability. Paper presented at the International Symposium on Transportation Network Reliability, The Hague, July 2007.

Liu L., Chen R.. 2017. A novel passenger flow forecast model using deep learning methods, Transportation Research Part C, 84, 74-91.

Lin L., He Z., Peeta S., 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach, Transportation Research Part C, 97, 258-276.

Lin L., John C. Handley, Gu Y., Zhu L., Wen X., and Sadek A., 2018. Quantifying uncertainty in short-term traffic prediction and its application to optimal staffing plan development. Transportation Research Part C, 92, 323-348.

Liu Z., Yan Y., Qu X., Zhang Y., 2013. Bus stop-skipping scheme with random travel time. Transportation Research Part C, 35(9), 46-56.

Liu Y., Liu Z., Jia R., 2019. DeepPF: A deep learning based architecture for metro passenger flow prediction. Transportation Research Part C, 101, 18-34.

Ma Z., Xing J., Mesbah M., Ferreira L., 2014. Predicting short-term bus passenger demand using a pattern hybrid approach, Transportation Research Part C, 39, 148-163.

Ma X., Ding C., Luan S., Wang Y., Wang Y., 2017. Prioritizing influential factors for freeway incident clearance time prediction using the Gradient Boosting Decision Trees Method, IEEE Transactions on Intelligent Transportation Systems, 18(9), 2303-2310.

Noursalehi P., Koutsopoulos H., Zhao J., 2018. Real time transit demand prediction capturing station interactions and impact of special events. Transportation Research Part C, 97, 277-300.

Shepard D. 1968. A two-dimensional interpolation function for irregularly-spaced data, in: Proceedings of the 23rd National Conference, ACM, New York, 517-523.

Shalaby A., Farhan A. 2003. Bus travel time prediction model for dynamic operations control and passenger information systems, Presented in the 82nd Annual Meeting of the Transportation Research Board, Washington DC, Jan. 2003.

Shi T., Xia Q., 2018. A cascadic multilevel optimization algorithm for the design of composite structures with curvilinear fiber based on Shepard interpolation. Composite Structures, 188, 209-219.

Tang, T., Liu, R, Choudhury, C, 2020. Incorporating the effect of weather conditions and travel history in alighting stop estimation using smart card data. Sustainable Cities and Society. 53, 101927.

Wei Y., Chen M., 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C, 21(1), 148-162.

Wu W., Liu R., Jin W., 2016. Designing robust schedule coordination scheme for transit networks with safety control margins. Transportation Research Part B, 93, 495-519.

Wu W., Liu R., Jin W., Ma, C., 2019. Simulation-based robust optimization of limited-stop bus service with vehicle overtaking and dynamics: A response surface methodology. Transportation Research Part E, 104, 175-197.

Yu B., Yang Z., Jin P., Wu S., Yao B., 2012. Transit route network design-maximizing direct and transfer demand density. Transportation Research Part C, 22, 58-75.

Yu B., Guo Z., Asian S., Wang H., Chen G., 2019. Flight delay prediction for commercial air transport: A deep learning approach. Transportation Research Part E, 125, 203-221.

Yao B., Hu P., Lu X., Gao J., Zhang M., 2014. Transit network design based on travel time reliability. Transportation Research Part C, 43, 233-248.