Short-term BRT Passenger Flow Prediction with A Deep Learning Method

Haitao XU, Jiaxue PENG, Ning ZHENG, Ying GAO

School of Computer Science and Technology, Hangzhou Dianzi University, 310018 Hangzhou, China

***Abstract* —Predicting passenger flow accurately can make full use of the existing traffic facilities to prevent emergencies from happening which is caused by heavy passenger flow. However, most of the existing methods to predict short-term passenger flow are shallow architecture, they cannot mine the relationship between short-term passenger and other external factors deeply enough. Therefore, we consider using a deep learning method to predict Bus Rapid Transit (BRT) passenger flow, which is different from previous methods to predict BRT passenger flow. We use a Prediction Stacked Auto Encoder (PSAE) model to mine BRT passenger flow features deeply. The top of the whole model establishes a prediction model for BRT passenger flow prediction from the extracted feature set. It uses a greedy layer-wise algorithm to train model parameters. Experimental results show that this method has good performance for predicting short-term BRT passenger flow.**

Keywords- predict short-term BRT passenger flow;deep learning, prediction stacked autoencoder

I. INTRODUCTION

With the rapid development of the intelligent transportation system (ITSs), passenger flow prediction has been captured more and more attention by researchers. In recent years, a large number of researchers use prediction model in this area [1]. Passenger flow is a typical time-series data and predicting short-term passenger flow exactly is an integral part in intelligent transportation fields. It contributes to individual passengers, bus companies, government departments and so on. It is also helpful to make optimal route choice decision for passengers, reduce congestion and emissions and improve the public transport operation efficiency potentially. Predicting passenger flow can be treated as a crucial element of the ITS. At the same time, it is not only an essential component of the traffic management system which needs to seize passenger flow efficiently and accurately, but also the advanced public transportation system and the commercial vehicle operation.

Passenger flow prediction is similar to the traffic flow prediction in the ITS extremely. The accuracy of predicting passenger flow highly depends on the collected data from the bus device and the real-time passenger flow data. Therefore, the study of traffic flow prediction methods have significant reference on the passenger flow prediction. In other words, a good deal of high-quality passenger flow data is crucial for this research. Because of historical data accumulation which is combined with environmental factors related to weather and holidays, efficient traffic management and control departments is becoming more and more data-driven [2][3]. Although there existed a lot of traffic flow predicting models, whose methods mostly use shallow architecture and model, and its performance is not meted the demand in some respects. Under the circumstance, this paper considers about using deep architecture to deal with the large number of complicated traffic flow data.

In recent years, the attention of academe and industry was attracted by deep learning as an active branch of machine learning fields [4]. Deep learning has been applied to dealing with classification problems, natural language processing, dimensionality reduction, image recognition, voice recognition, sentiment analysis successfully [5]-[9]. It is very widely used in image recognition, speech recognition filed [10]-[12].

Deep learning uses the deep architecture which is also called multi-layer architecture which can transform high-dimensional data into low-dimensional data by dimensionality reduction and extracts the inherent features of the data set from the high-dimensional data. Processing and analyzing passenger flow data is a complex task. Deep learning algorithms can help us get the intuitive feature expression of passenger flow data without any prior knowledge, and has good performance on passenger flow prediction.

The paper proposed a BRT passenger flow prediction method which is based on the deep learning method. Stacked autoencoder (SAE) is a deep architecture which can realize reducing the dimension of data and use greedy layer-wise algorithm to train the parameters of the model. In the BRT passenger flow prediction method, an improved stacked autoencoder model was used to predict BRT passenger flow, which is called Prediction Stacked Autoencoder (PSAE).The model contain two steps: feature learning and model learning. The feature learning step can mine spatiotemporal correlations between BRT passenger flow data and the external factors by using SAE model. After feature learning, the paper uses model learning to construct prediction layer.

The next part of this paper is as follows: The second part focuses on the study of some short-term traffic flow prediction method. The third section describes the model and training algorithm mentioned in my paper. The fourth section explains experiment results and gives comparative experiment results with other shallow architecture. Finally, the last part will give the summary and prospect.

II. RELATED WORKS

There are many similarities between BRT passenger flow prediction and traffic flow prediction, so the research of traffic flow prediction methods have significant reference on the BRT passenger flow prediction. Over the past decade, there have been a lot of research on traffic flow prediction. It may be inspired to handle BRT passenger flow data by continuing to study these prediction algorithm and model. To sum up, traditional traffic flow prediction approaches contains the following three categories:

The first category is Time-series model. This model is a widely used method to process traffic flow prediction. Integrated moving average model (ARIMA) [13] is a most widely used time-series model which is the mixture of self-mixing regression and moving average models. Unlike other methods, it doesn't need fixed initialization simulation, while it sees certain time traffic as more general non-stationary random sequence. [14] proposed a prediction scheme using Seasonal ARIMA (SARIMA) model for short term prediction of traffic flow using only limited input data. A different class of time-series models called structural time-series model (STM) has been introduced in [15] to develop a parsimonious and computationally simple multivariate short-term traffic condition forecasting algorithm. Most existed models are designed to improve the accuracy of prediction, such as seasonal ARIMA [16]. Three different time-series models, viz. random walk model, Holt-Winters' exponential smoothing technique and seasonal ARIMA model are used for modeling of traffic flow in Dublin[17].

The second category is nonlinear system theory and intelligent learning theory. Because of the stochastic and nonlinear feature of traffic flow data, more and more researchers begin to focus on using nonlinear ways to predict traffic flow. A good deal of study has shown that nonlinear methods usually have better performance, they can get the complicated nonlinear and uncertainty relationship of time-series data. [18] has proposed an accurate multi-steps traffic flow prediction model which was based on support vector machine (SVM). This method is different from traditional single-step prediction method. The representative ways are artificial neural networks [19], support vector regression [20] and partial weight learning [21]. [22] proposed a method for adaptive hybrid fuzzy rule-based system to predict traffic flow. [23] proposed an approach that combines fuzzy state transform and Kalman filter forecasting model and considered the advantage of the two models, a weight combination model is proposed. [24] used optimized BP neural network based on particle swarm optimization (PSO) to predict short-term traffic flow. [25] used a hybrid model that combines clustering with support vector machine to predict public bicycle traffic flow by exploiting complementary advantages of both approaches. [26] combined wavelet analysis method with support vector machine (SVM), then proposed a short-term traffic flow prediction combined model based on SVM.

The third category is deep learning method. This is a new way that is quite different from the previous method. All in all, due to the rapid development of ITS, more and more traffic data was generated and the features of the data sets appeared to be different. Therefore, researchers consider of attempting to predict traffic flow with more efficient ways. It is hard to say which way to predict the traffic flow is the best method and that there is a better performance than any others. The prediction accuracy of the algorithm and model highly depends on the size and the features of the historical data that we used. A large number of research results demonstrated that the neural network approaches generally have better prediction performance and accuracy.

Currently, deep architecture which contrasts with neural network (NN) has better performance on mining the features of data. At the same time, the traditional neural networks only contain one hidden layer and use the traditional way which is based on stochastic gradient descent method to train the whole network. But for multi-layer neural network, it is difficult to use stochastic gradient descent way to train the network.This kind of training method is easy to fall into local optimum. However, [27] proposed a greedy layer-wise training algorithm with unremitting efforts in such a way that training deep architecture becomes more and more easy by Hiton. Generally, the deep architecture has good effect at realizing dimensionality reduction [28], processing classification. Thankfully, a few scattered authors try to apply deep learning method to ITS. [29] proposed a novel deep temporal-spatial traffic flow feature learning mechanism with large scale Taxi GPS traces for traffic prediction. A deep Restricted Boltzmann Machine and Recurrent Neural Network architecture is utilized to model and predict traffic congestion evolution based on Global Positioning System (GPS) data from taxi [30]. Deep belief network (DBN) is a typical deep architecture. It can learn effective features for traffic flow prediction in an unsupervised way. [31] is the first time that combined DBN with multitask regression and applied the deep learning approach to transportation research. A stacked autoencoder model is used to learn generic traffic flow features, and it is trained in a greedy layer-wise fashion. This is the first time that a deep architecture model is applied using autoencoder as building blocks to represent traffic flow features for prediction [32]. To the best of our knowledge, this several studies is the most innovative line of research which directly apply the deep learning method in ITS. In this paper, we use a deep architecture model PSAE to achieve dimensionality reduction of multi-dimensional data, and use a modified SAE which is called PSAE to perform BRT passenger flow prediction task.

III. MODEL AND ALGORITHM

Deep learning model can achieve dimensionality reduction and classification, but it is rarely directly used as prediction model. We use SAE as feature model firstly, SAE can find principal feature direction by nonlinear transformation, it can extract principal features from the original BRT passenger flow data, then we can construct prediction model according to the extracted features. On the basis of the extracted features $x=(x_1,x_2...,x_m)$ and prediction target $y=(y_1,y_2...,y_m)$ construct $\{(x_1,y_1),(x_2,y_2),...,(x_m,y_m)\}$,as pair of input and output and train the prediction model finally. Therefore, BRT

DOI 10.5013/ IJSSST.a.17.40.23 ISSN: 1473-804x online, 1473-8031 print

passenger flow prediction is divided into two parts, feature learning and model learning.

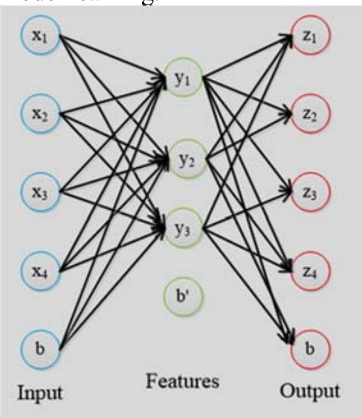

Figure 1. Autoencoder

A. Feature Learning

We can get the feature expression of the original data and achieve reducing data dimension by feature learning. Autoencoder is a kind of neural network architecture, such as the figure (1) above, it consists of the encoder and decoder. The two steps can be transformed as follows:

Encoder: During the encoder process, encode the original training set x into the hidden layer and generate the representation y. f_θ is called encoder. The procedure can be described as (1):

$$f_\theta(\mathbf{x}) = s(\mathbf{W}x + \mathbf{b}) \tag{1}$$

Its parameters is $\theta = \{\mathbf{W}, \mathbf{b}\}$, $\mathbf{W}$ is the weight matrix, $\mathbf{b}$ is the bias value.

Decoder: During the encoder process, the result of the hidden layer representation $\mathbf{y}$ is used to reconstructed the original input vector $\mathbf{z}$ in the original training set, $\mathbf{z} = g_{\theta'}(\mathbf{y})$. $g_{\theta'}$ is called decoder. The whole decoder process is as (2):

$$g_{\theta'}(\mathbf{y}) = s(\mathbf{W}' y + \mathbf{b}') \tag{2}$$

Its parameters is $\theta' = \{\mathbf{W}', \mathbf{b}'\}$. Logistic sigmoid function $1/(1+\exp(-x))$ is considered for $s(x)$.

By constructing the minimal reconstruction loss function, we can get the hidden layer's parameters θ of the model, and the minimal reconstruction loss function expression is as follows:

$$\begin{aligned} \theta &= arg\ \min_\theta L(X, Z) \\ &= arg\ \min_\theta \frac{1}{2} \sum_{i=0}^{N} \left\| x^{(i)} - z(x^{(i)}) \right\|^2 \end{aligned} \tag{3}$$

Under normal circumstances, the feature learning model contains more than one layer autoencoder. These autoencoder is stacked from the bottom to up. That is stacked autoencoder (SAE).

Suppose that we use the N layers deep architecture, the greedy layer-wise training algorithm for pretrain the SAE process is as follow:

(1) Train the first layer as an autoencoder in a purely unsupervised way, use the output and original input construct the miniming reconstruction error.

(2) Use the each output of the hidden layer as the input of the next layer, control the reconstruction error in a certain range.

(3) Iterate as in (2),compute the whole hidden layers training until the desired layer.

(4) Use the output of the last layer as input of the supervised layer and initialize parameters the all supervised layers

B. Model Learning

Model learning process is based on the extracted features and construct the prediction model. In machine learning, prediction is also a kind of classification, Logistic regression is a very efficient classifier. It can not only predict the categories of samples, but also calculate the probability of classification. Logistic regression is to learn such a function:

$$f(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \tag{4}$$

Among them,

$$g(z) = \frac{1}{1 + e^{-z}} \tag{5}$$

Known as the Logistic function or Sigmoid function.

On the basis of the feature model SAE, we construct the regression prediction model based on the extracted features, and construct the prediction model by using the characteristic value and the target value. This hybrid model is called PSAE, and the topology is shown below figure(2):

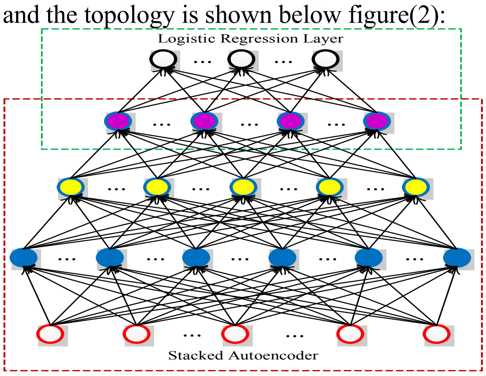

Figure 2. Prediction Stacked Autoencoder (PSAE)

The algorithm to train the combined feature learning and model learning can be described as follows:

---

(1)Use the original input data as the input firstly, and the first layer is trained by an unsupervised method, and the minimum error objective function is constructed;

(2)Use the output of the above layer as the input of the next layer, and construct an autoencoder.Continuing to use an unsupervised method to train the automatic encoder;

(3) Repeat iteration steps (2) until layer n-1;

(4) Build the logistic regression prediction layer by the extracted features and prediction target, and train the parameters of the prediction layer;

(5)Construct the cross entropy, use the back propagation algorithm, and use the gradient descent optimization technique to fine-tune the parameters of the whole network architecture, so that the parameters can be optimized.

IV. EXPERIMENTAL

In this paper, in order to verify the performance of the proposed model, the experiment used Chinese Shaoxing BRT IC card records of the citizen as original data. It can observed that the traffic situation of everyday by IC card records. This dataset contains IC card records between 25 June 2013 to 2 December 2014. It means that the experiment used the real data of the whole 525 days. Meanwhile, BRT passenger flow is related to weather, holidays, and many other factors. Thus, this experiment will combine weather, holidays and other factors as the feature of the experiment data except for the real history of IC card records. In the experiment, the IC card records was counted every 5 minutes, and the time interval of BRT passenger flow data was between 6 am to 5 pm of 12 hours every day. In the 525 days historical data, the first 500 days of date is used as training set and the data of final 25 days as a test set.

To evaluate the performance of the experiment model, we use three indicators mean absolute error, relative error and mean squared error .they can be computed as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=0}^{N} | x_i - \hat{x}_i |$$

$$\mathrm{MRE} = \frac{1}{N}\sum_{i=0}^{N} \frac{| x_i - \hat{x}_i |}{x_i}$$

$$\mathrm{RMSR} = \left[\frac{1}{N}\sum_{i=0}^{n} (| x_i - \hat{x}_i |)^2 \right]^{\frac{1}{2}}$$

Where x_i is the observed BRT passenger flow, $\hat{x}_i$ is the predicted BRT passenger flow.

A. *Model Architecture*

In the experiment, the principal task is to determine the size of the input layer, the number of hidden layers, the number of nodes in each hidden layer and the size of the output layer in the PSAE network architecture. There is no best solution to determine the best network architecture in the currently academic community. Getting the relative effective number of nodes and network layers can only be supported by providing a lot of comparative experiments. Based on historical experience, BRT passenger flow is related to objective historical weather factors and holidays.

Thus, during the experiment, the weather and someday which is a weekend day are used as features during the experiment. In the specific experiment procedure, the day's weather and whether someday is the weekend are set to (W_1, W_2) in order to predict the sequence of BRT passenger flow $(X^1, X^2, ..., X^m)$ during time interval t, excluding the use of raw sequence data of BRT passenger flow $(X^{t-1}, X^{t-2}, ..., X^{t-r})$ .The values of W_i are as follow:

$$W_1 = \begin{cases} 0, & workday \\ 1, & weekend \end{cases}$$

$$W_2 = \begin{cases} 0, & no \text{ heavy rain} \\ 1, & \text{heavy rain} \end{cases}$$

Taken together, the input vector of deep architecture is $(W_1, W_2, X^{t-1}, X^{t-2}, ..., X^{t-r})$, it can generate the target output finally.

To prove the reasonableness of adding extra features, we added comparative experiment to ensure that the added features can be recognized by the deep architecture and produce more accurate results. Through a large number of repeated comparison experiment, the ideal depth of the architecture was determined. Among them, the hidden layers is 4, the number of nodes in each hidden layer is set $\{100, 150, 90, 50, 33\}$. It must be affirmed that the first hidden layer node number must be consistent with the original input data dimensions in the PSAE model. In the process of comparative experiment, the actual dimensions of the original input data related with an added features. Then, after the addition of weather and holidays features, the number of hidden layers remained 4 constantly, but the number of the first input layer node becomes 102.Therefore, the hidden layers become $\{102, 150, 90, 50, 33\}$. During the comparative experiment, we can not get the desired results if hidden layers and the number of nodes are too big or too small, they can only be adjusted by the experiment experience.

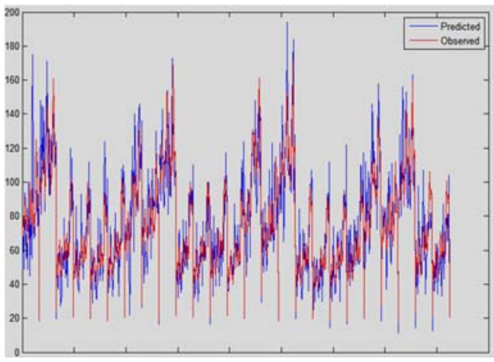

Figure 3. Experimental Results

B. *Results*

Observing the experimental results as figure (3) above, we find that the trend of the experimental results are consistent with observed results. The trend that it shows is accorded with the conditions of big passenger flow and moderate passenger flow. But the results of the experiment is not very good when passenger flow is small, it will cause large errors. It is consistent with the most results generated by the existing traffic prediction methods. The experiment will have a greater relative error in the small passenger flow. Therefore, the proposed method is more concerned about the case of large passenger traffic prediction.

TABLE 1. EXPERIMENTAL RESULTS

work	***PSAE***			***BP Neural Network***			***SVM***		
	MAE	MRE(%)	RMSE	MAE	MRE(%)	RMSE	MAE	MRE(%)	RMSE
5-min BRT flow prediction	11.8	17.32	2	15.6	20.3	12	14	19.8	11

The following table (1) shows that the model PSAE have a better performance compared with BP neural networks (BPNN) and support vector machine (SVM). Through the model proposed by experiments compared with BPNN and SVM, among these methods, BPNN is a classic non-linear learning model, but it is a shallow feature learning model. In this experiment, MAE predicted by neural network was 15.6%, while MRE was 20.3%, and RMSE was 12.As showed in the table. With respect to the PASE model, indicators were slightly inferior in all respects. SVM is a linear classifier defined on maximum interval feature space and a typical second-class classification model. In the comparative experiment, we use SVM to compare with PSAE model and found that the experiment results has a large gap with PSAE , MAE generated by SVM was 14%, while MRE was 19.8%, RMSE was 11. The Intuitive graphics display is as figure(4).

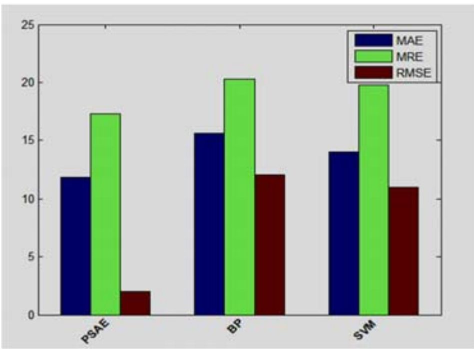

Figure 4. Comparing performance

By comparing the experiment results of the three ways, deep learning was found having better performance on processing BRT passenger flow data. It is hard to say which kind of algorithm is better performance than the other model, because the effect of the experiment is more relevant to the data. All we can do is to verify the effect of specific algorithm for data analysis results through experiments for real data.

V. CONCLUSION

This paper presents a PSAE model based on deep learning to predict BRT passenger flow. It is different from many previous passenger flow prediction method. The method presented in this paper can mine BRT passenger flow in spatiotemporal correlation unsupervised. We use greedy layer-wise training algorithm to pre-train the deep architecture and use back propagation algorithm to adjust the parameters of deep architecture, resulting in smaller error in network output. In this experiment, we use Chinese Shaoxing BRT passenger flow as experiment data, it was proved that the performance of the proposed method is relatively good with compared the BP neural network and support vector machine.

In future studies, the author will attempt to use other deep architecture to predict the BRT passenger flow, and try to use other common data set to verify the performance of the algorithm we proposed. In addition, logistic regression layer is used as the prediction layer in this paper, perhaps author will attempt to use other more efficient regression model to improve the experiment accuracy.

REFERENCES

[1] M. Chen, Y. Wei, Exploring time variants for short-term passenger flow,J. Transp. Geogr. 19 (4) (2011) 488–498.

[2] J. Zhang etal., “Data-driven intelligent transportation systems: A survey,IEEETrans.Intell.Transp.Syst.,vol.12,no.4,pp.1624–1639,Dec.2011.

[3] C. L. Philip Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” Inf. Sci.,vol. 275, pp. 314–347, Aug. 2014.

[4] Y. Bengio, “Learning deep architectures for AI,” Found. Trends Mach.Learn., vol. 2, no. 1, pp. 1–127, Jan. 2009.

[5] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507,Jul. 2006.

[6] R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in Proc. 25th ICML, 2008, pp. 160–167.

[7] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, “Multi-digit number recognition from street view imagery using deep convolutional neural networks,” arXiv preprint arXiv:1312.6082, 2013.

[8] B. Huval, A. Coates, and A. Ng, “Deep learning for class-generic object detection,” arXiv preprint arXiv:1312.6885, 2013.

[9] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach,“Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data,” IEEE Trans.Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1930–1943, Aug. 2013.

[10] Wang Y, Wang D L. Towards Scaling Up Classification-Based Speech Separation[J]. Audio Speech & Language Processing IEEE Transactions on, 2013, 21(7):1381-1390.

[11] Morgan N. Deep and Wide: Multiple Layers in Automatic Speech Recognition[J]. IEEE Transactions on Audio Speech & Language Processing, 2012, 20(1):7 - 13.

[12] Hinton G E, Osindero S, Yw. T. A fast learning algorithm for deep belief nets.[J]. Neural Computation, 2006, 18(7):1527-54.

[13] [13]. Yussof S, Razali R A, See O H. An investigation of using parallel genetic algorithm for solving the shortest path routing problem.[J]. Sciencepublications, 2011, :7(:2).

[14] Kumar S V, Vanajakshi L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data[J]. European Transport Research Review, 2015, 7(3):1-9.

[15] Ghosh B, Basu B, O'Mahony M. Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis[J]. Intelligent Transportation Systems IEEE Transactions on, 2009, 10(2):246-254.

[16] [16] B. M. Williams and L. A. Hoel, “Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results,” J. Transp. Eng., vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003.

[17] Ghosh B, Basu B, O'Mahony M. Time-Series Modeling For Forecasting Vehicular Traffic Flow in Dublin[J]. 2005.

[18] Zhang M, Zhen Y, Hui G, et al. Accurate Multisteps Traffic Flow Prediction Based on SVM[J]. Mathematical Problems in Engineering, 2013, 2013(6):91-109.

[19] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, “Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg Marquardt algorithm,” IEEE Trans. Intell. Transp. Syst., vol. 13, no. 2, pp. 644–654, Jun. 2012.

[20] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, “Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions,” Expert Syst. Appl., vol. 36, no. 3, pp. 6164–6173, Apr. 2009.

[21] M. Shuai, K. Xie, W. Pu, G. Song, and X. Ma, “An online approach based on locally weighted learning for short-term traffic flow prediction,” in Proc. 16th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst., 2008, p. 45.

[22] L. Dimitriou, T. Tsekeris, and A. Stathopoulos, “Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow,” Transp. Res. C, Emerging Technol., vol. 16, no. 5, pp. 554–573,Oct. 2008.

[23] Deng M J, Qu S R. Fuzzy State Transition and Kalman Filter Applied in Short-Term Traffic Flow Forecasting[J]. Computational Intelligence & Neuroscience, 2015, 2015(7):1-7.

[24] Li S, Wang L, Liu B. Prediction of Short-Term Traffic Flow Based on PSO-Optimized Chaotic BP Neural Network[C]// Proceedings of the 2013 International Conference on Computer Sciences and Applications. 2013:292-295.

[25] Xu H, Ying J, Wu H, et al. Public Bicycle Traffic Flow Prediction based on a Hybrid Model[J]. Applied Mathematics & Information Sciences, 2013, 7(2):667-674.

[26] Yang Y N, Lu H P. Short-Term Traffic Flow Combined Forecasting Model Based on SVM[C]// International Conference on Computational & Information Sciences. IEEE Computer Society, 2010:262-265.

[27] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural Comput., vol. 18, no. 7, pp. 1527–1554,Jul. 2006.

[28] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504-507.

[29] Niu X, Zhu Y, Zhang X. DeepSense: A novel learning mechanism for traffic prediction with taxi GPS traces[C]// Global Communications Conference. IEEE, 2014:2745- 2750.

[30] Ma X, Yu H, Wang Y, et al. Large-scale transportation network congestion evolution prediction using deep learning theory.[J]. Plos One, 2015, 10(3).

[31] Huang W, Song G, Hong H, et al. Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5):2191-2201.

[32] Lv Y, Duan Y, Kang W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2):865-873.

.