Predicting Daily Tourist Flow in Scenic Areas Using LSTM and Big Data from Baidu Index

Feng Li

School of Hotel Management, Qingdao Vocational and Technical College of Hotel Management, Qingdao, Shandong 266100, China

Email: lf_feng@outlook.com

The accurate prediction of tourist flow in scenic areas is crucial for effective tourist area management. This paper introduced the Baidu index associated with search engine data into relevant indicators that can be used to forecast daily tourist flow in scenic spots. The long short-term memory (LSTM) algorithm was employed for daily tourist flow prediction. Simulation experiments were conducted on the Longkou Nanshan Scenic Spot in Longkou City, Shandong Province, China. The experiment first verified the effectiveness of the feature indicators and then compared the predictive performance of the support vector machine (SVM), back-propagation neural network (BPNN), and LSTM models under the situation with or without the Baidu index. It was found that nine feature indicators exhibited significant correlations with daily flow, including date type, week type, month, number of holiday days, average tourist flow, antecedent daily flow, weather condition, standard deviation of daily flow, and Baidu index. The LSTM model demonstrated higher accuracy than SVM and traditional BPNN models. When using the same feature indicators, the p values between the LSTM model and the other two models were both 0.001, indicating significant differences. Furthermore, including the Baidu index as a feature significantly enhanced the accuracy of the prediction algorithm. When comparing within the same prediction algorithm, the p value between the algorithm using the Baidu index feature and the one without the Baidu index feature was less than 0.05, indicating a significant difference.

Povzetek: V raziskavi so za napovedovanje dnevnega turističnega toka uporabili podatke iz Baidu indeksa. LSTM algoritem je dosegel višjo točnost napovedi kot SVM in BPNN modeli, vključitev Baidu pa je bistveno izboljšala natančnost napovednega algoritma.

1 Introduction

Accurate tourist flow prediction is vital for managers of tourist areas to formulate effective resource allocation and operational strategies [1], thereby enhancing tourist satisfaction and safeguarding visitor safety [2]. The conventional approach to forecasting tourist flow is linear prediction. This method employs historical flow data and other quantifiable influencing factors for conducting linear regression analysis and leverages the derived linear pattern to forecast future flow in tourist areas. However, tourist flow in tourist attractions often exhibits complex dynamic characteristics [3], typically demonstrating nonlinear characteristics rather than linear ones, especially in specific periods or regions. As computer performance has advanced, machine learning algorithms, which simulate human brain thinking, mine the nonlinear patterns of tourist flow in tourist areas to realize nonlinear prediction. The relevant studies are reviewed in Table 1.

Table 1: Related studies

Literature	Author	Method	Result
[4]	Guo	The author proposed three types of transit-oriented development-based urban public transportation hubs according to various elements of the BRT transportation station.	The author designed a method and process for predicting hub passenger flow on the basis of studying the current scale and characteristics of hub passenger flow.
[5]	Li et al.	A nearest neighbor nonparametric	The case study demonstrated

		regression method for short-term passenger flow prediction	that the predicted values of the algorithm had a good fit with the measured values.
[6]	Marie-Sainte et al.	The author proposed to explore the firefly algorithm (FA) and particle swarm optimization (PSO) for finding the optimal values of coefficients in linear regression (LR) and used LR to predict air travel demand.	The results indicated that the LR prediction based on PSO performed the best, with a lower error rate compared to the LR based on FA and the LR alone.

2 Daily tourist flow prediction based on big data

The conventional method for predicting daily tourist flow at tourist attractions involves utilizing historical data on tourist flow and other quantifiable factors that can affect tourist flow to conduct regression analysis and using the linear regression equation obtained through fitting to predict daily tourist flow [7]. However, in practical scenarios, the fluctuation of tourist flow often exhibits nonlinear characteristics, and the requirement for a substantial amount of relevant data to ensure accuracy makes the fitting process challenging. As computer performance has advanced, machine learning algorithms have progressively found applications across various domains, including predicting tourist flow in tourist attractions [8]. The parallel processing capabilities of machine learning algorithms enable effective handling of big data, facilitating the extraction of nonlinear patterns.

The variation in daily tourist flow at tourist attractions can be viewed as a type of time series data, where past data influence present data. Consequently, a recurrent neural network (RNN) is employed to analyze time series data. However, RNN encounters challenges such as gradient explosion or gradient disappearance when dealing with long data sequences [9]. To address these issues, gated structures are introduced as memory units within the RNN architecture, forming LSTM.

To improve the accuracy of daily tourist flow prediction using LSTM, in addition to utilizing the flow at past time points as input features, other factors that can influence tourist flow will also be incorporated as input features [10]. This paper incorporated weather information and Internet search data for attractions as feature factors in predicting daily tourist flow. Weather

conditions can influence tourists' travel; typically, better weather conditions make travel easier. The Internet search data for attractions serves a dual purpose: firstly, it reflects the overall attention of tourists towards attractions; secondly, the attraction data provided by search engines can guide tourists in their decision-making process.
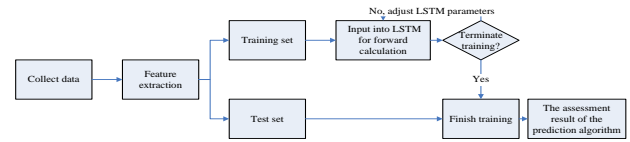

Figure 1: LSTM-based daily tourist flow prediction algorithm.

Table 2: Data features related to the prediction of daily tourist flow in tourist attractions

Feature name	Definition and representation
Date type s_t	A categorical variable, categorized as weekdays, weekends, and holidays, denoted by 0, 1, and 2, respectively
Week type w_t	A categorical variable, categorized into Monday to Sunday, represented by 0 to 6, respectively
Number of holiday days h_t	A numerical variable, depending on the nature of the date
Average tourist flow $\bar{y}_s$	A numerical variable representing the average daily flow for each date type
Month m_t	A categorical variable, categorized into January to December, represented by 1 to 12, respectively
Antecedent daily tourist flow y_{t-k}	A numerical variable, i.e., historical daily tourist flow
Weather condition a_t	A categorical variable, categorized into sunny, cloudy, rainy, snowy, and sandy, denoted by 0, 1, 2, and 3, respectively
The standard deviation of tourist flow v_s	A numerical variable, representing the standard deviation of daily tourist flow for each date type
Baidu index se_t	A numerical variable reflecting the extent to which tourists search for scenic spot-related content online.

The procedure of the LSTM-based daily tourist flow prediction is shown in Figure 1.

① Data related to tourist attractions are collected, and the type of data required is shown in Table 2.

② Feature extraction is performed on the collected data. The corresponding feature types selected through looking up relevant literature and interviews with professionals are shown in Table 1. A day at a tourist attraction is regarded as one sample, and then the feature in the t-th sample x_t is denoted as $\{s_t, w_t, h_t, \bar{y}_s, m_t, y_{t-k}, a_t, v_s, se_t\}$, where $\bar{y}_s$ and v_s depend on s_t and the number of y_{t-k} depends on the number of k required.

③ The training and testing sets were set up.

④ The features of the samples in the training set are input into the LSTM for forward computation [11], which is formulated as:

$$\begin{cases} f_t = \sigma(\omega_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(\omega_i [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(\omega_C [h_{t-1}, x_t] + b_C) \\ C_t = f_t \times C_{t-1} + i \times \tilde{C}_t \\ o_t = \sigma(\omega_o [h_{t-1}, x_t] + b_o) \\ h_t = o_t \times \tanh(C_t) \end{cases}, \quad (1)$$

where $\tilde{C}_t$ and C_t are the temporary state and updated state of the memory unit at the current moment [12], respectively, h_t is the hidden state of the sequence data at the present moment, x_t is the input at the present moment, f_t, i_t and o_t are the outputs of the three gated units, namely, forgetting, input, and output, at the current moment, respectively, ω_f, ω_i, and ω_i are weights in the corresponding gated unit, respectively, b_f, b_i and b_o are biases in the corresponding gated unit, respectively.

⑤ Whether the training is finished is determined. If not, then the parameters in LSTM are adjusted reversely, and then return to step ④; if it does, stop the training and obtain the trained prediction algorithm.

⑥ The performance of the prediction algorithm is tested by feeding the samples in the test set into the trained algorithm.

3 Simulation experiment

3.1 Study area

This study focuses on the Longkou Nanshan Scenic Spot in Longkou City, Shandong Province, China. It is a national 5A-level tourist attraction located in Longkou City, a coastal city on the Jiaodong Peninsula. This expansive scenic area is divided into multiple zones, including the Religious and Cultural Park, Historical and Cultural Park, East Sea Tourism Resort, etc. The Religious and Cultural Park serves as the core of the Nanshan Scenic Spot and consists of various temples. The Historical and Cultural Park was meticulously designed to showcase different dynasties in chronological order. The coastline of the East Sea Tourism Resort stretches for 20 kilometers. It is divided into various areas including the seaside tourist zone, wellness and leisure zone, villa residential zone, commercial and service zone, as well as cultural and educational zone. It is a comprehensive tourism resort that integrates living, tourism, leisure, and humanistic education with high technological content. The resort boasts a pristine ecological environment where people coexist harmoniously with nature.

3.2 Experimental data

The data required for the simulation experiments are outlined in Table 1 in the preceding section. Specifically, the tourist flow data for the Longkou Nanshan Scenic Spot was sourced from the Longkou City People's Government website. The chosen period for this data was from January 2017 to December 2022. Meteorological data was acquired from the National Meteorological Center website (nmc.gov.cn/index.html), spanning the same time frame. Date and month information was annotated according to the calendar for the specified time range, and holidays were designated based on official announcements during the selected period. Baidu index data related to this scenic spot was collected from the Baidu website, employing keywords in the same time period. Samples from January 2017 to December 2020 constituted the training set, while samples from January 2021 to December 2022 formed the test set. The values missing in the collected data were supplemented using the interpolation method.

When using the Baidu website to obtain the Baidu index related to scenic spots [13], the selection of keywords is essential. The processing steps in this study are as follows.

① A keyword thesaurus was established. "Longkou Nanshan Scenic Spot" and the name of its neighboring scenic spots were used as the root keywords, and then the beginning and end of these root keywords were expanded. The expansion of the beginning of the word means adding the name of the superior place, and the expansion of the end of the word means adding the vocabulary related to tourism elements. Taking "Longkou Nanshan Scenic Spot" as an example, the word beginning can be expanded to "Longkou City Longkou Nanshan Scenic Spot", and the word end can be expanded to "Longkou Nanshan Scenic Spot special food". ② The Baidu index of the keyword in the Baidu website, i.e., the overall trend of the search volume of the keyword, was crawled using a crawler program. The Baidu index was arranged in the order of the day. ③ The keywords with too low search volume and their corresponding Baidu index were deleted.

3.3 Experimental setup

The parameters of the LSTM algorithm obtained through the orthogonal experiment for predicting daily tourist flow in scenic spots are presented in Table 3. Additionally, experiments were also conducted on both support vector machine (SVM) and traditional back-propagation neural

network (BPNN) algorithms to assess the LSTM algorithm's performance. The relevant parameters for the SVM algorithm are as follows: the kernel function utilized the sigmoid function [14], and the penalty parameter was set to 1. As for the traditional BPNN algorithm, the parameters were configured as follows: the number of nodes in the input and output layers aligned with the LSTM algorithm, the hidden layer was set to one layer with 128 nodes, the activation function adopted sigmoid, and the learning rate and the number of training iterations were consistent with those of the LSTM algorithm.

Table 3: Parameter settings of LSTM.

Parameter	Setting	Parameter	Setting
Input layer	Nine nodes	Hidden layer	Three layers with 64 nodes per layer
Hidden layer activation function	Sigmoid	Output layer	One node
Learning rate	0.02	Maximum number of training sessions	500
Batch size	100	Optimizer	Stochastic gradient descent

In addition to the aforementioned experiments, robustness tests were also conducted on the LSTM algorithm. In these tests, data samples from Longkou People's Park and Huangshui River Wetland Park in Longkou City were collected from January 2021 to December 2022 using the method described earlier. Then, the performance of the LSTM algorithm in other datasets was evaluated by applying it to predict data based on these data samples.

3.4 Evaluation indicators

For the performance of the three prediction algorithms tested in this paper, the root-mean-square error (RMSE) and mean absolute percentage error (MAPE) [15] was used to measure the performance, which is formulated as:

$$\begin{cases} RMSE = \sqrt{\dfrac{\sum_i (y_i - \hat{y}_i)^2}{N}} \\ MAPE = \dfrac{1}{N} \sum_{i=1}^{N} \dfrac{|y_i - \hat{y}_i|}{y_i} \times 100\% \end{cases}, (2)$$

where y_i is the actual daily tourist flow, $\hat{y}_i$ is the predicted daily tourist flow, and N is the total number of test samples.

After computing the RMSE of the three prediction algorithms based on the sample data with and without the Baidu index feature, a t-test was employed to assess the influence of the Baidu index feature on algorithm performance and compare the performance of the three

prediction algorithms. A t-test p value below 0.05 indicates a significant difference, while a value below 0.01 signifies an extremely significant difference.

3.5 Experimental results

Before utilizing the samples to train each machine learning algorithm, this paper analyzed the correlation between the selected feature indicators and daily tourist flow to ensure that these indicators can affect the flow. Table 4 shows that all p values were less than 0.01, signifying a significant correlation between the nine feature indicators and daily tourist flow. The result verified the feasibility of using these nine feature indicators for predicting daily flow.

Table 4: Results of the correlation analysis between the sample feature indicators and daily tourist flow.

Feature indicator	Correlation coefficient	t	p
Date type s_t	1.11	1.36	0.001
Week type w_t	0.98	1.11	0.002
Number of holiday days h_t	1.35	2.14	0.001
Average tourist flow $\bar{y}_s$	0.87	0.58	0.001
Month m_t	0.52	0.87	0.002
Antecedent daily tourist flow y_{t-k}	1.37	1.14	0.001
Weather condition a_t	1.25	0.87	0.001
The standard deviation of tourist flow v_s	1.12	0.99	0.002
Baidu index se_t	2.47	1.58	0.000

Due to space limitations, only the prediction results of the three prediction algorithms for April 2021 are presented in Figure 2. The daily tourist flow at Longkou Nanshan Scenic Spot fluctuated between 10,000 person-time and 30,000 person-time during April 2021. Upon comparing the prediction outcomes of the three algorithms, it was found that the results of the SVM and traditional BPNN algorithms exhibited a noticeable deviation, with the SVM algorithm displaying a relatively more significant deviation. In contrast, the results obtained by the LSTM algorithm generally aligned with the actual daily flow, with only occasional deviations on some days. Moreover, these deviations were smaller compared to the other two algorithms.
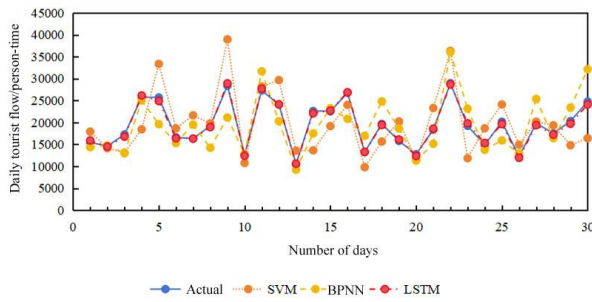
Figure 2: Results of the three forecasting algorithms for daily tourist flow in April 2021.

Comparing the mean square error of the three algorithms with or without the Baidu index feature and assessing the differences' significance, the results are summarized in Table 5, with specific values omitted here. Table 5 shows that when predicting daily tourist flow using the same algorithm, the RMSE and MAPE in the case without the Baidu index feature was significantly higher than that with the Baidu index feature, especially the LSTM algorithm. When comparing different prediction algorithms under the same feature indicators, it is observed that the RMSE and MAPE of the three prediction algorithms differed significantly. Regardless of the presence of the Baidu index feature, the SVM algorithm exhibited the highest RMSE and MAPE, followed by the traditional BPNN algorithm, while the LSTM algorithm had the lowest.

Table 5: The RMSE and MAPE of the three prediction algorithms with and without the Baidu index feature and the significant degree of differences.

Algorithm	Without the Baidu index feature		With the Baidu index feature	
Indicator	RMSE	MAPE	RMSE	MAPE
SVM	35837.65	22.89	29638.56*	20.13*
BPNN	26789.66	17.54	21547.74*	15.22*
LSTM	15787.58	12.11	12589.87*	10.14*
P value between SVM and BPNN	0.043	0.041	0.048	0.045
P value between SVM and LSTM	0.001	0.000	0.001	0.000
P value between BPNN and LSTM	0.001	0.001	0.001	0.001

Note: * indicates a significant difference within the same kind of algorithm with or without the Baidu index feature, i.e., the p value is less than 0.05.

The robustness test results of the LSTM algorithm are shown in Table 6. It can be seen that there was no significant difference in the errors of daily tourist flow forecasts for different scenic areas using the LSTM algorithm. In other words, the LSTM algorithm algorithm can be applied to predict daily tourist flow for other scenic areas.

Table 6: The robustness test results of the LSTM algorithm.

	RMSE	MAPE
Longkou Nanshan Scenic Spot	12589.87	10.14
Longkou People's Park	12578.57	10.21
Huangshui River Wetland Park	12547.98	10.24
The p value between the comparison of the Longkou Nanshan Scenic Spot and Longkou People's Park	0.125	0.214
The p value between the Longkou Nanshan Scenic Spot and Huangshui River Wetland Park	0.215	0.187
The p value between the Longkou People's Park amd Huangshui River Wetland Park	0.236	0.158

4 Discussion

This article used the LSTM algorithm, a deep learning algorithm, to forecast the daily tourist flow. In order to further enhance the accuracy of the prediction, the Baidu index feature was also introduced. Then, simulation experiments were conducted on the proposed algorithm and compared with the SVM and BPNN algorithms. Firstly, a correlation analysis was performed on nine indicators related to daily tourist flow, and the analysis results showed that all the indicators were significantly related to daily tourist flow. The date type includes working days, weekends, and holidays. Typically, tourists have more time to travel on weekends or holidays, resulting in an increase in visitor flow at scenic spots. Therefore, there is a correlation between date type and visitor flow. The weekday type refers to a specific day of the week, similar to the date type. Generally, tourists are more likely to travel on Saturdays and Sundays, so there is a correlation between weekday type and tourist flow. The holiday duration refers to the length of vacation time. Usually, longer vacations tend to promote tourist travel and increase tourist flow at scenic spots; hence there is a correlation between holiday duration and visitor flow. The average tourist flow represents the average number of visitors at different time periods and directly reflects changes in visitor flow at scenic spots; therefore it has a correlation with tourist flow. "Month" refers to the month to which a date belongs, and for scenic spots, there are

peak seasons and off-peak seasons. The number of tourists is higher during the peak season, so there is a correlation between the month and visitor flow. The antecedent daily tourist flow refers to the number of visitors in a certain period before the current day, which also correlates with tourist flow prediction. The weather condition refers to the weather on that day; better weather conditions can promote tourist activities. The standard deviation of tourist flow reflects the fluctuation in tourist numbers during a specific time period. Baidu index, an indicator reflecting online user behavior data, was used in this article to reflect tourists' searches and evaluations for scenic spots. The more searches and evaluations there are, the more attractive the scenic spot is to tourists.

In comparison with the other algorithms, the results also showed that the introduction of Baidu index as a feature improved the accuracy of the prediction algorithm. Additionally, the LSTM algorithm used in this study demonstrated higher predictive accuracy compared to the other two algorithms. The reason for this is that Baidu Index reflects internet users' behavior data in information retrieval, which can reflect tourists' interest in scenic spots on the Internet and further supplement the features of tourist behavior data, thereby enhancing the accuracy of predictions. Furthermore, compared to the SVM and BPNN algorithms, the LSTM algorithm not only uses the activation function in hidden layers to fit non-linear patterns but also utilizes the gate mechanism to incorporate historical information, resulting in higher predictive accuracy.

The novelty of this article lies in the use of LSTM algorithm to forecast daily tourist flow with time series characteristics. To enhance the precision of prediction, Baidu index was introduced as an effective reference for accurately predicting daily tourist flow in scenic areas. The limitation of this article lies in the insufficient span of the algorithm's prediction for daily tourist flow, or rather, its instability in predicting large-span daily tourist flow. Although robustness tests have been conducted, the dataset lacks diversity in terms of types. Therefore, future research directions include further improving the accuracy and generalization ability of the algorithm.

5   Conclusion

This paper provides a concise overview of relevant indicators for predicting daily tourist flow at tourist attractions. The Baidu index data from search engines was introduced into these indicators. The LSTM algorithm was employed for prediction. Moreover, simulation experiments were conducted to compare the LSTM algorithm with the traditional BPNN and SVM algorithms using the Longkou Nanshan Scenic Spot, spanning January 2017 to December 2022. Key findings are as follows. Nine feature indicators, including date type, week type, month, number of holiday days, average tourist flow, antecedent daily tourist flow, weather condition, the standard deviation of daily tourist flow, and Baidu index, exhibited significant correlations with daily tourist flow. The SVM and traditional BPNN algorithms displayed noticeable deviations in daily tourist flow prediction, with

the former showing a more significant deviation; in contrast, the prediction results of the LSTM algorithm generally aligned with the actual daily tourist flow, exhibiting only occasional deviations. When the identical prediction algorithm was used to predict daily tourist flow, the RMSE and MAPE were significantly higher for feature indicators without the Baidu index feature than those with the Baidu index feature. Additionally, among the three prediction algorithms under the same feature indicators, the SVM algorithm had the highest RMSE and MAPE, followed by the traditional BPNN algorithm, and the LSTM algorithm had the lowest.

The article used the LSTM algorithm to forecast the daily tourist flow with time series characteristics. In order to enhance the accuracy of prediction, it also introduced the Baidu index feature, which provides an effective reference for accurately predicting daily tourist flow in scenic areas.

6   References

[1]   Zhang Z, Wang C, Gao Y, Chen Y, Chen J (2020). Passenger Flow Forecast of Rail Station Based on Multi-Source Data and Long Short Term Memory Network. *IEEE Access*, 8, pp. 28475-28483. https://doi.org/10.1109/ACCESS.2020.2971771

[2]   Guo Z, Zhao X, Chen Y, Wu W, Yang J (2019). Short-term passenger flow forecast of urban rail transit based on GPR and KRR. *IET Intelligent Transport Systems*, 13(9), pp. 1374-1382. https://doi.org/10.1049/iet-its.2018.5530

[3]   Liu W, Tan Q, Wu W (2020). Forecast and Early Warning of Regional Bus Passenger Flow Based on Machine Learning. *Mathematical Problems in Engineering*, 2020(1), pp. 1-11. https://doi.org/10.1155/2020/6625435

[4]   Guo Z, Wu HY, Yang CH (2015). Study on Method of BRT Transportation Station Passenger Flow Forecast with the TOD Mode. *Applied Mechanics & Materials*, 744-746, pp. 2059-2062. https://doi.org/10.4028/www.scientific.net/AMM.744-746.2059

[5]   Li Y, Ma C (2023). Short-Time Bus Route Passenger Flow Prediction Based on a Secondary Decomposition Integration Method. *Journal of Transportation Engineering, Part A. Systems*, 149(2), pp. 4022132.1-4022132.10.

[6]   Marie-Sainte SL, Saba T, Alotaibi S (2019). Air Passenger Demand Forecasting Using Particle Swarm Optimization and Firefly Algorithm. *Journal of Computational and Theoretical Nanoscience*, 16(9), pp. 3735-3743. https://doi.org/10.1166/jctn.2019.8242

[7]   Lu T, Yao E, Liu S, Zhou W (2020). Short-time Forecast of Entrance and Exit Passenger Flow for New Line of Urban Rail Transit During Growth Period. *Tiedao Xuebao/Journal of the China Railway Society*, 42(5), pp. 19-28. https://doi.org/10.1049/iet-its.2018.553010.3969/j.issn.1001-8360.2020.05.003

[8]   Yao E, Zhou W, Zhang Y (2018). Real-Time Forecast of Entrance and Exit Passenger Flow for

Newly Opened Station of Urban Rail Transit at Initial Stage. *Zhongguo Tiedao Kexue/China Railway Science*, 39(2), pp. 119-127. https://doi.org/10.3969/j.issn.1001-4632.2018.02.15

[9] Wang Y, Ma J, Zhang J (2020). Metro Passenger Flow Forecast with a Novel Markov-Grey Model. *Periodica Polytechnica Transportation Engineering*, 48(1), pp. 70-75. https://doi.org/10.3311/PPtr.11131

[10] Wen K, Zhao G, He B, Ma J, Zhang H (2022). A decomposition-based forecasting method with transfer learning for railway short-term passenger flow in holidays. *Expert Systems with Applications*, 189, pp. 1-16. https://doi.org/10.1016/j.eswa.2021.116102

[11] Ni M, He Q, Gao J (2017). Forecasting the Subway Passenger Flow Under Event Occurrences With Social Media. *IEEE Transactions on Intelligent Transportation Systems*, 18(6), pp. 1623-1632. https://doi.org/10.1109/TITS.2016.2611644

[12] Zhang J, Shen D, Tu L, Zhang F, Xu C, Wang Y, Tian C, Li X, Huang B, Li Z (2017). A Real-Time Passenger Flow Estimation and Prediction Method for Urban Bus Transit Systems. *IEEE Transactions on Intelligent Transportation Systems*, 18(11), pp. 3168-3178. https://doi.org/10.1109/TITS.2017.2686877

[13] He Y, Li L, Zhu X, Tsui KL (2022). Multi-Graph Convolutional-Recurrent Neural Network (MGC-RNN) for Short-Term Forecasting of Transit Passenger Flow. *IEEE Transactions on Intelligent Transportation Systems*, 23(10), pp. 18155-18174.

[14] Sajanraj TD, Mulerikkal J, Raghavendra S, Vinith R, Fabera V (2021). Passenger flow prediction from AFC data using station memorizing LSTM for metro rail systems. *Czech Technical University in Prague - Central Library*, 31(3), pp. 173-189. https://doi.org/10.14311/NNW.2021.31.009

[15] Sha S, Li J, Zhang K, Yang Z, Wei Z, Li X, Zhu X (2020). RNN-based subway passenger flow rolling prediction. *IEEE Access*, 8, pp. 15232-15240. https://doi.org10.1109/ACCESS.2020.2964680