2018 International Conference on Security, Pattern Analysis,
and Cybernetics (SPAC)

Short-term Bus Passenger Flow Forecast Based On Deep Learning

1st Xiaoshuang Li
The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
lixiaoshuang2017@ia.ac.cn

2nd Ziyang Chen
Traffic police detachment of Hefei Public Security Bureau
Hefei, China
710887219@qq.com

3rd Fenghua Zhu
The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
Fenghua.zhu@ia.ac.com

4th Wei Chang
iFLYTEK CO.LTD
Hefei, China
weichang2@iflytek.com

5th Chang Tan
iFLYTEK CO.LTD
Hefei, China
changtan2@iflytek.com

6th Gang Xiong
The State Key Laboratory for Management and Control of Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
gang.xiong@ia.ac.com

***Abstract*—The public transportation system is an essential part of the life of the citizens and it's the basis of intelligent transportation system(ITS). This paper tries to predict short-term bus passenger flow by using deep learning approach that called SAE model and DBN model. The model training and evaluation were carried out using the credit card records of the Suzhou bus IC card. The experimental results show that the SAE and DBN models can reduce the prediction error by 9.51% and 10.48%, respectively, compared with the traditional method. The methods of deep learning show a good application prospect in the short-term bus passenger flow forecasting.**

***Index Terms*—intelligent transportation; short-term bus passenger flow forecasting; deep learning ; SAE model ; DBN model**

I. Introduction

Bus passenger flow forecast is an important part of intelligent public transport system [1], is one of the key links to realize intelligent public transportation. The results of the forecast for the bus system will have an important impact on the operation and resource allocation. The existing prediction models mainly include regression prediction [2], [3] and time series analysis [4], [5] based on macroscopic traffic parameters, neural network method [6], [7] based on knowledge discovery, prediction method based on chaos theory and forecasting method based on combinatorial theory [8], these methods have yielded some results in terms of actual traffic flow forecasts. Regression prediction analysis with multiple linear regression as an example, selecting the corresponding variable as the argument, configuring different weights based on historic data, determing the future flow of traffic. Although this method can have a higher accuracy, but heavily dependent on the selected variables and artificial experience. In the time series analysis method, exponential smoothing method is widely used on periodic data, but it can not reflect the advantages of large-scale data. The improvement of data size can not effectively improve the performance of model prediction. The short-term bus passenger flow system is a random system with strong uncertainty and complex nonlinearity. Many factors will have an influence on it, which makes it highly non-linear. And traditional methods have some difficulties in dealing with such problems.

The arrival of the digital consumer age, which is represented by IC card, help the formation of large traffic data [9]. Data driven [10] model is showing a stronger vitality than the traditional model-driven approach. Deep learning, which is based on neural network, has strong ability to study and fit nonlinear function, good fault tolerance and is able to find out the complex structure concluded in big data [11]. Pekel [12] proposed a parlimenttary optimization algorithm based on POA-ANN algorithm and IWD-ANN algorithm. And this algorithm has a very good performance when it's used to forecast the passenger flow. Zhou [13] successfully used BP neural network and RBF neural network for short-term passenger flow forecasting in Chengdu. Duan [14] used SAE model to forecast traffic flow at different time and Huang [15] and his parterner used DBN to solve the same problem. Lv [16] using deep learning approach to forecast traffic flow and get good effect. We find that the research of deep learning in public passenger flow is still less. At the same time, the feature of traffic flow and bus passenger flow some place is similar and some place is different to each other. This paper try to use SAE model and DBN model to forecast the short-term bus passenger flow with the work above.

II. Model Construction

The SAE (stack autoencoder) model is a special deep neural network model that superimposes multiple AE models. It is an unsupervised learning artificial neural network. AE has the same numbers of input and output nodes. The number of hidden layer nodes is generally not less than the number of nodes in the input layer [17].

An AE consists of two parts: an encoder and a decoder. encoder f maps an input vector x into its hidden layer and decoder g maps the hidden representation y into its output layer. The mapping process is shown in equation (1) and (2).

$$\boldsymbol{y = f(W_1 x + b)} \tag{1}$$

$$\boldsymbol{z = g(W_2 x + c)} \tag{2}$$

Where x is the input vector, z is the output vector, y is the hidden layer output vector, W_1 and W_2 are the coding matrix and the decoding matrix, f and g are the activation function of the hidden layer and the output layer. This paper use sigmoid function as the activation function, and $f(x) = g(x) = 1/(1+e^{-x})$. AE transforms the input of the visual layer to the hidden output layer and then reconstructs the hidden layer to achieve the target that the output of the autoencoder is almost equal to the original input itself, so the training goal is to minimize the reconstruction error as shown in equation (3).

$$\boldsymbol{L(X, Z) = \frac{1}{2}\sum_{i=1}^{N} \parallel x^i - z^i \parallel^2} \tag{3}$$

The training algorithm is as algorithm1:

Algorithm 1 Training the AE

Hyperparameters: Given data set $\boldsymbol{X = x, x \in R^{KM}}$. Set the number of hidden units $\boldsymbol{n_h}$ and the iterations $\boldsymbol{T}$

Initialize: system coding matrix $\boldsymbol{W_1(n_h \times KM)}$, decoding matrix $\boldsymbol{W_2(KM \times n_h)}$ and biases $\boldsymbol{b(1 \times n_h)}$, $\boldsymbol{c(1 \times KM)}$ randomly;

while $\boldsymbol{i < T}$ **do**
Perform forward propagation to compute $\boldsymbol{Y}$, the output of AE;
Compute output error:$\boldsymbol{X - Y}$;
Perform backward propagation to compute $\boldsymbol{\Delta\theta}$ with BP algorithm;
Update $\boldsymbol{\theta = \theta + \Delta\theta}$
$\boldsymbol{j = j + 1}$;
end while

After stacking several layers of AE, a regression layer is added to the top of the SAE model and used as a predictor. Then it's necessary to training the network to minimize the reconstruction error. The training algorithm is as algorithm2.

DBN(Deep belief network) is one kind of deep neural networks and it is formed by cascade restrict boltzmann machine. RBM(Restrict Boltzmann Machine) is a special form of the boltzmann machine (BM). It is a random network[11].

Algorithm 2 Training the SAE

Hyperparameters: Given data set $\boldsymbol{X = x, x \in R^{KM}}$,$\boldsymbol{Y = y}$,$\boldsymbol{y \in R^N}$

Initialize: Set the number of hidden layers $\boldsymbol{H}$, the number of hidden units $\boldsymbol{n_h}$, $\boldsymbol{h = 1, 2, 3, ..., H}$, the pretraining iterations $\boldsymbol{T}$, the fine-tuning iterations $\boldsymbol{S}$, $\boldsymbol{W_{L+1}(N \times n_h)}$,$\boldsymbol{b_{L+1}(1 \times N)}$;

Pretraining:
$\boldsymbol{X}$ is fed to the input layer;
Train the first AE using the algorithm1
for $\boldsymbol{h = 2 : H}$
The output of the $\boldsymbol{h - 1}$ layer AE is used as the input of the $\boldsymbol{h}$ layer AE, train this AE with algorithm1.
end for
fine-tuning:
for $\boldsymbol{i = 1 : S}$
Perform forward propagation to compute $\boldsymbol{Y}$
Compute output error $\boldsymbol{X - Y}$
Perform backward propagation to compute parameter change $\boldsymbol{\Delta\alpha}$ with BP algorithm
Update $\boldsymbol{\alpha = \alpha + \Delta\alpha}$
end for

The standard RBM consists of a binary hidden layer and a visible layer (input layer). There is no connection in the same layer of RBM, but BM is a feedback neural network composed of random neuron connections, symmetrical and non-self-feedback[12].

The RBM is an unsupervised learning method, a probability graph model. The ultimate goal of learning is to maximize the fitting of the distribution of input data, so that the Gibbs distribution represented by RBM and the distribution of the input data are maximally consistent. Using the KL distance to measure the distance between the distribution of the sample representation and the edge distribution of the network represented by the RBM network. And the problem is transformed into a solution to minimize the KL distance as the eqution (4)

$$\begin{aligned} KL(q \parallel p) &= \sum_{x \in \Omega} q(x) ln \frac{q(x)}{p(x)} \\ &= \sum_{x \in \Omega} q(x) lnq(x) - \sum_{x \in \Omega} q(x) lnp(x) \end{aligned} \tag{4}$$

Where Ω is the sample space, q is the distribution of input data, p is the edge distribution of the Gibbs distribution represented by the RBM. The training algorithm of RBM is as algorithm3.

After stacking several layers of RBM, a regression layer is added to the top of the DBN model and used as a predictor. Then it is necessary to train the network to minimize the KL distance. The training algorithm is as algorithm4.

III. Evaluation criteria

In order to evaluate the performance of the final model for short-term bus passenger flow prediction, this paper adopt

978-1-7281-0551-2/18/$31.00 ©2018 IEEE

Algorithm 3 Training the RBM

Hyperparameters: Given data set $\boldsymbol{X}$,number of hidden units $\boldsymbol{n}$, learning rate $\boldsymbol{\alpha}$,iterations $\boldsymbol{T}$
Initialize: $\boldsymbol{v_1} = \boldsymbol{X}$,weights matrix $\boldsymbol{W}$,biases $\boldsymbol{a}, \boldsymbol{b}$(samll random numbers)
for $\boldsymbol{j = 1 : T}$
for $\boldsymbol{s = 2 : n}$(for all hidden units)
Compute $\boldsymbol{P(h_s \mid v_1)}$, where $\boldsymbol{P(h_s \mid v_1) = sigmoid(b_s + \sum_i v_{1i} w_{i,s})}$
Sample $\boldsymbol{h_s} \in \boldsymbol{\{0, 1\}}$ from conditional distribution $\boldsymbol{P(h_s \mid v_1)}$
end for
for $\boldsymbol{r = 1 : m}$(for all visible units)
Compute $\boldsymbol{P(v_{2r} = 1 \mid h)}$, where $\boldsymbol{P(v_{2r} = 1 \mid h) = sigmoid(a_r + \sum_i w_{r,i} h_r)}$
Sample $\boldsymbol{v_2 r} \in \boldsymbol{\{0, 1\}}$ from conditional distribution $\boldsymbol{P(v_2 r = 1 \mid h)}$
end for
for $\boldsymbol{t = 1 : n}$(for all hidden units)
Compute $\boldsymbol{P(h'_t = 1 \mid v_2)}$, where $\boldsymbol{P(h'_t = 1 \mid v_2) = sigmoid(b_t + \sum_i v_{2i} w_{i,t})}$
Sample $\boldsymbol{v_2 r} \in \boldsymbol{\{0, 1\}}$ from conditional distribution $\boldsymbol{P(v_2 r = 1 \mid h)}$
end for
Update:
$\boldsymbol{W = W + \alpha(P(h = 1|v_1) v_1^T - P(h' = 1|v_2) v_2^T)}$
$\boldsymbol{a = a + \alpha(v_1 - v_2)}$
$\boldsymbol{b = b + \alpha(P(h = 1|v_1) - P(h' = 1|v_2))}$
end for

Algorithm 4 Training the DBN

Hyperparameters: Set the number of hidden layers in the DBN model l , the number of hidden units n, train data set X, Y
Unsupervised pretrain: X is fed to the first RBM, train the first RBM with the algorithm3, the result is h_1
for $i = 2 : l$ Using h_{i-1} as input data of the $i - th$ RBM, train the RBM with algorithm3, the result is h_i
end for
fine-tuning:
h_l is fed to the predictor, set cost function, train the predictor with gradient descent method
Using the parameters in the network which has been pretrained as the initial parameters, fine tune the whole network in which the predictor is included.

three different criteria to measure the error of predicted data. They are Mean absolute error (MAE), Root mean square error, (RMSE), Mean relative error, (MRE) and they are defined as eqution(5)(6)(7).

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \mid y_i - \hat{y_i} \mid \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y_i})}{N}} \tag{6}$$

$$MRE = \frac{1}{N} \sum_{i=1}^{N} \frac{\mid y_i - \hat{y_i} \mid}{y_i} \tag{7}$$

Where y_i is the observed bus passenger flow at time i and $\hat{y_i}$ is the predicted one. MAE measures the average error predictions over the entire test set. The smaller MAE in the whole test set, the smaller error ,the better model. RMSE measures whether the distribution of error data is stable and the smaller RMSE indicate that the singularity of the predicted value, which is far from the true value, be less, the prediction is more accurate. MRE measures the average size of the error data relative to the true data and the smaller MRE show that the absolute value of error relative to the real passenger data is smaller. These three criteria measure different points and they are used in this paper at the same time.

IV. Experiments and analysis

Before the prediction experiment, we need to decide the structure of the network. There exists some hyper parameters such as the number of hidden layers, the numbers of hidden layer units, activation function, the times of iteration and so on and they need a large number of pre-experiments to decide the best combination of different hyper parameters. TableI shows the possible value of hyper parameters and after pre-experiments, we find the best combination of hyper parameters for every criterion, the tabelII show the parameters value with the best results.

TABLE I: hyper parameters setting

Hyper parameters	Range	Experiment value	
		SAE	DBN
Activation function	sigmoid	sigmoid	sigmoid
Number of hidden layers	1–5	1,2,3,4,5	1,2,3
Number of hidden units	5–15	5,7,10,15	5,6,7,8,9,10,15
Times of pretraining iterations	100–1000	500,1000	1000
Times of fun-tuning iterations	100–1000	500,1000	1000
Time delay steps	1–7	7	7

Figure1 shows the contrast of the passenger forecast and the true value at different time in May 26th for SAE model. It also shows the distribution of predicting data and corresponding error data when the criteria of MAE/RMSE and MRE get the best results separately (the network structures of MAE and RMSE are same when RMSE and MAE get the best result separately). These two methods both behave well in forecasting the bus passenger flow. The time of morning and evening peak, the trend of flow change have been forecasted accurately. The structure which RMSE/MAE gets the best result perform better than the structure which MRE gets the best result, on predicting the time of morning and evening peak.

Figure2 shows the contrast of the passenger forecast and the true value at different time in May 26th for DBN model. It also shows the distribution of predicting data and corresponding

TABLE II: The optimal parameters of each model network structure

Network name	SAE			DBN		
Criterion	MAE	RMSE	MRE	MAE	RMSE	MRE
Number of hidden layers	4	4	5	2	1	3
Number of hidden units	10	10	10	5	9	5
Times of iterations	1000	1000	500	10000	10000	10000
Value for criterion	10.9808	14.044	17.11%	10.826	13.8625	16.14%

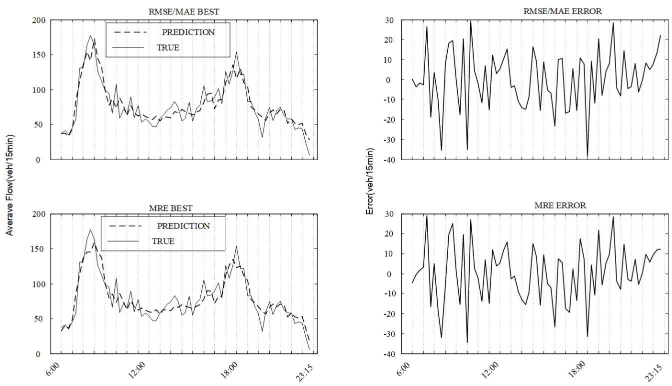

Fig. 1: SAE model forecast experiment and the prediction error

error data when the criteria of MAE,RMSE and MRE get the best results separately. The left side of fig.1 and fig.2 show the comparison of forecast data and true value. It's obvious that forecast data in the peak, the trend of flow change and so on can be consistent with the true value well. The right side of figure1 and figure2 show the error data in the test dataset and the difference of forecast data and true value is between -20 and +20 in most instances, evenly distributed in the 0 scale.

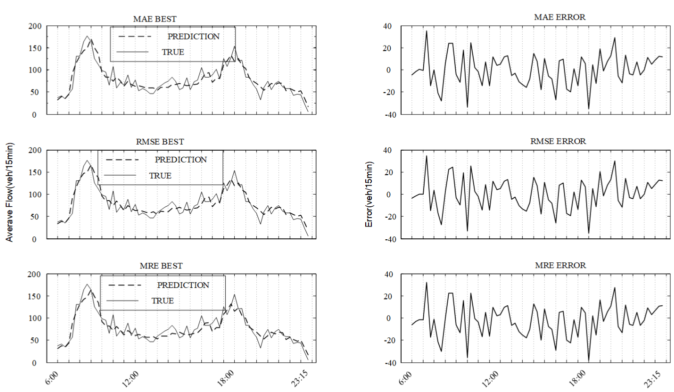

Fig. 2: DBN model forecast experiment and the prediction error

In the experiment, the last 15% of the whole data is selected as a test set. There are 491 test points and calculate the mean and variance. The result is as tableIII.

TableIII shows that DBN model has a better performance than other models whether we compare mean and variance. Less absolute mean indicates that the data predicted by DBN model is closer to the true value and less variance indicates the error is steadier, the number of singularity point is fewer.

At the same time, we compare the experiment results with multiple linear regression method and Holt-went exponential smoothing method. We did the same forecast experiments using the same training dataset and test dataset, evaluate the methods with MAE, RMSE and MRE. TableIV shows the results of different method.

TableIV shows that deep learning method: SAE model and DBN model both perform well than traditional method on mostly all criteria and they have better performance when they are used in the actual forecasting task. The performance of two different deep learning model: SAE and DBN are similar, but the performance of the traditional model is obviously different and it also indicates that traditional approach relies heavily on the model itself so human experience has a significant impact on the final prediction accuracy

Short-term bus passenger flow has non-stationary randomness [18] and complex nonlinear characteristics. The traditional methods through a different combination of ways to achieve a higher degree of data fitting and they almost are linear methods, which is not able to solve nonlinear problem perfectly. Multiple linear fitting method influences the final result by change the weights of every arguments and when the data size is much larger than the number of model arguments, the role of data size is not obvious. The weight coefficient of the exponential smoothing method changes according to the exponential law, and the weight coefficient is corrected according to the real data and the prediction error of the previous time. Exceptional points also have a direct effect on the final result, for example the wrong data couldn't be recognized and the model fault tolerance is low, however it is important for model itself.

The basic unit of the deep learning network is neurons and there existing a nonlinear transformation process when calculating, which make it good at non-linear fitting. There exist large scale parameter matrix compared with the traditional model in the deep learning model. After the non-linear transformation of the neurons, the increasing of the data scale can effectively improve the performance and fault tolerance of the model. It explains the reason why deep learning model has better performance than traditional model in short-term bus passenger flow forecast.

V. Conclusion and Prospect

In this paper, as the original data set, We use SAE and DBN deep learning models to forecast bus passenger flow with Suzhou City bus operating dataset. For MAE, RMSE and MRE three different evaluation indicators, respectively, finding the corresponding best combination of hyper parameters. At the

TABLE III: Statistical analysis of error data

Model	SAE		DBN		
	MRE	RMSE/MAE	MAE	RMSE	MRE
average value	-0.4758	-0.4212	-0.0614	-0.4819	1.9702
variance	197.7713	197.4529	192.5991	192.3274	194.0466

same time, this paper also compares deep learning methods with multiple linear regression and exponential smoothing method. The results show that the deep learning model can realize the forecast of bus passenger flow and the performance is superior to the traditional methods. In the short-term bus passenger flow forecasting, the deep learning has broad application prospect.

TABLE IV: The model performance indicators

Method	MAE	RMSE	MRE
SAE	10.9808	14.0438	17.11%
DBN	10.826	13.8625	16.14%
multiple linear regression	11.0736	14.2402	16.63%
Holt-went exponential smoothing	17.2673	23.9811	26.62%

Deep learning is more suitable for large-scale data than traditional method and in this article a total of only 2829 data is used to train the models. The data size is relatively small and the strong learning ability of deep learning may not be able to be shown. A larger scale of data which will be used to train and evaluate the models is of great significant for improving the performance of deep learning models. On the other hand, this paper divides the dates into two categories: weekdays and non-weekdays, to predict the bus passenger flow in weekdays. But in practice, weekdays and non-weekdays are alternating. So there is an idea that forecast the bus passenger flow with the passenger flow data for consecutive dates, making use of the learning ability of deep learning models to verifies whether the overall accuracy of the bus passenger flow forecast can be improved.

Acknowledgment

This work was supported in part by National Key R&D Program of China 2018YFB1004803, National Natural Science Foundation of China 61533019 and 61773381.

References

[1] S. Tao, D. Rohde, and J. Corcoran, "Examining the spatial–temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap," *Journal of Transport Geography*, vol. 41, pp. 21–36, 2014.

[2] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.

[3] S. Clark, "Traffic prediction using multivariate nonparametric regression," *Journal of transportation engineering*, vol. 129, no. 2, pp. 161–168, 2003.

[4] R. Xue, D. J. Sun, and S. Chen, "Short-term bus passenger demand prediction based on time series model and interactive multiple model approach," *Discrete Dynamics in Nature and Society*, vol. 2015, 2015.

[5] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results," *Journal of transportation engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[6] T.-H. Tsai, C.-K. Lee, and C.-H. Wei, "Neural network based temporal feature models for short-term railway passenger demand forecasting," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3728–3736, 2009.

[7] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.

[8] X. Shenglong, "Research and application of dynamic prediction method for urban road short-term traffic flow," Master, Changan University, 2015.

[9] F. Z. T. J. Gang Xiong, Xisong Dong, "Urban traffic big data technology and intelligent application system," *BigData*, vol. 1, no. 4, p. 2015042, 2017.

[10] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: http://www.nature.com/articles/nature14539

[12] E. Pekel and S. Soner Kara, "Passenger flow prediction based on newly adopted algorithms," *Applied Artificial Intelligence*, vol. 31, no. 1, pp. 64–79, 2017.

[13] H. Zhou, R. Yao, and X. Xu, "Predicting Short-Term Bus Passenger Flow Using Neural Network Models," Tech. Rep., 2017.

[14] "Performance evaluation of the deep learning approach for traffic flow prediction at different times - IEEE Conference Publication." [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7551691/

[15] W. Huang, G. Song, H. Hong, and K. Xie, "Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning." *IEEE Trans. Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.

[16] Y. Lv, Y. Duan, W. Kang, Z. Li, F.-Y. Wang *et al.*, "Traffic flow prediction with big data: A deep learning approach." *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2015.

[17] L. Weixi, "Medical image retrieval based on sae feature extraction," Master, Guangdong University of Technology, 2016.

[18] S.-l. QIAO, R.-c. SUN, and J. LIU, "Short-term Traffic Flow Forecast Based on Deep Learning," *Journal of Qingdao University (Natural Science Edition)*, vol. 4, p. 013, 2017.