applied sciences

MDPI

Article

Metro Emergency Passenger Flow Prediction on Transfer Learning and LSTM Model

Jingye Ma [1], Xin Zeng [1,*], Xiaoping Xue [1] and Ranran Deng [2]

[1] Department of Information and Communication Engineering, Tongji University, Shanghai 200070, China; 2033081@tongji.edu.cn (J.M.); xuexp@tongji.edu.cn (X.X.)
[2] Technical Center of Shanghai Shentong Metro Group Co., Ltd., Shanghai 200070, China; deng2ran@163.com
* Correspondence: zengxin1@tongji.edu.cn

Abstract: The metro transportation system will have emergency passenger flow for various reasons, resulting in passenger flow congestion, affecting efficiency and risks. In this paper, the LSTM network is applied to predict the normal passenger flow and emergency passenger flow of metro transportation based on transfer learning to solve the imbalanced data set problem when the amount of emergency samples is too small. The results show that under normal and emergency conditions, the average prediction error is less than 5%, which provides an alarm for the operating company to take preventive measures in advance. Compared with the strategy without transfer learning, it proves that the strategy proposed in this paper has advantages in predicting emergency conditions.

Keywords: LSTM; transfer learning; passage flow prediction; metro station

Citation: Ma, J.; Zeng, X.; Xue, X.; Deng, R. Metro Emergency Passenger Flow Prediction on Transfer Learning and LSTM Model. *Appl. Sci.* **2022**, *12*, 1644. https://doi.org/10.3390/app12031644

Academic Editor: Alexandre Carvalho

Received: 28 December 2021
Accepted: 27 January 2022
Published: 4 February 2022

1. Introduction

Compared with private car travel, public transport has greater traffic volume and communication efficiency, lower energy consumption and air pollution. As an effective measure to alleviate the pressure of urban traffic, the urban metro transportation system is widely used in the world. However, it is common for passenger flow congestion to occur during the peak period of transportation. Therefore, it is necessary to predict the passenger flow of metro transportation.

1.1. Passenger Flow Forecast

Passenger flow forecast is divided into two categories. One is analyzing the time sequence characteristics according to the existing data and then predicting the future passenger flow. In this way, a short-term passenger flow forecast can be obtained. Traditional methods can realize the analysis of time series characteristics, such as regression model, Bayesian network, Markov method, etc.

In ref. [1], Sun et al. present a traffic flow prediction method based on a Bayesian network, and the joint probability distribution between the cause node (data for prediction) and the effect node (data to be predicted) is described as a Gaussian mixture model (GMM). Cai et al. [2] use a hybrid regression model, considering both local and global information, to solve the limited training sample size and achieve short-term passenger flow forecasting.

With the development of machine learning and neural network technology, machine learning technology has also been applied to metro passenger flow forecasting. Sun et al. [3] use SVM learning combined with wavelet analysis technology to divide passenger flow data into high-frequency and low-frequency sequences for learning. Li et al. [4] use a dynamic radial basis function (RBF) neural network to forecast passenger flow and analyze its performance in different periods. LSTM network is often used in time series prediction. Many researchers combine it with different methods to predict subway passenger flow and have achieved good results [5–7].

In the operation process of metro transportation, the best design is generally considered to make the passenger flow close to the maximum transport capacity of the metro transportation network, to maximize the operating efficiency. When the external environment is affected by some unpredictable factors, the potential transport passenger demand increases, or the transport capacity of metro transportation decreases. It will lead to passenger congestion and potential safety problems, which is called emergency passenger flow. If we can have an accurate forecast of emergency passenger flow, we can take targeted measures to alleviate the pressure of the metro transportation system.

1.2. Emergency Passenger Flow

Li et al. [8] established the mathematical prediction model of emergency passenger flow by analyzing the passenger flow data within 35 days. Gao et al. [9] analyze the change of passenger flow and give the guidance method of passenger flow evacuation according to the special situation of subway service interruption. However, with this analysis method using mathematical modeling, it is difficult to take full use of the huge amount of passenger flow data obtained from the subway.

Compared with normal passenger flow, the characteristics of emergency passenger flow lie in its uncertainty and contingency. The causes of emergency passenger flow are often different, and the impact of different reasons on passenger flow fluctuation is also different. Emergency passenger flow is often rare. Compared with normal passenger flow, the sample size is much smaller, and the distribution is more discrete, so it is difficult to summarize the regularities.

Based on the above problems, the traditional learning method is not suitable for emergency passenger flow forecasting, and the data enhancement strategy usually used for small sample learning is not suitable because the sample spaces of emergency cases are unknown.

1.3. Transfer Learning

Transfer learning is to apply the prior knowledge obtained in primary fields to other target fields. Pan [10] published a survey paper on transfer learning in 2010, summarized the history of the existing transfer learning, and defined the classification of different transfer learning methods. In fact, transfer learning can not only be used in machine learning but also has outstanding performance in data mining and other fields. Weiss, K. et al. summarized the application of transfer learning in data-mining-related tasks such as classification, regression and clustering, and further standardized the definition of terms related to transfer learning in different fields.

Generally, the network comprising feature extraction layers and classifier layers has been trained in the primary field. Then the feature extraction part is reused and fixed in the target field training where only the classifier will be trained based on the samples in target fields.

In this way, it can solve the problem of insufficient sample size in the target field. It has been proved to be one of the effective methods to deal with small sample learning. It has been widely used in natural language processing [11–13], computer vision field [14,15].

The biggest advantage of transfer learning is that it can be further trained based on existing high-performance models to adapt to different situations. This usage is widely used in the field of natural language processing. BERT [16] is one of the most advanced natural language processing models. Based on it, researchers have produced a large number of research results [17–19] using transfer learning. Based on the original Bert model, they retrained the model with data sets in different target fields. In the case of retaining part of the original network structure, some parameters are added or deleted to optimize the effect in the corresponding target field.

Based on this feature, transfer learning can solve the problem of small sample learning, especially fault-related problems. Transfer learning can extract features from massive source data and apply them to similar but smaller fault data. Xiao et al. [20] use the CNN network based on transfer learning to solve the problem of the small amount of mechanical fault data in industrial applications, and uses motor fault data to verify the accuracy of its fault analysis model.

This paper takes Shanghai Metro as an example and uses the transfer learning method based on LSTM to establish a prediction model for metro transportation emergency passenger flow. The problem of small samples in the training process is solved by using transfer learning. The results show that the model has good accuracy under normal and emergency conditions.

The rest of this paper is organized as follows. Section 2 introduces the relevant knowledge and work; Section 3 introduces the data set and the method used in this paper; Section 4 shows the experimental results. Finally, Section 5 summarizes and discusses the future research.

2. Materials and Methods

2.1. Time Series Analysis

Considering that the focus of this paper is the decline of metro transportation capacity, this paper uses the section passenger flow in each time period of the metro as the analysis data set. The so-called section passenger flow is defined as the number of people who are transported from one station to the adjacent station by the metro every hour. It is assumed that there are $n + 1$ stations in a metro line, which is represented as Sat_i. Then the whole line can be divided into n sections, and the passenger flow per unit time of the i section can be expressed by k_i, that is $k_i = [Sat_{i-1} \rightarrow Sat_i]$.

The section passenger flow has directivity to reflect the current metro transportation capacity in a certain section for either uplink or downlink direction. The characteristics of uplink and downlink passenger flow are usually different, so it needs to be considered separately. The passenger flow of the section in a certain period of the whole station interval can be expressed as $w_t = \left(k_1^t, \cdots, k_n^t\right)$ (t is the time period, n is the total number of sections).

2.2. Factors of Emergency Passenger Flow

Metro faults are one of the main causes of emergency passenger flow. On 25 April 2018, Shanghai Metro Line 2 was delayed for more than 2 h from 6 a.m. due to signal system fault, resulting in a large number of passengers being stranded and having to choose other ways of transportation, which also increased the burden of other ground public transportation and brought potential risks.

As shown in Figure 1, the passenger flow in the fault period of the day is compared with that in the same period of a week. It can be seen that the transportation passenger flow has decreased significantly.

According to the different metro faults, we can use the severity of the fault to measure its impact on the metro operation. Fault severity can be measured from two dimensions of time and space. The time dimension of fault refers to the duration of fault, including the duration of the fault itself, the time needed to alleviate or repair the fault, and the time needed to clear the impact caused by the fault; The spatial dimension of fault refers to the location of the fault, including the type of fault, the equipment with the fault, the operation section affected by the fault, etc.

The metro fault situation is complex and changeable, and the situation of different lines is also different, so it is difficult to use a unified algorithm to describe it. In addition, we do not need accurate quantitative descriptions. Therefore, this paper uses the expert evaluation method to quantify the fault severity qualitatively. The advantage is that the relevant personnel dealing with the line fault are familiar with the situation of the corresponding metro line, and have sufficient experience and knowledge of metro faults, which can more accurately measure the severity of the metro fault.

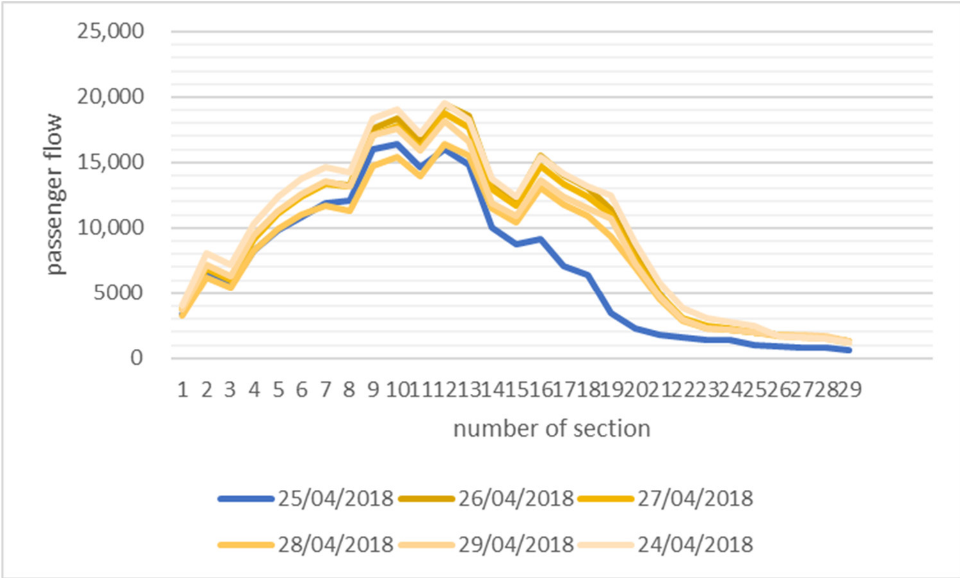

Figure 1. Impact of an emergency event on passenger flow.

3. Prediction Model for Emergency Event

The purpose of this model is to realize passenger flow prediction under normal and emergency conditions. Therefore, this paper is divided into two parts. The first part is to complete the normal passenger flow forecast of metro transportation, and the second part is to forecast the emergency passenger flow based on the normal passenger flow forecast, as shown in Figure 2.

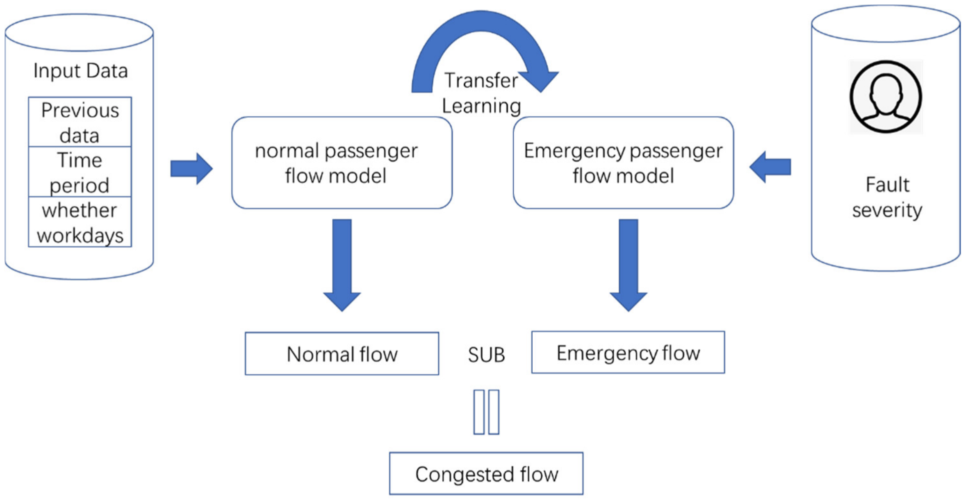

Figure 2. The proposed transfer learning procedure.

3.1. Calculation Process

Under normal conditions, according to the surrounding environment, line distribution, and other factors of the metro, there will be a stationary passenger flow demand R_0 per unit time, that is, $R_0 = f(X)$. X is the set of environmental factors, which is only related to the physical space environment of the metro and has nothing to do with the operation behavior.

Similarly, according to the Metro operator's strategy, such as operation speed, departure interval, number of passengers in the carriage, there will be a stationary maximum transport capacity Cap_{max} per unit time, that is, $Cap_{max} = g(Y)$. Y is the metro operation strategy set, and Cap_{max} can change with the operator's strategy. Obviously, when

Cap_{max} approaches to R_0 and is slightly greater than R_0, the metro operation efficiency is the highest. Here, we might as well assume that under normal conditions, the transport capacity of the metro can fully meet the demand of transport passenger flow. At this time, the normal transport passenger flow is the demand of normal transport passenger flow and just reaches the limit of transport capacity, that is, $R_0 \approx Cap_{max}$.

As shown in Figure 3, when the metro breaks down suddenly, the transportation capacity of the metro will decrease, but the passenger flow demand will not be affected. Currently, R_0 remains unchanged, $Cap'_{max} = g(Y') < Cap_{max}$. The difference N between the two is the passenger flow stranded in the station due to the emergency, that is, $N = Cap_{max} - Cap'_{max}$.

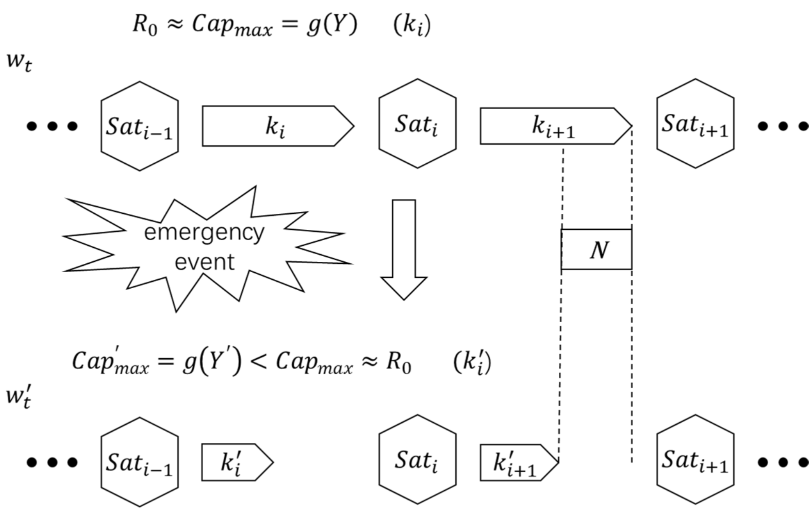

Figure 3. Process for calculating congested passenger flow.

To simplify the model and improve the universality of the model, we use the LSTM network to forecast the normal passenger flow, taking into account the working day and time period; For the emergency passenger flow, we use the method of expert evaluation to quantify the severity of the fault qualitatively. Taking the severity of the fault as a reference factor, we use the method of transfer learning to predict the emergency passenger flow based on the normal passenger flow prediction network. The input parameters are shown in Table 1.

Table 1. Input parameter list.

Input	Symbol	Form
passenger flow	w_t	$3 \times n$ matrix
time period	t	0~24
whether workdays	wd	0 (no) or 1 (yes)
fault level	f_{lv}	0~9, 0 means no fault.

3.2. Loss Function

It can be found that even in the peak period, for the departure and terminal stations in remote areas of the city, the passenger flow will not be very high. If the traditional percentage error function is used, even if there is a small error at these stations, it will have a great impact on the whole network. In addition, we are not very concerned about the situation of small passenger flow, so in order to reduce the impact of small passenger flow on the results, we use the rewritten weighted percentage error function.

The traditional mean percentage error function ($MAPE$) is as follows:

$$MAPE = \frac{1}{n}\Sigma|y' - y|/y, \tag{1}$$

It can be seen that when y is small or even 0, $MAPE$ cannot produce appropriate results. Therefore, we use the improved weighted mean percentage error function ($WMAPE$).

$$WMAPE = \Sigma|y' - y|/\Sigma y, \tag{2}$$

Table 2 shows an example to compare the advantages and disadvantages of the two algorithms, if the passenger flow and predicted value of each station in a certain period are shown in the table below, the corresponding $MAPE$ and $WMAPE$ are calculated according to the data. We can see that when the passenger flow of the station itself is large, such as station 5 and station 6, the impact of the fluctuation of the forecast passenger flow deviation on the two algorithms is basically the same. When the passenger flow of the station itself is small, such as station 2, the predicted number of people is only one person different, but the result of the $MAPE$ algorithm has a huge change. In addition, for unmanned sites, such as site 1, the $MAPE$ algorithm cannot give results.

Table 2. Comparison of different error calculation methods in a hypothetical case.

Station	**1**	**2**	**3**	**4**	**5**	**6**
real flow	0	5	50	500	5000	50,000
predict flow	1	4	49	498	4998	49,998
error	1	1	1	2	2	2
MPE	err	0.2	0.02	0.004	0.0004	0.00004
WMPE	0.00002	0.00002	0.00002	0.00004	0.00004	0.00004
MAPE						0.03741
WMAPE						0.00016

3.3. LSTM Network Structure

In this paper, the input of the network is the time of the passenger flow to be predicted (time), the workday (workday), the fault severity (LV), and the section passenger flow in the first few time periods of all stations. The output result is the passenger flow of the whole station section, that is, the transportation capacity of the whole station.

The input and output of the traditional machine learning model based on time series are usually time-series data. However, in this paper, the time period, working day and fault severity data are the properties of the passenger flow, and they are not time series. Therefore, they cannot be used as the input of the LSTM network, and need to add a dense layer. A multi-input and single output neural network is constructed, in which the LSTM part is responsible for extracting the time-series features of section passenger flow, and the dense layer part is responsible for mixed judgment of time period, working day and fault severity. The whole network is shown in Figure 4.

The first half of the network is a two-layer LSTM network. The sliding window algorithm is adopted. Taking the passenger flow of each station section in the first three hours of the predicted time period as the input data, the input is $w_{in1} = (w_{i-3}, w_{i-2}, w_{i-1})$. Finally, after the integration of the whole connection layer of the first layer, the output of wout1 is achieved.

Assuming that no other non-time-series input is considered, the above w_{out1} can output the final prediction result w_{out}. Due to the addition of other input variables, LSTM does not output the result but outputs 256-dimensional time series eigenvalues. After merging with another time period, working day and fault severity three input data w_{in2} through a merging layer, it is input into a network composed of two dense layers, the final output of the current period of the whole station passenger flow forecast, as shown in Figure 5.

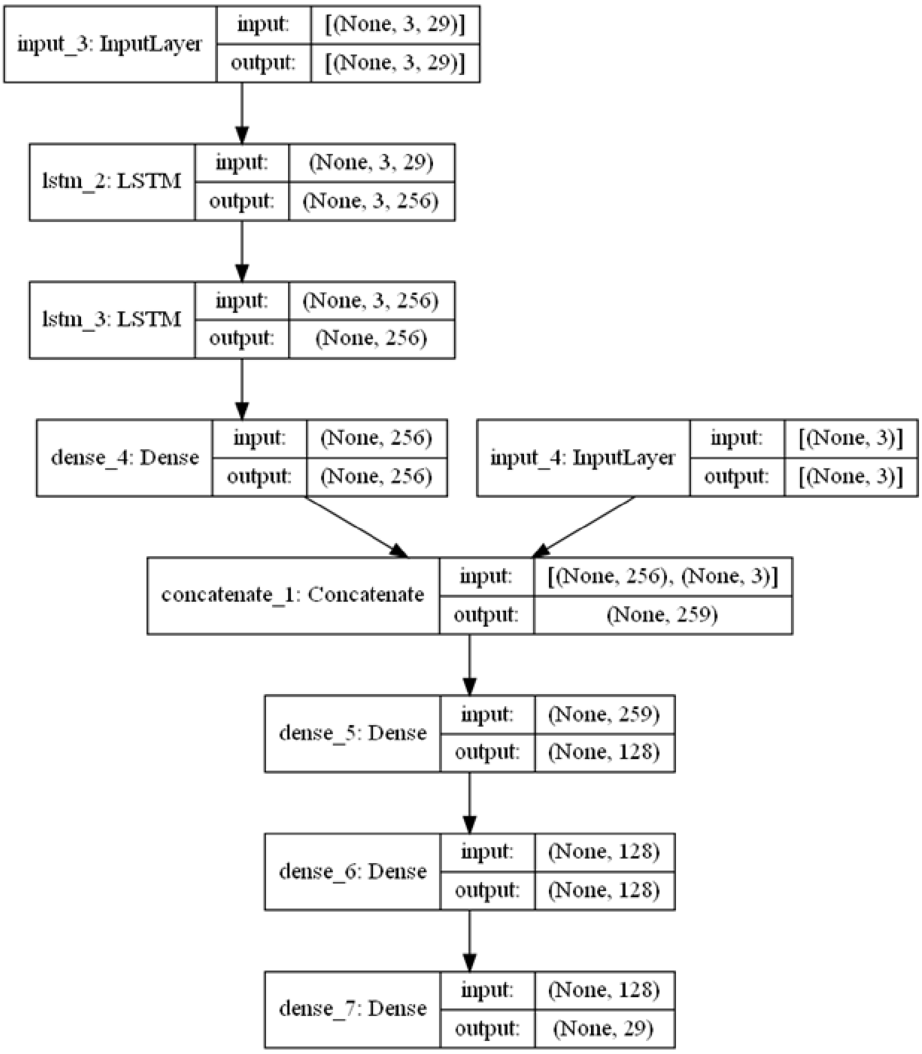

Figure 4. The modified LSTM network structure for transfer learning.

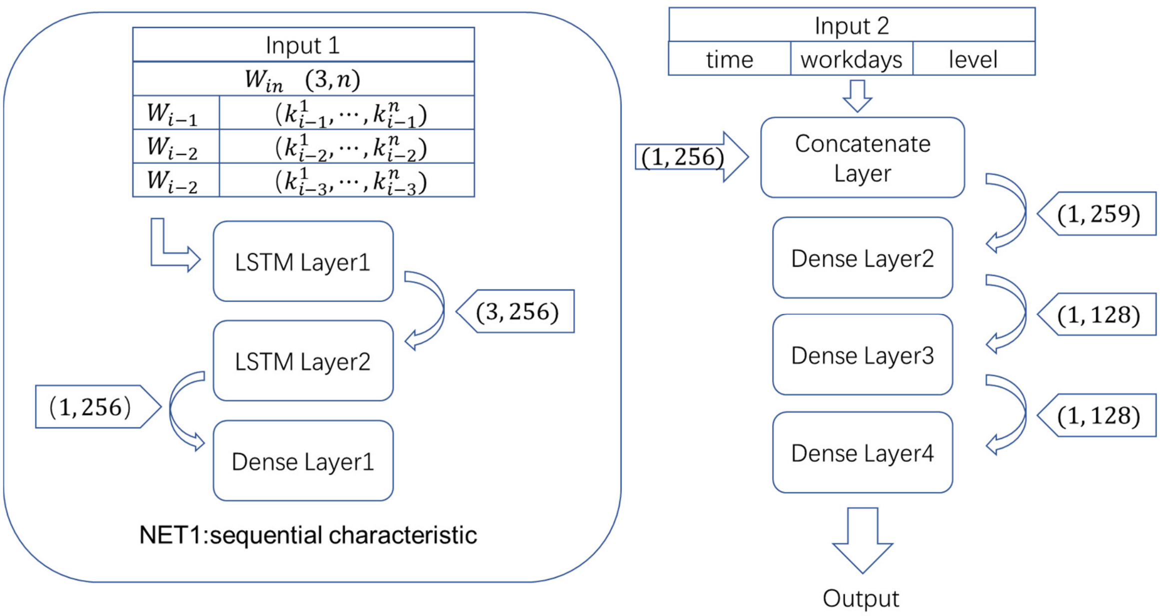

Figure 5. The training procedure for proposed LSTM network.

3.4. Model Training Procedure

Based on the above work, the network is trained with data without fault conditions. At this time, the input of fault severity is disabled by setting to zero to get the first network K. $K = K_1 + K_2$, where K_1 is the weight matrix of the LSTM network part and K_2 is the weight matrix of the dense layer network. The prediction results can be calculated by the following formula:

$$y = f(K_2[g(K_1 w_{in1}); w_{in2}], \tag{3}$$

where [:] represents merging operation, g represents LSTM partial operation, and f represents full connection layer partial operation.

Let us assume that the network $K' = K_1' + K_2'$ for the prediction of passenger flow in emergencies.

For the passenger flow forecast in the period of fault occurrence, since the fault will not affect the passenger flow in the period before the fault occurs, the occurrence of emergency will not affect the characteristics of the first half of the network, that is, $K_1' = K_1$. Therefore, for K', we can fix the weight matrix of the first half of the network, and only change the weight of the second half of the network through training, That is $K' = K_1 + K_2'$. Then the forecast passenger flow can be expressed as

$$y' = f(K_2'[g(K_1 w_{in1}); w_{in2}], \tag{4}$$

4. Simulation and Results

The data set adopted is the passenger flow data of Shanghai Metro Line 2 from 2017 to 2019. The operation time of line 2 is from 5:28 to 23:30.

4.1. Data Pretreatment

Considering the data integrity, the two time periods of 5:00–6:00 and 23:00–24:00 are excluded when selecting the data set. Only 17 time periods between 6:00–23:00 are selected as the data set, and the passenger flow in the non-operation time period of 0:00–5:00 is zero. There are 29 sections of data in each time period. Finally, about 27,000 normal passenger flow data and 20 available emergency passenger flow data are obtained.

For data with particularly large scales such as people flow, the maximum-minimum normalization method is used to normalize the data.

$$x_{nor} = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{5}$$

x_{min} is the minimum value of passenger flow, which is 0 in this paper, x_{max} is the maximum value of passenger flow.

4.2. Simulation Setup

Because the neural network based on LSTM is one of the common schemes for passenger flow prediction, we take the LSTM network without migration learning as the control group. In this experiment, we compare the results of using transfer learning and not using transfer learning to prove the effectiveness of the experimental design.

For this model, we build it through the following steps:

1. All data sets were randomly divided into a training set and test set, accounting for 70% and 30% respectively;
2. The normal passenger flow training set data are used as input to train the network, and the normal passenger flow test set is used to verify the network performance to obtain the normal passenger flow prediction network K;
3. Freeze the LSTM part of network K, use the training set of emergency passenger flow, train the network again based on network K, and then use the test set of emergency passenger flow to verify the network performance, and get the prediction network K'.

The trigger uses Selu and Adam optimizer. Considering that there are 17 samples in a whole day, the learning step is set to 34, and other parameters are adjusted according to experience and learning effect. The above-mentioned weighted average percentage error function was used for model evaluation.

4.3. Results and Analysis

4.3.1. Normal Passenger Flow

For the normal passenger flow forecast results, although the $WMAPE$ algorithm used in the model can well measure the forecast error, we cannot see the intuitive performance of the model from it. Therefore, after the training, we use the mean percentage error function to evaluate the performance of the model site by site.

Consider the following questions:

1. The purpose of this paper is to predict in advance the passenger flow congestion problem that may occur when there is an emergency passenger flow;
2. The samples with small passenger flow are often in the starting station or the early morning when the subway is running. For these samples, most of the input passenger flow is 0, so accurate results cannot be obtained.

Based on the above two points, we investigate the performance of the model in different situations, and the results are shown in Table 3.

Table 3. Prediction effect under different passenger flow ranges.

	Sample Size	Error Rate
>500	585,289	12.95%
<1000	44,254	147.82
1000~5000	230,557	19.75%
>5000	325,854	4.91%

For all the samples with passenger flows of more than 500, there are 585,289 samples, accounting for 97.31% of the total, and the average error is 12.95%.

Although for the sample with less than 1000 passengers (7.36% of the total sample), the average error of the model is larger, reaching 147.83%, which is nearly double the error, but for the sample with more than 5000 passengers (54.18% of the total sample), the average error is only 4.91%. For the situation we focus on, this model has high accuracy. Here we divide into different sections to observe the accuracy of the prediction results.

Section 1 shown in Figure 6 is from Nanjing East Road station to people's Square Station, which is a representative large passenger flow interval of line 2. A total of 14 days are randomly selected, and the prediction results are shown in the figure below. It can be seen that when the passenger flow is large, the prediction of normal passenger flow has high accuracy.

Section 2 shown in Figure 7 is the downward start section from Pudong International Airport Station to Haitian 3rd road station. Similarly, 14-day data are selected. The prediction results are as follows. It can be seen that for the low passenger flow range, the prediction error is large, especially in the first hour and the last hour of the operation period, but on the whole, except for individual stations, the error remains within a certain acceptable range, and the model performance is still acceptable.

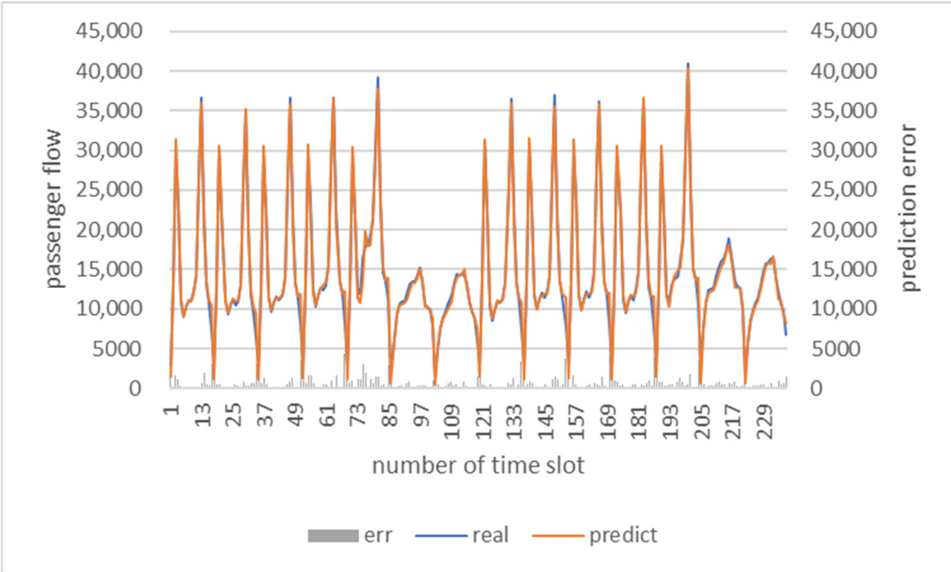

Figure 6. Normal prediction results from Nanjing East Road station to people's Square Station.

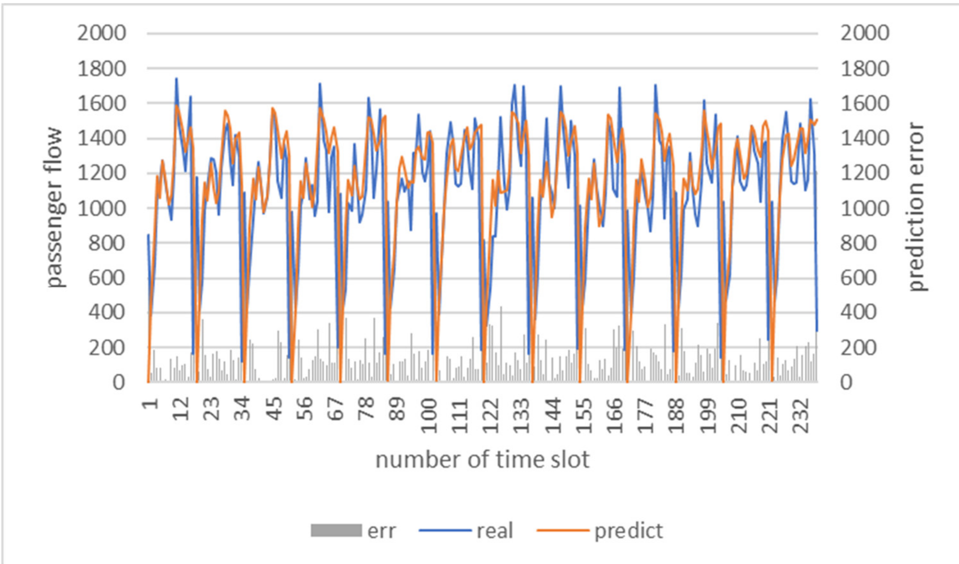

Figure 7. Normal prediction results of Nanjing East Road from Pudong International Airport Station to Haitian 3rd road station.

4.3.2. Emergency Passenger Flow

For emergency passenger flow, there are 20 available data and 580 samples, and the overall prediction error is 4.29%.

Figures 8–10 shows the forecast results of some emergency passenger flow. It can be seen that due to the small number of samples, compared with the normal passenger flow, the prediction error of emergency passenger flow is larger, but the overall error is still less than 10%, which has a certain guiding role.

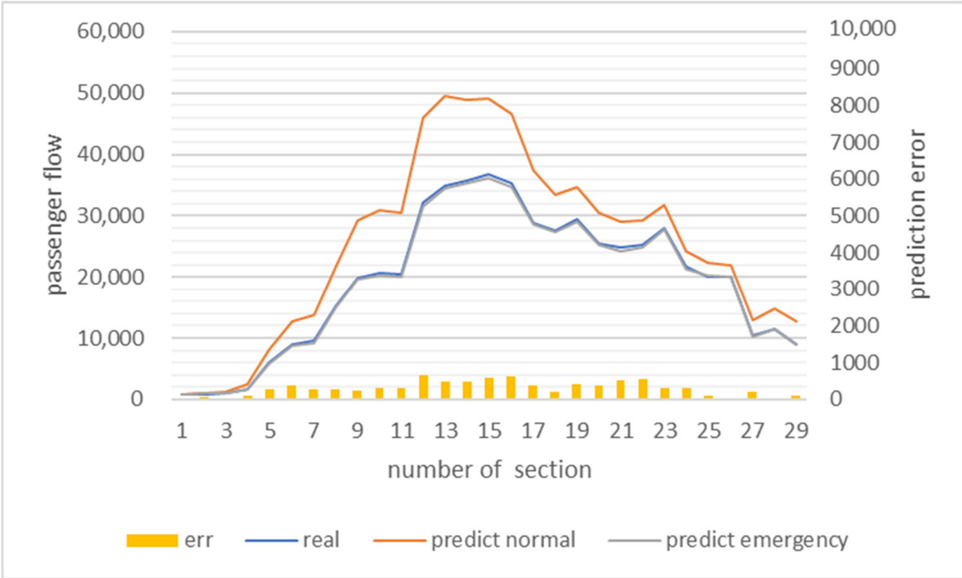

Figure 8. Emergency passenger flow forecast on 22 August 2017.

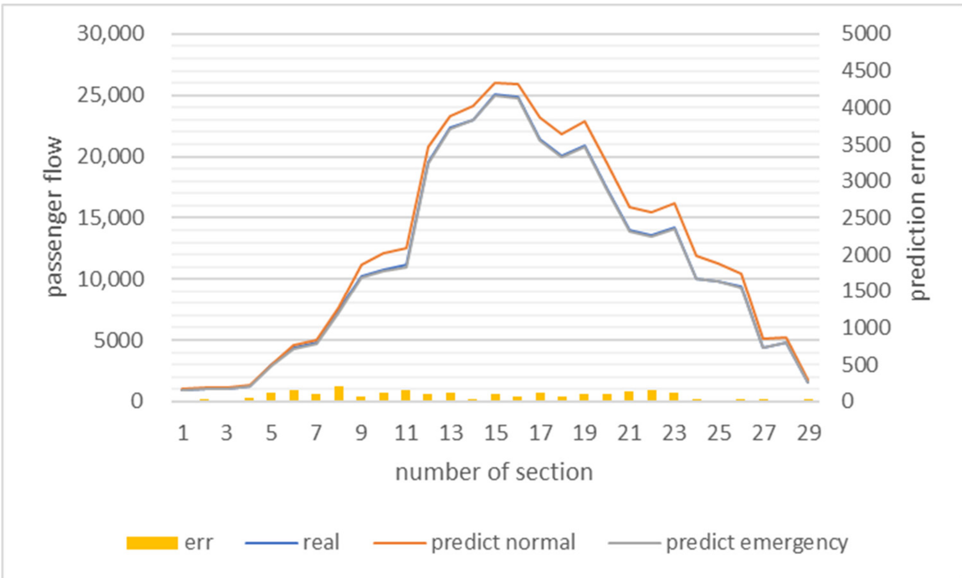

Figure 9. Emergency passenger flow forecast on 12 September 2018.

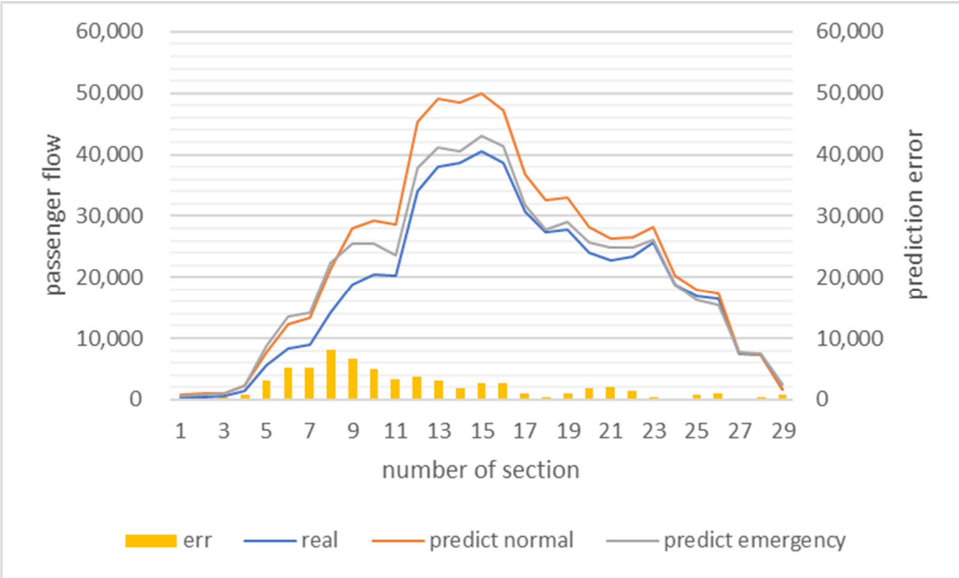

Figure 10. Emergency passenger flow forecast on 23 April 2019.

4.3.3. None Transfer Learning

As mentioned above, when the transfer learning strategy is not used, the accuracy of normal passenger flow prediction is still high when the passenger flow is high, and low when the passenger flow is low, but the overall network performance decreases. For the grouping of different passenger flows, the error increases by about 1% on average.

However, for the prediction of emergency passenger flow, the network accuracy is greatly reduced without transfer learning, with an average error of 12.6%, but the average error is only 9.74% compared with the predicted normal passenger flow, which means that the predicted value of emergency passenger flow is actually closer to the predicted normal passenger flow and cannot correctly predict the emergency passenger flow we need. The following Figure 11 is a representative example.

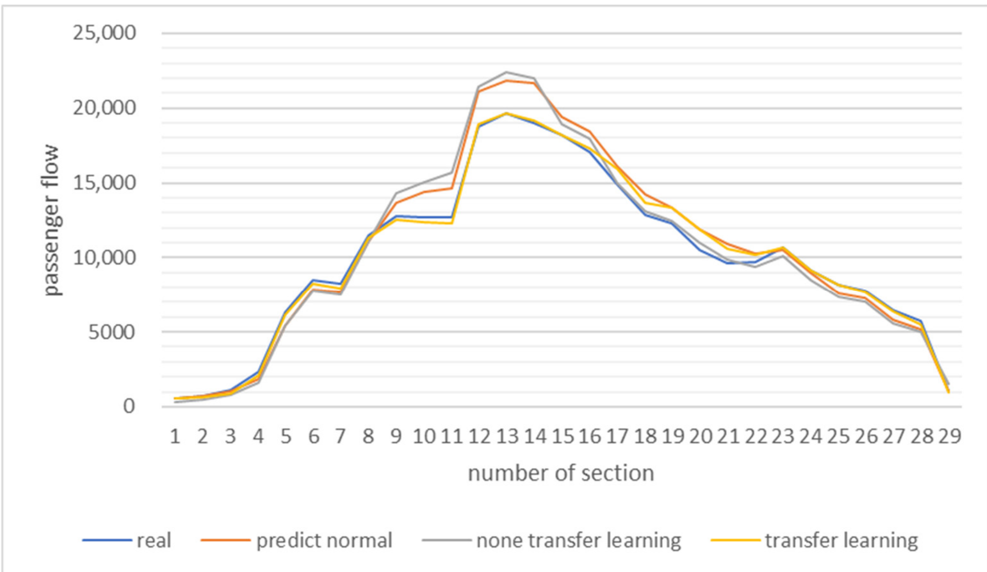

Figure 11. Comparison of the effects of using transfer learning and not using transfer learning.

5. Summary

This paper presents an LSTM prediction network based on transfer learning, which can accurately predict the normal passenger flow and emergency passenger flow of the metro transportation system. Compared with the traditional research using entering and leaving data or IC card data, this paper creatively uses the section passenger flow data which can better reflect the transportation capacity of the metro transportation system. In addition, compared with the traditional algorithm, this paper uses transfer learning to solve the problem of the too-small sample size of emergency passenger flow. In the case of an algorithm without using the transfer learning, due to the interference of the sample size of normal passenger flow, the prediction result will be closer to the normal passenger flow rather than the emergency passenger flow reflecting the real situation. The error rate of the final prediction result is less than 10%, which can provide help for the decision-making of the operating company.

Author Contributions: Conceptualization, J.M. and X.Z.; methodology, J.M.; software, J.M.; validation, J.M. and X.Z.; formal analysis, J.M.; investigation, J.M.; resources, X.Z. and R.D.; data curation, J.M.; writing—original draft preparation, J.M.; writing—review and editing, J.M., X.Z. and X.X.; visualization, J.M.; supervision, X.Z. and X.X.; project administration, X.Z.; funding acquisition, X.Z. and X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shanghai Sailing Program (Grant No. 19YF1451500) and the Program of Shanghai Science and Technology Innovation Action Plan (Grant No. 19DZ1201100).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from Technical Center of Shanghai Shentong Metro Group Co., Ltd. and are available from Ranran Deng and Xin Zeng with the permission of Technical Center of Shanghai Shentong Metro Group Co., Ltd.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Sun, S.; Zhang, C.; Yu, G. A Bayesian Network Approach to Traffic Flow Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2006**, *7*, 124–132. [CrossRef]
2. Cai, L.; Yu, Y.; Zhang, S.; Song, Y.; Xiong, Z.; Zhou, T. A Sample-Rebalanced Outlier-Rejected-Nearest Neighbor Regression Model for Short-Term Traffic Flow Forecasting. *IEEE Access* **2020**, *8*, 22686–22696. [CrossRef]
3. Sun, Y.; Leng, B.; Guan, W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing* **2015**, *166*, 109–121. [CrossRef]
4. Li, H.; Wang, Y.; Xu, X.; Qin, L.; Zhang, H. Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network. *Appl. Soft Comput.* **2019**, *83*, 105620. [CrossRef]
5. Zhang, J.; Chen, F.; Shen, Q. Cluster-Based LSTM Network for Short-Term Passenger Flow Forecasting in Urban Rail Transit. *IEEE Access* **2019**, *7*, 147653–147671. [CrossRef]
6. Yu, W.; Zhifei, W.; Hongye, W.; Junfeng, Z.; Ruilong, F. Prediction of passenger flow based on CNN-LSTM hybrid model. In Proceedings of the 12th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 14–15 December 2019; pp. 132–135.
7. Jiao, F.; Huang, L.; Song, R.; Huang, H. An Improved STL-LSTM Model for Daily Bus Passenger Flow Prediction during the COVID-19 Pandemic. *Sensors* **2021**, *21*, 5950. [CrossRef] [PubMed]
8. Zhao, M.; Zhang, X.; Jin, Y. Wavelet embedded attentive Bi-LSTM for short-term passenger flow forecasting. In Proceedings of the IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, UK, 23–26 August 2021; pp. 177–183.
9. Gao, H.; Xu, J.; Li, S.; Xu, L. Forecast of passenger flow under the interruption of urban rail transit operation. In *Lecture Notes in Electrical Engineering: Proceedings of the 4th International Conference on Electrical and Information Technologies for Rail Transportation (EITRT)*; Springer: New York, NY, USA, 2019; pp. 283–291.
10. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
11. Wang, C.; Mahadevan, S. Heterogeneous domain adaptation using manifold alignment. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; Volume 2, pp. 541–546.

12. Zhou, J.T.; Tsang, I.W.; Pan, S.J.; Tan, M. Heterogeneous domain adaptation for multiple classes. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Reykjavic, Iceland, 22–25 April 2014; pp. 1095–1103.
13. Zhou, J.T.; Pan, S.; Tsang, I.W.; Yan, Y. Hybrid heterogeneous transfer learning through deep learning. In Proceedings of the National Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014; Volume 3, pp. 2213–2220.
14. Raghu, M.; Zhang, C.; Kleinberg, J.; Bengio, S. Transfusion: Understanding Transfer Learning for Medical Imaging, NeurIPS. *arXiv* **2019**, arXiv:1902.07208.
15. Guo, Y.; Shi, H.; Kumar, A.; Grauman, K.; Rosing, T.; Feris, R. SpotTune: Transfer learning through adaptive fine-tuning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4805–4814.
16. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
17. Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; De Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; Gelly, S. Parameter-Efficient Transfer Learning for NLP. In Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research), Long Beach, CA, USA, 9–15 June 2019; Volume 97, pp. 2790–2799.
18. Gordon, M.; Duh, K.; Andrews, N. Compressing BERT: Studying the Effects of Weight Pruning on Transfer Learning. In Proceedings of the Rep4NLP 2020 Workshop at ACL 2020 Conference, Online, 19 July 2020.
19. Mozafari, M.; Farahbakhsh, R.; Crespi, N. A BERT-Based Transfer learning approach for hate speech detection in online social media. In *COMPLEX NETWORKS 2019: Complex Networks and Their Applications VIII*; Springer: Berlin, Germany, 2019; Volume 1, pp. 928–940.
20. Xiao, D.; Huang, Y.; Qin, C.; Liu, Z.; Li, Y.; Liu, C. Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2019**, *233*, 5131–5143. [CrossRef]