This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.2986995, IEEE Transactions on Fuzzy Systems

T2F-LSTM Method for Long-term Traffic Volume Prediction

Runmei Li, Yongchao Hu, Qiuhong Liang

Abstract— **Long Short-Term Memory (LSTM) neural network shows excellent performance in learning, processing and classifying time series data but with some limitations such as high computational cost and lack of interpretability. Fuzzy neural networks, which combine the complementary capabilities of both neural networks and fuzzy system, thus constitute a more promising technique for processing traffic flow. This paper presents a Type-2 Fuzzy Long Short-Term Memory (T2F-LSTM) neural network model for long term traffic volume prediction. Type-2 Fuzzy Sets (T2FSs) provide more freedom to describe membership information and process data with higher uncertainty better than traditional fuzzy system does. In this paper, an interval T2FSs is introduced to extract the probability distribution and spatial-temporal characteristics of traffic volume. Using parameters of Closure of Support (*CoS*) obtained in interval T2FSs, weights of input gate in LSTM neural network are updated and converged to the region with larger slope of the sigmod function fast. The network interpretability is also increased by better control the information flow using motivational factors constructed from the parameters. Experiment conducted with real traffic volume data shows that the proposed model achieves more accurate prediction result and shorter network training time.**

Index Terms—**Traffic volume, Long-term prediction, Type-2 fuzzy sets, Closure of Support, Long Short-term Memory Neural Network**

I. INTRODUCTION

ACCURATE traffic flow prediction provides information for users to make decisions in advance. For example, it is an important reference for vehicle path planning. Predicting where and when congestion will occur is greatly beneficial for transportation management, as practitioners would be able to allocate resources to the roads most at risk for congestion and ultimately reduce the traffic congestion [1]. Based on the length of time, traffic flow prediction can be divided into short-term, medium-term, and long-term. The one with the time span from 30-second to 15-minute is usually regarded as the short-term prediction. Medium-term prediction generally allows a range from 30-minute to a couple of hours, and long-term prediction denotes longer period, such as 24 hours in advance [2].

Since 1980s, researchers began to study traffic flow prediction. A couple of different prediction methods were proposed. Time series analysis model [3]， Kalman filters [4], support vector regression [5], K-nearest neighbor [6], the gradient boosting tree regression [7], and hybrid models [8] were widely used in traffic flow prediction.

Most of the above methods are short-term prediction. In the event of a traffic accident, a short-term prediction model should be precise providing the vicinity measurements of traffic flow in the subsequent instants of the aforementioned accident. By contrast, long-term predictions allow users to have a global insight of the traffic at any time and are useful for designing public transportation policies or road planning. This is the reason why current research in traffic flow forecasting models is mainly focused on predicting road traffic from minutes to hours into the future [9-11]. Lana et al. [9] clustered the historical traffic data to seek the data patterns and parameters of each cluster. Combining the historical averaging methods with random forest methods helps improve the prediction accuracy. Thomas et al. [10] proposed an auto-regressive integral moving average model to predict 24-hour traffic volume using traffic flow data of urban roads in Almelo. Note that these time series models are ideal traffic volume forecasting schemes in relatively stable traffic conditions. Drastically varying traffic conditions require excessive calculation and may yield slower forecasting accuracy.

The field of transportation studies has in recent years seen an increased interested in neural network applications due to the inability of traditional methods to address the complexity of traffic flow characteristics. Results that are better than that obtained using mathematical models have thus been achieved using neural networks. Lana et al.[11] used spiking neural networks for adaptation to changes in traffic forecasting using similarity-based clustering of daily traffic volume data. Jiang et al. [12] developed a time-delay recurrent wavelet neural network model to forecast traffic flow and highlighted the importance of periodicity on long-term forecasting. Yi et al. [13] applied the multi-layer feed forward neural network to traffic volume prediction.

Now, deep learning technology has attracted a lot of interest in academic and industrial world. It has been applied successfully in classification tasks, natural language processing, dimensionality reduction, object detection, motion modeling, and forecasting [14]. Wu et al. [15] proposed a Deep Neural Network (DNN) based traffic flow prediction model to improve the prediction accuracy using weekly/daily periodicity and spatial-temporal characteristics of traffic flow. Lv et al. [16] proposed a deep-learning-based traffic flow prediction method which considered the spatial and temporal correlations inherently. A stacked autoencoder model was used to learn generic traffic flow features, and it was trained in a greedy layer-wise fashion. Yang et al. [17] proposed a stacked autoencoder Levenberg-Marquardt model, which is a type of deep architecture of neural network approach aiming to improve forecasting accuracy. The proposed model was designed using the Taguchi method to develop an optimized structure and to learn traffic flow features through layer-by-layer feature granulation with a greedy layerwise unsupervised learning algorithm. These methods all belong to fully-connect structure. It is difficult for a fully-connected neural networks to capture representative features from the dataset with plentiful characteristics [15].

A Recurrent Neural Network (RNN) is an extension of regular artificial neural networks that add connections feeding the hidden layers of the neural network back into themselves, these are called recurrent connections. RNNs have been widely used for processing sequential data. However, RNNs are commonly difficult to train due to the well-known gradient vanishing and exploding problems and hard to learn long-term patterns. LSTM was developed to address these problems [18-20]. Ma et al. [21] applied LSTM to forecast traffic speed, and demonstrated that LSTM could capture the long-term temporal dependence of traffic data. Liu et al. [22] proposed an end-to-end deep learning architecture which combined convolution and LSTM being able to extract the spatial-temporal information from the traffic flow. Liu et al. [23] established a series of LSTM with deep neural layers using 16 settings of hyper-parameters and investigates their performance on a 90-day travel time dataset from Caltrans Performance Measurement System.

Although some results are promising, the difficulties in the design and implementation of LSTM neural networks remain unresolved and the opaqueness of the trained networks prevents understanding the underlying models. Moreover, the use of hyperbolic tangent and the sigmoid action functions results in gradient decay over layers in LSTM neural networks. Consequently, construction of an efficiently trainable deep network is challenging [24]. A possible solution is to adopt a knowledge-based approach to model the problem and subsequently using the knowledge independently. In the case of traffic flow processing, such knowledge can represent those characteristics that are invaluable to traffic engineer. Fuzzy neural networks, which combine the capabilities of both neural networks and fuzzy logic, are seen as a very promising technique for automatically deriving from experimental data an approximate knowledge-based model. The results obtained highlight the capability of fuzzy system to add interpretability as well as generalization ability to the input data [25][26]. Traffic information or data availability go with uncertainty and randomness, which should be taken into account during data processing and forecasting. Studies have shown the ability of fuzzy sets to capture data uncertainty, especially type-2 fuzzy sets [27-31]. The combination of type-2 fuzzy set theory and deep learning method will consider both precision and interpretability of model [32,33].

In our previous works [27-29], the central limit theorem or K-means clustering method were adopted to convert multi-day traffic flow point data in one sampling period into interval data of the same period. Then the interval data were used as the input data of Type-2 fuzzy sets forecasting model. These type-2 fuzzy sets-based methods had strong capability of anti-noise, but the forecast precision was relatively low. This motivates us to seek an improved approach in this work.

This paper describes a novel approach of traffic flow processing and prediction using a specific class of fuzzy neural networks known as the Type-2 Fuzzy Long Short-term Memory neural network (T2F-LSTM) model. This model shows how to introduce the knowledge extracted by type-2 fuzzy sets from raw traffic volume data into deep neural network. The bulk of the proposed prediction model is a LSTM neural network. The prediction results obtained highlight the capability of the model to extract a working fuzzy knowledge sets as well as generalize from the input data. The key work of this paper can be generalized as follows:

1) We use confidence interval to transform multi-day traffic volume point data in one sampling period into interval data of the same period which describe the fluctuation characteristic of traffic flow more properly, and lay the data foundation for T2FSs. We use the interval data to obtain the embedded type-1 fuzzy sets (T1FSs).

2) We use Adaptive Network-based Fuzzy Inference System (ANFIS) to verify that the current traffic flow is most closely related to the previous 15-minute. ANFIS helps to determine the number of T1FSs to perform union operation to obtain *CoS*. The spatial-temporal feature of traffic flow are extracted by *CoS* which parameters are used in LSTM neural network as motivational factors.

3) We choose LSTM neural network as the main algorithm for long-term traffic volume prediction. *CoS* information is added to the weight update formula of the input gate as motivational factors on the activation function output to improve the network interpretability and computational expense. The proposed approach encourages activation function outputs to satisfy a target pattern with the weight have the ability of learning the fluctuation and spatial-temporal characteristic of data more accurately.

Experimental results demonstrate that the model proposed in this paper possesses a higher degree of prediction capability as well as better noise tolerance than tradition LSTM neural network. The model gives full play to the ability of type-2 fuzzy set theory to deal with uncertainty data and the ability of LSTM to implement prediction accuracy.

II. RELATED WORKS

A. *LSTM Neural Network*

Fig. 1 shows the architecture of LSTM neural network. Unlike the traditional neural networks, the basic unit of the hidden layer of LSTM is memory block [32]. The memory block contains memory cells with self-connections memorizing the temporal state, and a pair of adaptive, multiplicative gating units to control information flowing into the block. Two additional gates named input gate and output gate respectively control the input and output activations into the block. The core of memory cell is a recurrently self-connected linear unit-Constant Error Carousel (CEC), and the activation of the CEC represents the cell state. Due to the present of CEC, multiplicative gates can learn to open and close, and thus LSTM can solve the problem of vanishing error problem by keeping the network error constant. To prevent the internal cell values growing without bound when processing the continual time series that are not previously segmented, a forget gate is added to the memory block. This treatment enables the memory blocks to reset by itself once the information flow is out of date, and replaces the CEC weight with the multiplicative forget gate activation [21].

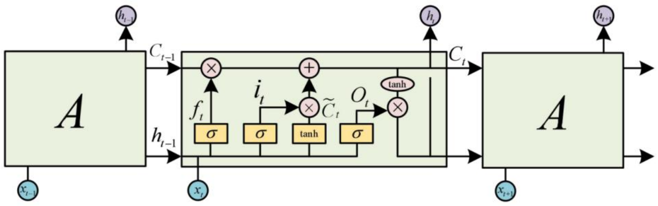


Fig. 1. LSTM neural network architecture.

Set the input traffic volume data in a sliding window be $\{x_1, x_2, \ldots, x_t, x_T\}$, the hidden vector sequence $\mathbf{H} = \{h_1, h_2 \cdots, h_t, h_T\}$ is calculated then an output sequence $\mathbf{Y} = \{y_1, y_2 \cdots, y_t, y_T\}$ is obtained by formula (1) and (2).

$$h_t = H(\mathbf{W_h} \cdot [h_{t-1}, x_t] + \mathbf{b_h}) \tag{1}$$

$$y_t = \mathbf{W_y} h_t + \mathbf{b_y} \tag{2}$$

where $\mathbf{W_h}$ is the input-hidden weight matrix, $\mathbf{b_h}$ is the input-hidden bias vectors. $\mathbf{W_y}$ is the hidden layer output weight matrix, $\mathbf{b_y}$ is the hidden layer output bias vectors. $H(\cdot)$ is hidden layer function and obtained by following equations (3)-(8):

$$f_t = \sigma(\mathbf{W_f} \cdot [h_{t-1}, x_t] + \mathbf{b_f}) \tag{3}$$

$$i_t = \sigma(\mathbf{W_i} \cdot [h_{t-1}, x_t] + \mathbf{b_i}) \tag{4}$$

$$\widetilde{C}_t = \tanh(\mathbf{W_c} \cdot [h_{t-1}, x_t] + \mathbf{b_c}) \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{6}$$

$$O_t = \sigma(\mathbf{W_o} \cdot [h_{t-1}, x_t] + \mathbf{b_o}) \tag{7}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{8}$$

where $\sigma(\cdot)$ is sigmoid function, $\sigma(\cdot)$ and $\tanh(\cdot)$ are defined by formula(9) and (10). f, i, o, c are forget gate, input gate, output gate and cell update gate respectively. $\mathbf{W_f}$ is the forget gate weight matrix, $\mathbf{W_i}$ is the input gate weight matrix, $\mathbf{W_c}$ is the cell update gate weight matrix, and $\mathbf{W_o}$ is the output gate weight matrix. $\mathbf{b}$ is the bias vectors(e.g. $\mathbf{b_i}$ is the input gate bias vector)

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{9}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{10}$$

B. Interval Type-2 Fuzzy Sets

In 1975, Zadeh presented the definition of type-n fuzzy sets [34], followed by a definition of T2FSs in 1976 [35]. Then relatively new definitions for T2FSs were presented by Mendel and others [36]. Now the study of T2FSs, especially interval type-2 fuzzy sets (IT2FSs), has been one of a popular direction in artificial and computational intelligence [37]. T2FSs have been successfully used in many areas, such as computing with words [38,39],forecasting of time-series [40], clustering [451], pattern recognition [42], fuzzy logic controller [43], industrial application [44], simulation [45] and stock market index prediction [46]. In 2016, Mo Hong et al. presented the latest definition, which is shown as follows [47].

Preliminaries: Let X, Y be two sets, and for any $x \in X, y \in Y$, there is an ordered pair (x, y). The set of all these ordered pairs called Cartesian Produce of X, Y, written as $X \times Y$: $X \times Y = \{(x, y) | x \in X, y \in Y\}$. X is the universe of discourse, for every $x \in X$,there is $L_x = I$. Let μ_ω be a membership function on X defined as:

$$\begin{aligned} \mu_\omega\colon\ & X \to I^I \\ & x \mapsto g_x \end{aligned} \tag{11}$$

Definition 1. Let ω be a T2FS on X, defined as:

$$\omega = \{(x, u, z) | \forall x \in X, \forall u \in L_x \in C(2^I), z = \mu_\omega^2(x, u) \in \mathrm{I}\}$$

where x, u, z are master variable, sub variable and third variable respectively, and L_x is primary membership function which can be obtained by :

$$\mu_\omega^1\colon\ X \to C(2^I) \tag{12}$$

That is, for $\forall x \in X$, there is $L_x \in C(2^I)$ that makes $\mu_\omega^1(x) = L_x$. Where μ_ω^1 is called primary membership function. μ_ω^2 is second membership function shown as formula (13).

$$\begin{aligned} \mu_\omega^2\colon & \bigcup_{x \in X} x \times L_x \to I \\ & x \times u \mapsto z \end{aligned} \tag{13}$$

For every $x \in X$ and $u \in L_x$, there is $\mu_\omega^2(x, u) = 1$, then ω is an interval type-2 fuzzy sets.

Definition 2. Let ω be a T2 FS on X. The Cartesian product of X *and* I is called the Closure of Support (*CoS*).

$$CoS(\omega) = \overline{\{(x, u) | \mu_\omega^2(x, u) > 0\}} \tag{14}$$

III. MODEL DEVELOPMENT

A. Traffic Flow Data Preprocessing

Liu and Mendel [38] and Wu and Mendel[39] presented a very practical type-2-fuzzistics methodology for obtaining interval type-2 fuzzy set models for words, which was called Interval Approach (IA). The basic idea of the IA is to collect interval endpoint data for a word from a group of subjects, map each subject's data interval into a prespecified type-1 person membership function, interpret the latter as an embedded T1 FS

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.2986995, IEEE Transactions on Fuzzy Systems

of an IT2 FS, and obtain a mathematical model for the footprint of uncertainty for the word from these T1FSs. By definition of Mo[47], Interval Approach is used in this paper to obtain *CoS*. The central limit theorem is adopted to convert single traffic volume data of mass traffic flow in some time range into interval data of the same time range (also called confidence interval data).

Let $x_1, x_2, ..., x_n$ be random sequences that are independently and identically distributed, and: $E(x_i) = \mu, D(x_i) = \sigma^2 > 0, i = 1,2,3,...$, then $\{x_i\}$ are normally distributed: $\frac{x_n - n\mu}{\sqrt{n}\sigma} \to N(\mu, \sigma^2)$.

For such a normal distribution, the axial quantity can be constructed according to formula (15):

$$G = G(x_1, x_2, ..., x_n, \mu) = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \tag{15}$$

Given the variance σ^2, the expected confidence interval of μ is:

$$(\overline{x} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{x} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) \tag{16}$$

Confidence interval is the estimation interval of population parameter of sample statistics estimates. It is applied to obtain confidence interval and confidence level of the true value of traffic flow data from a large amount of data here. When $\alpha = 0.05$, this interval contains 95% of the actual traffic flow data. The interval data describe the fluctuation characteristics and the possible range of traffic flow in a specific location. They also express the spatial characteristics of traffic flow.

B. *Correlation Analysis of Traffic Flow in Adjacent Time*

Considering the spatial and temporal correlations inherently and highlighting the importance of periodicity on long-term forecasting, an Adaptive Network-based Fuzzy Inference System (ANFIS) [48] is used to determine how many previous traffic volume data have strong correlation with current traffic volume. Thus we can determine the maximum number of corresponding type-1 fuzzy sets which will be used to aggregate *CoS*.

ANFIS method was proposed by Jang. It was a fuzzy inference system implemented in the framework of adaptive networks. By using a hybrid learning procedure, ANFIS can construct an input-output mapping based on both human knowledge and stipulated input-output data pairs. ANFIS architecture was employed to model nonlinear functions, identify nonlinear components on-line in a control system, and predict a chaotic time series, all yielding remarkable results [48].

Fig.2 shows the structure of ANFIS used here which is used as a correlation analysis model. It is essentially a prediction model with different number input data leading to output with different error value which we define as loss value. Loss value is used to evaluate the strength of correlation between the current traffic flow and the previous traffic flow at the same monitoring point.

According to literature on traffic flow proceeding [49], researchers often use the sampling traffic flow of the previous 10 moments at most to obtain the characteristics of the current traffic flow data. So we calculated the correlation between the traffic flow from the previous 1 to the previous 10 moments and the current moment. In other words, when we use at most the previous 10 traffic flow data $(x(k-1), x(k-20,..., x(k-10))$ to get the characteristics of the current traffic flow ($y(k)$), do we necessarily get the minimum loss value?

The process is briefly described below:

Step1 Construct single input ANFIS. The input variables are combined with the output variables to obtain a binary group (x_1, y). Then input the binary group to ANFIS for learning and the corresponding prediction errors e_{x1} are obtained which are used as loss value $Loss_{x_1}$.

Step2 Construct double input ANFIS. The selected combinations (x_1, x_2, y) are put into the system again. Calculate the corresponding loss value $Loss_{x_1 x_2}$.

Step3 Increase the number of input variables until $Loss_{x_1 x_2 ... x_{10}}$ obtained.

Step4 All loss values are compared, and the smallest loss value corresponds to the best number of input variables.

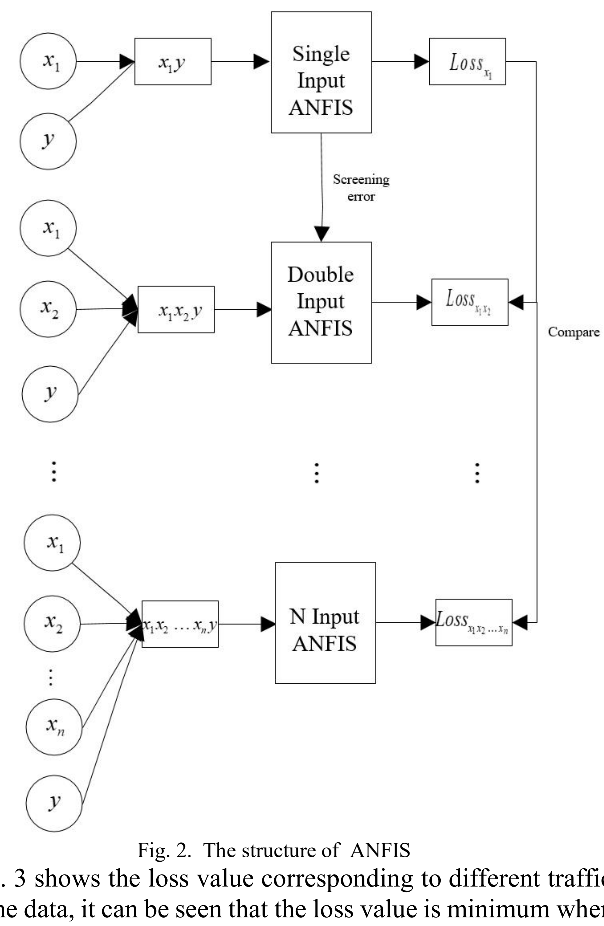


Fig. 2. The structure of ANFIS

Fig. 3 shows the loss value corresponding to different traffic volume data, it can be seen that the loss value is minimum when using 3 previous traffic volume data. If the number of input variables is more or less than 3, the loss value is greater. That means that traffic volume data at time $k-1, k-2, and \quad k-3$

have strong correlation with current time (k) traffic volume. So the type-1 fuzzy sets corresponding to traffic volume at time of $k-1, k-2, and \quad k-3$ can be used to aggregate *CoS*.

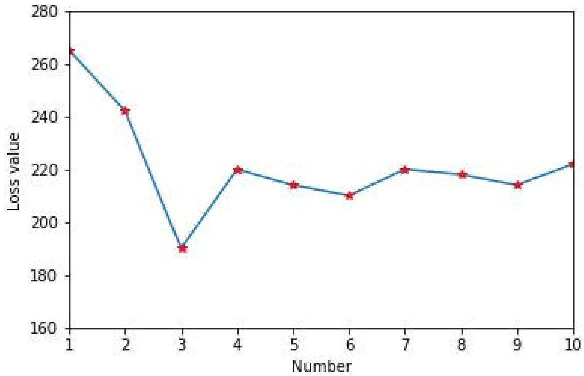


Fig. 3. Loss value analysis of ANFIS

C. Generating CoS of Interval Type-2 Fuzzy Sets

The parameters of the respective embedded T1FSs are determined using the data statistics and uncertainty measures from intervals obtained above. A uniform distribution is assigned to each of the intervals $[a^{(i)}, b^{(i)}]$, and its mean and standard deviation are computed as follows[43]:

$$m^{(i)} = \frac{a^{(i)} + b^{(i)}}{2} \tag{17}$$

$$\sigma^{(i)} = \frac{b^{(i)} - a^{(i)}}{\sqrt{12}} \tag{18}$$

There are different methods to map a data interval into T1FS. In IA, it is achieved by equating $m^{(i)}$ in (17) and $\sigma^{(i)}$ in (18) to the mean and standard deviation of a T1FS. Each confidence interval has a corresponding type-1 fuzzy set model. By a couple of type-1 fuzzy sets corresponding to traffic volume data in adjacent sampling periods, the interior *CoS* is obtained exactly needed in type-2 fuzzy sets as shown in Fig.4 [39].

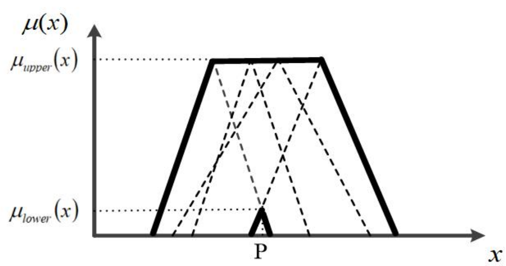


Fig. 4. An example of the union of type-1 fuzzy sets.

The upper boundary of *CoS* is the upper membership function, with the lower membership being similar. The upper and lower boundary values $\mu_{upper}(x), \mu_{lower}(x)$ are obtained which express the spatial-temporal characteristics of traffic flow. They are used in LSTM network as motivational factors.

D. Updating Weights of LSTM Neural Networks

LSTM input gate weight $\mathbf{W_i}$ are updated by formula(19):

$$\mathbf{W_i} = \mathbf{W_i} + X_i(\mu_{upper}(x) - \mu_{lower}(x)) \tag{19}$$

where $\mathbf{W_i}$ is the weight of the input gate which is calculated by back propagation algorithm, X_i is the input traffic flow of the input gate at time t.

Chai et al. [50] proposed above approach to update the weights. After the learning process, these weights represent the strengths of the fuzzy rules. Among the links, at most one link with the highest weight is chosen and the others are deleted. When all the weights of the links between a rule node and the output-label nodes of an output linguistic node are very small or almost equal to each other, this means that this rule node has little or no relation to that particular output variable. That just gives an interpretation of the network link relationship. With the similar method, formula (19) uses $u_{upper}(x)$ and $u_{lower}(x)$ to limit the data in the region with larger slope of the sigmod function. The validity of this work can also be supported by [51]. Make it a little bit more intuitive, this method aims to use gates to control information flow in the recurrent computations, although its practical implementation based on soft gates only partially achieves this goal. This method pushes the output values of the gates towards 0.5 quickly as shown in Fig.5. By doing so, we can better control the information flow: the gates are mostly in a middle state, which makes the results more interpretable. Empirical studies show that (1) We achieve better or comparable performances due to its better generalization ability; (2) With the weights entering around 0.5, the convergence speed of training set is also been dramatically increased.

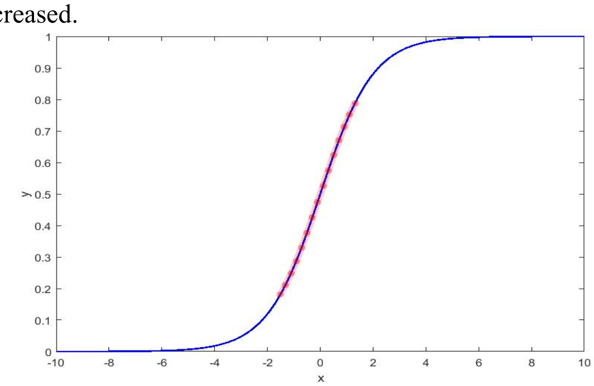


Fig.5 The sigmoid function

IV. Algorithm Verification Using Real Data

The weekday traffic volume data collected from Mountain Henglong Tunnel, Shenzhen are selected to validate the proposed model. The raw data are processed as the standard traffic volume in 5-minute sampling period. Then, the data of one day preserver 288 traffic volume data points (24 hours per day and 12 data points per hours). Fig. 6 shows the traffic volume data on 21 workdays in March 2014.

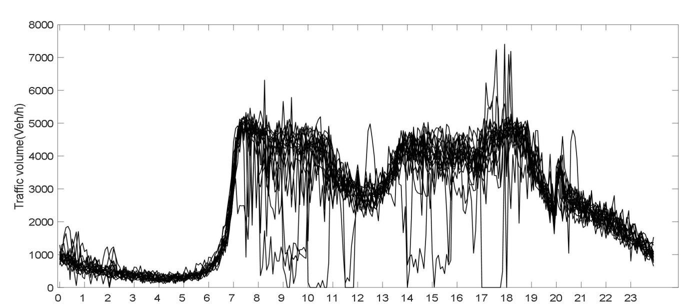


Fig. 6. Traffic volume on 21 workdays in March 2014.

A. *Determining CoS*

To describe the fluctuation range of traffic flow data and convert data into intervals, this paper applies central limit theorem for traffic flow data preprocessing with 90% confidence level. Confidence interval [1278, 1550] of sampling period 15:01-15:05 means there is 90% possibility for the real value to fall in the interval. Table I shows all intervals in 15:00-15:55.

TABLE I
CONFIDENCE INTERVALS OF 15:01-15:55

Time	Interval left end-point	Interval right end-point
15:00-15:05	1278	1550
15:06-15:10	1269	1540
15:11-15:15	1232	1516
15:16-15:20	1202	1486
15:21-15:25	1208	1516
15:26-15:30	1213	1511
15:31-15:35	1215	1460
15:36-15:40	1219	1461
15:41-15:45	1225	1464
15:46-15:50	1189	1362
15:51-15:55	1179	1353

Feature scaling is applied to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Here the range is set 0 to 5 according the actual value. As shown in Table II, right endpoints of confidence intervals after this part should fall into [0, 5]. Scale factor is equal to maximum value (right end-point at 15:00-15:05)/5 in this paper.

TABLE II
CONFIDENCE INTERVALS ON A CONTINUOUS SCALE

Time	Interval left end-point	Interval right end-point
15:00-15:05	4.121	5.000
15:06-15:10	4.094	4.966
15:11-15:15	3.972	4.889
15:16-15:20	3.876	4.794
15:21-15:25	3.897	4.890
15:26-15:30	3.913	4.874
15:31-15:35	3.918	4.707
15:36-15:40	3.932	4.711
15:41-15:45	3.950	4.723
15:46-15:50	3.836	4.392
15:51-15:55	3.804	4.365

Each confidence interval has a corresponding type-1 fuzzy sets model. Applying union operation to three type-1 fuzzy sets of $x(k-1), x(k-2), and \quad x(k-3)$, *CoS* of $x(k)$ is obtained. Fig.7 demonstrates *CoS* of 15:00-15:15, 15:06-15:20, 15:11-15:25, 15:16-15:30, 15:21-15:35 and 15:26-15:40.

B. *LSTM Algorithm for Predicting Traffic Flow*

The LSTM prediction model consists of an input layer with 288 nodes, a hidden layer with 18 nodes, and an output layer with 288 nodes, which is shown in Fig. 8. The initial weights $\mathbf{W_i}, \mathbf{W_o}, \mathbf{W_f}$ are randomly generated. For example, we need to predict the traffic flow at time t . The weights of the original LSTM input gate are updated in each iteration using back propagation algorithm. In this paper, we integrate the upper and lower boundary values of T2FSs generated by the previous 15-minute traffic flow into the weight update formula to further accurately calculate the scope of the data at time t , the weight updating formula is shown in (18). The training data we use in this paper is the first 17 workdays of March 2014.

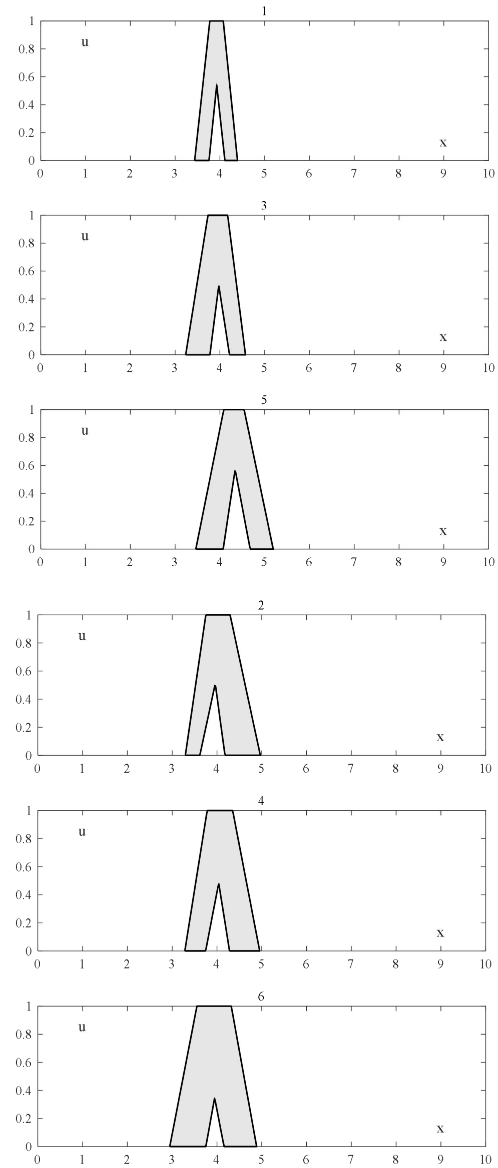


Fig. 7. Six *CoS* of 15-minutes.

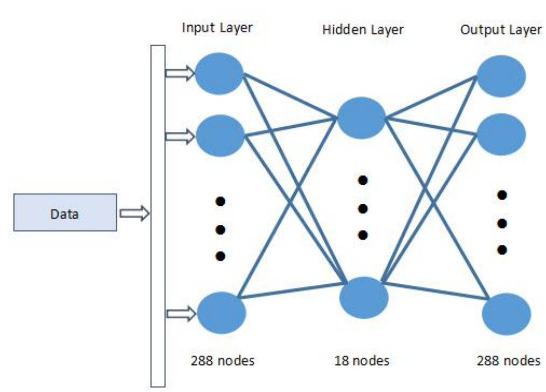


Fig. 8. LSTM prediction model structure.

C. *Simulation and Error Analysis*

We use the Mean Absolute Percentage Error(MAPE) and Mean Square Error(MSE) to evaluate the forecasting accuracy and compare it with models given by other researchers. The MAPE calculation formula is as formula (19) and the MSE calculation formula is as formula (20).

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\frac{\left|\hat{v}_k - v_k\right|}{v_k}\cdot 100\% \qquad (20)$$

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(v_k - \hat{v}_k)^2 \qquad (21)$$

where $\hat{v}_k$ is the predicted traffic volume at kth time period, v_k is the real traffic volume at kth time period and *N* is the number of time intervals one day.

Fig. 9 shows the differences between the typical LSTM prediction model and T2F-LSTM model by prediction data on March 26th. Fig.9 (a) shows the prediction result of typical LSTM model and Fig.9 (b) shows the prediction result of T2F-LSTM model. The red curve is the forecasting data, and the black curve represents the real data of traffic volume. The MAPE of the forecasting data is 8.14% for typical LSTM prediction model and the MAPE of the T2F-LSTM model is 4.02%. Fig. 10 and Fig.11 shows the T2F-LSTM model by prediction data on March 27th and 28th.

Table III shows the detail MAPE and MSE value of forecasting result by the T2F-LSTM model for several days. Then, we conduct simulations to compare the prediction performance of our method with the other existing deep learning methods including BP Neural Network (BPNN), Deep Neural Network (DNN), random forest with the same data set shown as Table IV.

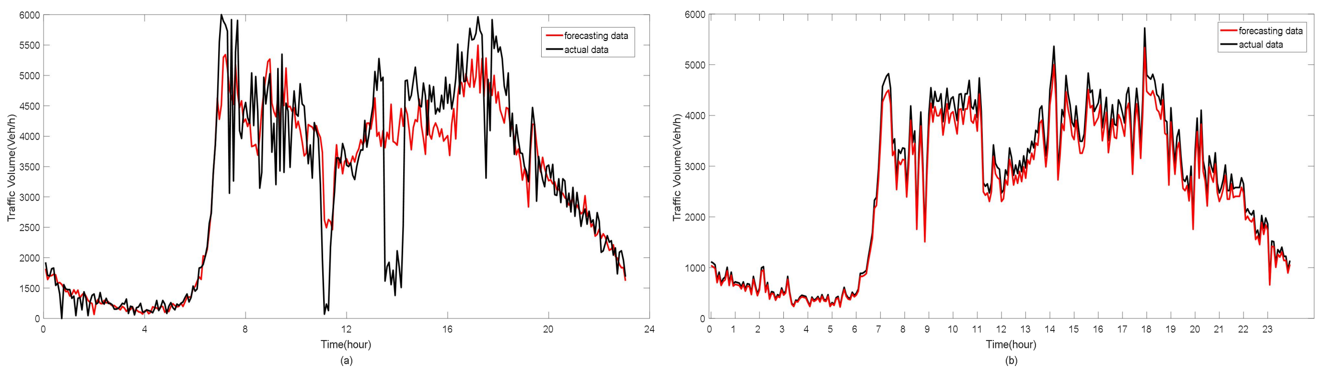
TABLE III
PERFORMANCE OF TRAFFIC VOLUME PREDICTION OF THE TYPICAL LSTM MODEL AND T2F-LSTM MODEL

Forecast model	March 26	March27	March28	Mean
T2F-LSTM	MAPE (%)			
	4.02	3.67	3.33	3.67
	MSE			
	15664	15210	14756	15210
LSTM	MAPE (%)			
	8.31	7.88	8.23	8.14
	MSE			
	19888	17883	19507	19092

Compared to the typical LSTM, the MAPE has been reduced by more than half. In the meantime, by adding motivational factors in input gate weight, the convergence speed of training set is almost doubled. Typical LSTM takes 1125.8seconds, and T2F-LSTM takes only 725seconds when sets the same 100 iterations.

TABLE IV
TRAFFIC FLOW PREDICTION PERFORMANCE OF DIFFERENT DEEP LEARNING METHODS

Algorithm	BPN	DNN	Random Forest	LSTM	T2F-LSTM
MAPE(%)	9.83	8.96	14.06	8.14	3.67
MSE	27829	23121	56933	20255	15468



(a) Typical LSTM prediction model (b) Type-2 fuzzy LSTM prediction model

Fig. 9. Forecasting result of typical LSTM model and type-2 fuzzy LSTM model on March 26th.

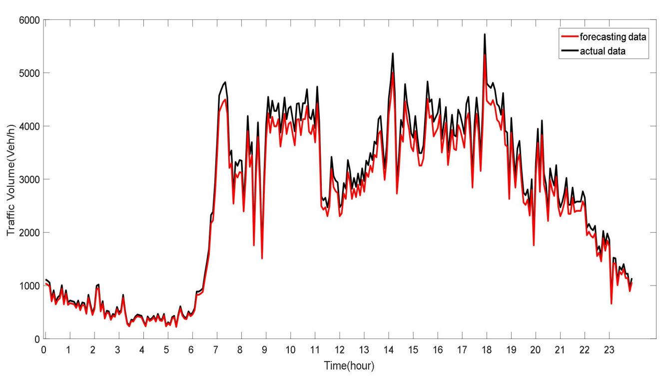


Fig. 10. T2F-LSTM model forecasting data on March 27th.

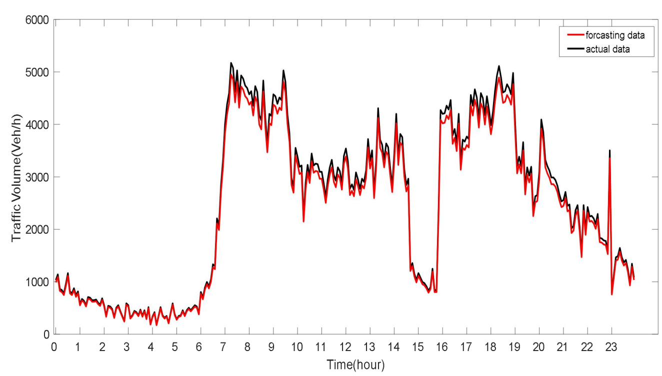


Fig. 11. T2F-LSTM model forecasting data on March28th.

V. Conclusion

In this paper, a T2F-LSTM traffic volume forecasting model is proposed. Simulation results indicate that the model gives full play to the ability of type-2 fuzzy set theory to deal with uncertainty data and the ability of LSTM to implement prediction accuracy. The confidence interval is used to transform point data into interval data which describe the fluctuation characteristic of traffic volume more properly, and lay the data foundation for T2FSs. The spatial-temporal feature of traffic flow are extracted by union operation of several embedded type-1 fuzzy which are from traffic volume in adjacent sampling period. Meanwhile, *CoS* information is added to the weight update formula of the input gate of LSTM neural network as motivational factors on the activation function output to improve the network interpretability and computational expense. The proposed approach encourages activation function outputs to satisfy a target pattern with the weight have the ability of learning the fluctuation characteristic of data more accurately. The experiment results indicate that the proposed models can provide precise traffic volume prediction results.

In the future study, based on existing methods on how to generate type-2 fuzzy set from type-1 fuzzy sets, different types of membership functions should be discussed.

References

[1] Yuan J, Zheng Y, Xie X et al., "Driving with knowledge from the physical world," *17th ACM/SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp.316–324, 2011.

[2] Hou Z S, Li X Y, "Repeatability and similarity of freeway traffic flow and long-term prediction under big data," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 6, pp. 1786–1796, 2016.

[3] S. V. Kumar, L. Vanajakshi, "Short-term traffic flow prediction using seasonal Arima model with limited input data," European Transport Research Review, vol.7, no.3, pp.1–9, 2015.

[4] Guo J H, Huang W, Williams et al., "Adaptive kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol.43, pp.50–64, 2014.

[5] Wang J, Shi Q X, "Short-term traffic speed forecasting hybrid model based on chaos-wavelet analysis-support vector machine theory," *Transportation Research Part C: Emerging Technologies*, vol.27, pp.219–232, 2013.

[6] Wu Y K, Tan H C, J. Peter et al., "Short-term traffic flow prediction based on multilinear analysis and k-nearest neighbor regression," *15th COTA International Conference of Transportation Professionals*, 2015.

[7] Zhang Y R, A. Haghani, "A gradient boosting method to improve travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol.58, pp.308–324, 2015.

[8] P. Lopez-Garcia, E. Onieva, E. Osaba et al., "A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy," *IEEE Transactions on Intelligence Transportation Systems*, vol.17, no.2, pp.557–569, 2016.

[9] I. Laña, J. D. Ser , I. I. Olabarrieta, "Understanding daily mobility patterns in urban road networks using traffic flow analytics," *Network Operations & Management Symposium*, pp.1157-1162, 2016.

[10]T. Thomas, W. Weijermars, E. V. Berkum, "Predictions of urban volumes in single time series," *IEEE Transactions on Intelligent Transportation Systems*, vol.11, no.1, pp.71-80, 2010.

[11] I. Lana, J. L. Lobo, E. Capecci et al., "Adaptive long-term traffic state estimation with evolving spiking neural networks", *Transportation Research Part C*, 101, pp.126-144, 2019.

[12] Jiang X, H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *Journal of Transportation Engineering*, vol. 131, no. 10, pp. 771-779, 2005.

[13] Hou Y, P. Edara, Sun C, "Traffic flow forecasting for urban work zones," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1761-1770, 2015.

[14] G. E. Hinton, R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol.13, no. 5786, pp. 504–507, 2006.

[15] Wu Y K, Tan H C, Qin L Q, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C*, vol. 90, pp. 166-180, 2018.

[16] Lv Y S, Duan Y J, Kang W W et al., "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865-873, 2015.

[17] Yang H F, T. S. Dillon, Chen Y P P, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2371-2381, 2017.

[18] Y. Adi, J. Keshet et al., "Sequence segmentation using joint RNN and structured prediction models," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2422-2426, 2017.

[19] Z. Zhene, P. Hao et al., "Multiple RNN method to prediction human action with sensor data," *2017 International Conference on Virtual Reality and Visualization (ICVRV)*, pp. 419-420, 2017.

[20] Z. Zhene, P. Hao et al., "Deep convolutional mesh RNN for urban traffic passenger flows prediction," *2018 IEEE SmartWorld*, pp. 1305–1310, 2018.

[21] Ma X L, Tao Z M, Wang Y H, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[22] Liu Y P, Zheng H F, Feng X X, "Short-term traffic flow prediction with Conv-LSTM," *International Conference on Wireless Communications & Signal Processing*, pp. 1-6, 2017.

[23] Liu Y D, Wang Y Z , Yang X G et al., "Short-term travel time prediction by deep learning: A comparison of different LSTM-DNN models," *IEEE 20th International Conference on Intelligent Transportation Systems*, pp. 1-8, 2017.

[24] Liu H, Shi X M, Guo D M et al., "Feature selection combined with neural network structure optimization for HIV-1 protease cleavage site prediction," *Biomed Research International*, vol. 2015, no. 1, pp. 263-586, 2015.

[25] Zhen X, G. Panoutsos, "Interpretable machine learning: convolutional network with RBF fuzzy logic classification rules," *2018 International Conference on Intelligent Systems*, pp. 448-454, 2018.

[26] Chai Q, Michel P, Bernard B S L, "POP-TRAFFIC: a novel fuzzy neural approach to road traffic analysis and prediction," *IEEE Transactions on ITS*,vol.7, no. 2,pp.133-146, 2006.

[27] Li R M, Jiang C Y, Zhu F H, "Traffic flow data forecasting based on Interval Type-2 Fuzzy Sets theory," *Journal of Automatica Sinica*, vol. 2, no. 2, pp. 141-148, 2016.

[28] Li R M, Huang Y F, Wang J, "Long-term traffic volume prediction based on type-2 fuzzy sets with confidence interval method," I*nternational Journal of Fuzzy Systems*, vol. 21, no. 7, pp. 2120-2131, 2019.

[29] Li R M, Huang Y F and Wang J, "Long-term traffic volume prediction based on K-means Gaussian interval type-2 fuzzy sets," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1344-1351, 2019.

[30] A. Segatori, F. Marcelloni, W. Pedrycz, "On distributed fuzzy decision trees for big data," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 174-192, 2018.

[31] Zhao H, Wang P, Hu Q H, "Fuzzy rough set based feature selection for large-scale hierarchical classification," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 10, pp. 1891-1903, 2018.

[32] Abigogun, Olusola and Adeniyi, "Data mining, fraud detection and mobile telecommunications: call pattern analysis with unsupervised neural networks master thesis," *University of the Western Cape*, 2005.

[33] Deng Y, Ren Z Q, Kong Y Y et al., "A hierarchical fused fuzzy deep neural network for data classification," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 1006-1012, 2017.

[34] L. Zadeh, "The concept of a linguistic variable and its application to approximate reason-1," *Information Sciences*, pp. 199-249, 1975.

[35] M. Mizumoto, K. Tanaka, "Some properties of fuzzy sets of type-2," *Information and Control*, vol. 31, no. 4, pp. 312-340, 1976.

[36] Li Y M, Huang J, "Type-2 fuzzy mathematical modeling and analysis of the dynamical behaviors of complex ecosystems," *Simulation Modelling Practice and Theory*, vol. 16, no. 9, pp. 1379-1391, 2008.

[37] S. Greenfield, F. Chiclana, "Defuzzification of the discretized generalized type-2 fuzzy set: Experimental evaluation," *Information Sciences*, vol. 244, no. 7, pp. 1-25, 2013.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.2986995, IEEE Transactions on Fuzzy Systems

[38] Liu F, Mendel J M. "Encoding words into interval type-2 fuzzy sets using an interval approach", *IEEE transactions on fuzzy systems*, vol. 16, no.6,pp.1503-1521,2008.
[39] Wu D, Mendel J M, Coupland S. "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis". *IEEE transactions on fuzzy systems*,vol. 20,no.3,pp.499–513,2012.
[40] N. N. Karnik, J. M. Mendel, "Applications of type-2 fuzzy logic systems to forecasting of time-series," *Information Sciences*, vol. 120, no. 1-4, pp. 89-111, 1999.
[41] R. Aliev, W. Pedrycz, B. Guirimov et al., "Type-2 fuzzy nueral networks with fuzzy clustering and differential evolution optimization," *Information Sciences*, vol. 181, no. 9, pp. 1591-1608, 2011.
[42] B. Choi, F. Rhee, "Interval type-2 fuzzy membership function generation methods for pattern recognition," *Information Sciences*, vol. 179, no. 13, pp. 2102-2122, 2009.
[43] Wu D R, Tan W W, "A simplied type-2 fuzzy logic controller for real-time control," *ISA Transactions*, vol. 45, no. 4, pp. 03-516, 2006.
[44] T. Dereli, A. Baykasoglu, K. Altun et al., "Industrial applications of type-2 fuzzy sets and systems: A concise review," *Computers in Industry*, vol. 62, no. 2, pp. 125-137, 2011.
[45] C. Leal-Ramirez, O. Castillo, P. Melin et al., "Simulation of the bird age-structured population growth based on an interval type-2 fuzzy cellular structure," *Information Sciences*, vol. 181, no. 3, pp. 519-535, 2011.
[46] S. Chakravarty, P. K. Dash, "A PSO based integrated functional link net and interval type-2 fuzzy logic system for predicting stock market indices," *Applied Soft Computing*, vol. 12, no. 2, pp. 931-941, 2012.
[47] Mo H, Wang F-Y, Zhou M et al., "Footprint of uncertainty for type-2 fuzzy sets," *Information Sciences*, vol. 272, no. C, pp. 96-110, 2014.
[48] Jang J S R, "ANIFS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems Man and Cybernetics*, vol 23, no. 3, pp.664-685, 1993.
[49] Smith B L, Demetsky M J, "Traffic Flow Forecasting: Comparison of Modeling Approaches," *Journal of Transportation Engineering*, vol. 123, no.4,pp.261-266, 1997.
[50] Chai Q, Michel P, Bernard B S L, "POP-TRAFFIC: a novel fuzzy neural approach to road traffic analysis and prediction," IEEE Transactions on ITS, vol. 7, no. 2, pp. 133-146, 2006.
[51] Li Z H, He D, Fei Tian, et al., "Towards Binary-Valued Gates for Robust LSTM Training," ICML 2018.

Runmei Li is an associate professor at Beijing Jiaotong University. She received Ph. D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, China, 2005. In July 2005, she started her career at the School of Electronic and Information Engineering of Beijing Jiaotong University as a teacher. Her research interests include Cooperative Vehicle-Road System, traffic data mining，Type-2 fuzzy sets theory and application.

Yongchao Hu received the B.S. degree in Rail Transportation Signal and Control from Beijing Jiaotong University, China, in 2018. He is currently working toward the M.S. degree at the Institute of Advanced Control System, Beijing Jiaotong University, Beijing, China. His research interests include traffic data mining，autonomous vehicle.

Qiuhong Liang is currently working toward the M.S. degree at the Advanced Control System Laboratory, Beijing Jiaotong University, Beijing, China. Her research interests include traffic flow prediction and traffic information processing.