Research Article

Short-Term Traffic Flow Prediction with Recurrent Mixture Density Network

Mingjian Chen (ID)**, Rui Chen, Fu Cai, Wanli Li, Naikun Guo, and Guangyun Li**

Institute of Geospatial Information, Information Engineering University, Zhengzhou 450000, China

Correspondence should be addressed to Mingjian Chen; zzcmj@139.com

Traffic situation awareness is the key factor for intelligent transportation systems (ITS) and smart city. Short-term traffic flow prediction is one of the challenging tasks of traffic situation awareness, which is useful for route planning, traffic congestion alleviation, emission reduction, and so on. Over the past few years, ubiquitous location acquisition techniques and sensors digitized the road networks and generated spatiotemporal data. Massive traffic data provide an opportunity for short-term traffic flow prediction in a data-driven manner. Most of the existing short-term traffic flow prediction methods can be divided into two categories: nonparametric and parametric. Traditional parametric methods failed to obtain accurate prediction, due to the nonlinear and stochastic characteristics of short-term traffic flow. Recently, deep learning methods have been studied widely in the fields of short-term prediction. These nonparametric methods yielded promising results in practical experiments. Motivated by the current study status, we dedicate this paper to a short-term traffic flow prediction approach based on the recurrent mixture density network, the combination of recurrent neural network (RNN), and mixture density network (MDN). This approach is implemented on real-world traffic flow data and demonstrates the prominent superiority. To the best of our knowledge, this is the first time that the recurrent mixture density network is applied to a real-world short-term traffic flow prediction task.

1. Introduction

Over the past few decades, the number of megalopolises has increased rapidly, especially in developing country as China. In the process of city development, projects in various aspects such as urban layout plan, environment protection, and traffic management are still tricky challenges [1]. Hence, great demand for smart city construction and intelligent transportation systems (ITS) emerges [2]. Short-term traffic flow prediction aims to provide future traffic information of road network, which is useful for both individual users and transportation management [3]. Timely and accurate traffic situation awareness contributes to a wide spectrum of applications, such as route planning, traffic congestion alleviation, and emission reduction. With the rapid development of sensors, communication techniques, and location acquisition techniques, such as Global Navigation Satellite System (GNSS) and 5G technology, massive multisource data including trajectory data and traffic flow data are generated and communicated by prevalent traffic information acquisition devices [4].

The evolution of traffic conditions has both spatial and temporal characteristics [5]. There is abundant potential knowledge about traffic conditions contained in traffic data, which implies the spatiotemporal patterns of traffic flow. Ubiquitous traffic data lead to the increasing advent of data-driven traffic flow prediction methods. Hence, it is feasible to predict the evolution trend of traffic flow via knowledge discovery. Predicting short-term traffic flow refers to estimating the traffic flow indices of a certain road or transverse section at one or more following time instances, based on the recent and current traffic flow data. Most of the existing traffic flow prediction methods can be categorized into parametric methods and nonparametric methods [6].

In the category of parametric methods, the prediction models maintain a fixed structure. The model parameters are

estimated by the empirical way based on historical data. Widely used parametric methods include the time-series model, Autoregressive Integrated Moving Average (ARIMA) model [7], and Kalman Filter [8]. There are three core parameters to determine the structure of ARIMA (p, d, q), where p denotes the autoregressive order, d denotes the integrated order, and q is the moving average polynomial order, respectively. ARIMA model converts the time-series data into stationary series without tendency and captures both the autocorrelation features and noise features of the stationary series. Kalman Filter based methods rely on transition matrix and dynamic equation. However, the short-term traffic flow is highly nonlinear and stochastic. The evolution of traffic flow is actually an uncertain event. Hence, traditional parametric methods are limited in capturing the inherent features of traffic flow and fail to achieve expected performance.

In order to improve prediction precision, plenty of literature focuses on applying nonparametric models to traffic flow prediction. Nonparametric methods are mainly based on machine learning and deep learning technologies, which do not involve predefined parameters and structure [9]. Deep neural networks (DNN) [10], including convolutional neural network (CNN) [11] and recurrent neural network (RNN) [12], have achieved remarkable performance in a wide range of fields such as computer vision (CV) and natural language processing (NLP). The rapid development of deep learning technologies brings opportunities to improve the precision of traffic flow prediction more efficiently. RNN has the ability to learn historical information and contextual dependency from traffic flow series. Furthermore, RNN excels at fitting nonlinear time series by virtue of the nonlinear activation units. Actually, the concrete value of traffic flow indices is nonrepetitive and usually has a certain distribution. The way to predict concrete values does not match with the fact that the future evolution of traffic flow is an uncertain event. Hence, the methods introduced above regarding the anticipated traffic flow indices as deterministic results could be confused by the continuous and non-repetitive values.

Motivated by the current study status, we dedicate this paper to a short-term traffic flow prediction approach based on the recurrent mixture density network. The recurrent mixture density network is constructed by integrating long short-term memory (LSTM) network and mixture density network (MDN) [13]. LSTM network is a variant of RNN with a special unit structure, which is competent for learning historical information and contextual dependency in time series and avoiding the issue of vanishing gradient. The main purpose of introducing MDN is to parameterize mixture distributions by the outputs of the neural network. In this way, the prediction model generates the probable distributions of anticipated traffic flow indices instead of concrete values. The ultimate prediction results can be obtained by sampling from the distributions. The manner of fuzzy prediction accords with the actual behavior of traffic flow evolution. To the best of our knowledge, this is the first time that the recurrent mixture density network is applied to a real-world short-term traffic flow prediction task. Additionally, the recurrent mixture density model is easy to be transplanted to different application scenarios, due to the prominent advantage of generalization and learning ability.

The prediction model is implemented on a real-world traffic flow prediction task. Different from most of the other existing work, which predicts the value of only one certain traffic flow index of next one time step at once, we predict both the traffic time indices (TTI) and the average speed (AS) of a certain road section at several following time steps simultaneously. Since it is reasonable to consider the coherent relationship between TTI and AS, predicting traffic flow at more than one time step also provides more valuable information for traffic situation awareness. The remainder of this paper is organized as follows: the previous work related to short-term traffic flow prediction is reviewed in Section 2; Section 3 elaborates the prediction model; Section 4 presents the experimental results and the corresponding evaluation; and, finally, the discussion and conclusions are presented in Section 5.

2. Related Work

The manner of short-term traffic flow prediction methods can be concluded as follows: build a mathematic model based on history series data and then use the model to predict the future value of traffic flow indices. In general, the existing prediction methods can be categorized into two groups: parametric methods and nonparametric methods, according to the way of modeling.

In the category of parametric methods, the model parameters are estimated by the empirical way based on historical data. ARIMA model is a typical parametric model. As early as 1979, ARIMA has been applied to represent freeway traffic time-series data and make forecasts one time-interval in advance [14]. In recent years, various ARIMA models and the ARIMA-based models have been proposed. A vector autoregressive model was found appropriate and better than the traditional ARIMA model in practical experiments [15], which introduced the traffic flow information upstream as well as downstream. The ARIMA model and its variants such as autoregressive moving average (ARMA) and seasonal ARIMA (SARIMA) are also studied comparatively with other widely used machine learning methods, to evaluate the efficiency in traffic flow prediction tasks [16, 17].

However, short-term traffic flow is difficult to predict due to the nonlinear and stochastic feature. The performance of parametric methods is limited. Plenty of existing work attempts to develop nonparametric methods with flexible structure and parameters. Nonparametric methods mainly include machine learning and deep learning techniques [18]. RNN and its variants such as LSTM and gated recurrent units (GRU) network are widely used in short-term traffic flow prediction tasks. One of the advantages of the LSTM network is that the format of the input data is flexible. Hence, in order to improve the prediction performance, several traffic flow indices such as flow, speed, and occupancy can be used as

input simultaneously [19]. LSTM network and GRU network were also evaluated comparatively with ARIMA. The experiments demonstrated that the RNN based methods perform better than the ARIMA model [20]. Experiments including several prevailing parametric and nonparametric algorithms also suggest the excellent prediction ability of the LSTM network [21]. Experimental results of multiple prediction tasks with different time intervals verified the generalization ability of the LSTM network [22].

Additionally, other deep learning techniques are also widely used for traffic flow prediction. CNN is an excellent technique in image recognition fields, which is utilized to learn traffic flow as an image and make predictions [23]. A deep architecture combined by CNN and LSTM is developed and implemented on traffic flow forecast [24]. Stacked autoencoders (SAE) also perform better than other machine learning techniques [25].

In this work, we construct the recurrent mixture density network and apply it to a real-world short-term traffic flow prediction task, taking advantage of the LSTM network and meanwhile integrating the parameterization function of the MDN.

3. Methodology

The goal of short-term traffic flow prediction can be described as follows: given the recent part of traffic flow indices series $\{x_j, x_{j-1}, x_{j-2}, \ldots, x_{j-k+1}\}$ of a certain road section at the present time j and $k-1$ previous time stamps, where $x = [\text{TTI}, \text{AS}]$ is a two-dimensional vector containing the traffic time indices and the average speed information, predict the traffic flow indices $\{x_{j+1}, x_{j+2}, \ldots, x_{j+h}\}$ at next h time steps.

The recurrent mixture density network and data processing process will be elaborated in this section as follows.

3.1. Recurrent Mixture Density Network. The recurrent mixture density network is constructed by integrating the LSTM network and MDN, which has the ability of learning historical information and contextual dependency and generating predicted mixture distributions. Unlike the ordinary feedforward neural network that only has connections between adjacent layers, the LSTM network allows recurrent connections between adjacent cells of the same layer. Hence, the LSTM network is deep in time-series dimension and has the ability to map the historical input series to the output via recurrent connections. The most distinct characteristic of the LSTM network is the unique gate structure of the LSTM cell including input gate, forget gate, and output gate as shown in Figure 1.

The three gates work collaboratively to encode the input signal and generate outputs, which enable the LSTM network to maintain historical information contained in series data. The LSTM network encodes the input signal as follows:
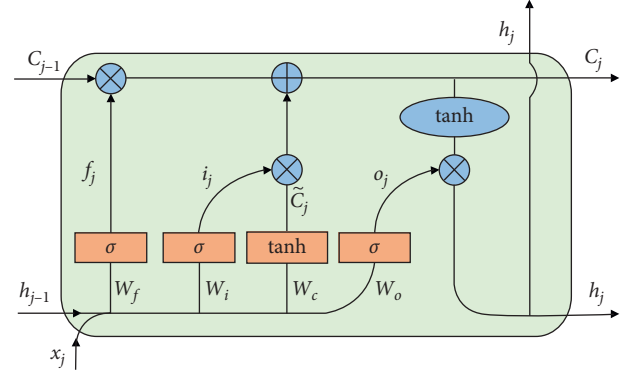

FIGURE 1: The architecture of the long short-term memory (LSTM) unit.

$$
\begin{aligned}
f_j &= \sigma\left(W_f\left[x_j, h_{j-1}\right] + b_f\right), \\
i_j &= \sigma\left(W_i\left[x_j, h_{j-1}\right] + b_i\right), \\
o_j &= \sigma\left(W_o\left[x_j, h_{j-1}\right] + b_o\right), \\
\tilde{C}_j &= \tanh\left(W_c\left[x_j, h_{j-1}\right] + b_c\right), \\
C_j &= f_j \cdot C_{j-1} + i_j \cdot \tilde{C}_j, \\
h_j &= o_j \cdot \tanh\left(C_j\right),
\end{aligned}
\tag{1}
$$

where f_j, i_j, and o_j denote the forget gate, input gate, and output gate, respectively. σ and tanh denote the logistic sigmoid function and hyperbolic tangent function, respectively, both of which are used as the nonlinear activation function. h_{j-1} is the output of the $j-1$ th unit in the LSTM layer. x_j denotes the input signal from the preceding layer. The weight matrices W_f, W_i, W_o and the corresponding bias b associated with these gates are variables to be optimized during the training process. LSTM cells retain and update the cell state while receiving an input signal. Then, the updated cell state is activated by the nonlinear activation function and fed to the next layer.

Outputs of all the LSTM layers are concatenated as the integrated output of hidden layers and sent to the MDN layer, as shown in Figure 2. The MDN layer is a full-connected layer as follows:

$$
\tilde{y}_j = \sum_{n=1}^{N} W_{h^n y} \cdot h_j^n + b_y,
\tag{2}
$$

where N denotes the number of stacked LSTM layers and h_j^n is the output of the nth bidirectional LSTM layer. $W_{h^n y}$ denotes the weight matrix connecting the LSTM layer and MDN layer, b_y is the corresponding bias, $\tilde{y}_j$ denotes the outputs of the MDN layer. The outputs of the MDN layer are used to parameterize mixture distributions. In order to capture the probable range of the two-dimensional vector $x = [\text{TTI}, \text{AS}]$ simultaneously, a bivariate Gaussian mixture model is adopted as the mixture distributions. Individual components of the mixture distributions are denoted by a
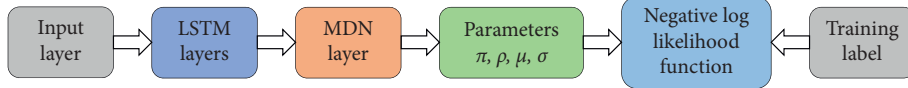
FIGURE 2: Schematic of recurrent mixture density network.

bivariate Gaussian model, which are parameterized by sets of means μ, standard deviations σ, and correlations ρ. Additionally, the Gaussian mixture model also contains the weight parameter π:

$$\widetilde{y}_j = \left\{\widetilde{\pi}_j^m, \widetilde{\rho}_j^m, \widetilde{\mu}_j^m, \widetilde{\sigma}_j^m\right\}_{m=1}^M, \tag{3}$$

where M denotes the number of individual components composed of the mixture distributions.

All the outputs generated by the network must be normalized to meaningful ranges as the final output before being used as parameters:

$$y_j = \left\{\pi_j^m, \rho_j^m, \mu_j^m, \sigma_j^m\right\}_{m=1}^M, \tag{4}$$

where

$$\pi_j^m = \text{Soft max}\left(\widetilde{\pi}_j^m\right) = \frac{\exp\left(\widetilde{\pi}_j^m\right)}{\sum_{m=1}^M \exp\left(\widetilde{\pi}_j^m\right)},$$

$$\rho_j^m = \tanh\left(\widetilde{\rho}_j^m\right), \tag{5}$$

$$\mu_j^m = \widetilde{\mu}_j^m,$$

$$\sigma_j^m = \exp\left(\widetilde{\sigma}_j^m\right).$$

The probability density function of the train label x_{j+l} ($l \geq 1$) against the output mixture distributions is defined as follows:

$$\Pr\left(x_{j+l} \mid y_j\right) = \sum_m^M \pi_j^m G\left(x_{j+l} \mid \mu_j^m, \sigma_j^m, \rho_j^m\right), \tag{6}$$

where G is the bivariate Gaussian function:

$$G(x|\mu, \sigma, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-Z}{2\left(1-\rho^2\right)}\right],$$

$$Z = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2}. \tag{7}$$

The negative log-likelihood function is employed as the loss function to be optimized in the training process:

$$\text{Loss} = -\log\left(\sum_{m=1}^M \pi_i^m G\left(x_{i+l}|\mu_i^m, \sigma_i^m, \rho_i^m\right)\right). \tag{8}$$

Adam optimizer is applied to the backpropagation through time process to fine-tune the network variants. We also adopt the dropout technique and gradient clip strategy to avoid the overfitting and exploding gradient issues.

3.2. Data Processing.

In order to predict the short-term traffic flow indices $x = [\text{TTI}, \text{AS}]$ at several following time steps, we prepare both the training input data and training label as sequence. All the training data are generated through sliding window strategy [13], as shown in Figure 3, where we hypothesized the input sequence length $k = 3$ and prediction length $h = 2$. The train label is in the same sequence length with the train input and lags $h = 2$ time steps. In the prediction stage, the prediction model also outputs the mixture distributions of $k = 3$ vectors, whereas we only pick out the last $h = 2$ mixture distributions for sampling. The ultimate prediction results are obtained by sampling from mixture distributions via roulette strategy based on the weights of each individual Gaussian component.

4. Experiments

In this section, we aim to complete a practical short-term traffic flow prediction task by utilizing the recurrent mixture density network. The optimal parameter configuration of the prediction model is determined after several experiments. In order to assess the performance of the prediction model, two widely studied methods are introduced for comparison.

4.1. Dataset and Experiment Setup.

Shenzhen is one of the first-tier cities in China, where the economy developed rapidly in recent years. As a megalopolis, Shenzhen is faced with typical city development challenges, such as transportation management. Shenzhen North Railway Station is one of the most important transportation hubs of the city. The road networks around Shenzhen North Railway Station are also the main transportation corridors of the city, which directly influence the urban traffic efficiency. The traffic flow dataset is generated by sensors installed on road networks which record the speed and traffic flow information of mobile vehicles. It contains TTI and AS information of twelve road sections around Shenzhen North Railway Station in two periods from 1 January 2019 to 31 March 2019 and from 1 October 2019 to 21 December 2019. The time granularity is ten minutes; that is, records at each timestamp denote the average indices in ten minutes. Figure 4 presents an example of typical traffic flow patterns overtime in a day of a certain road section.

In experiments, we select the records from 1 December 2019 to 21 December as test data and use the rest for model training. We use one-hour-long traffic flow information to predict the indices in the upcoming half-hour; that is, the input length is 6 and the prediction length is 3.

Three metrics are employed for performance evaluation: mean absolute error (MAE), mean relative error (MRE), and root mean square error (RMSE):
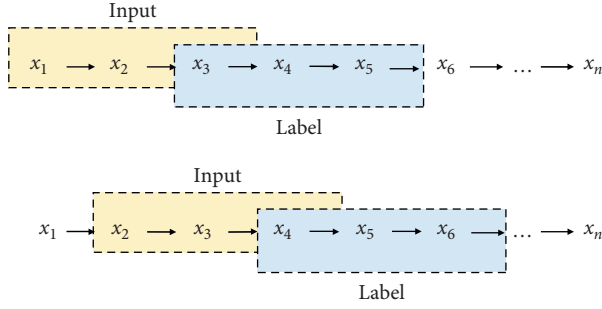
FIGURE 3: Schematic of generating the training data when input sequence length $k = 3$ and prediction length $h = 2$.
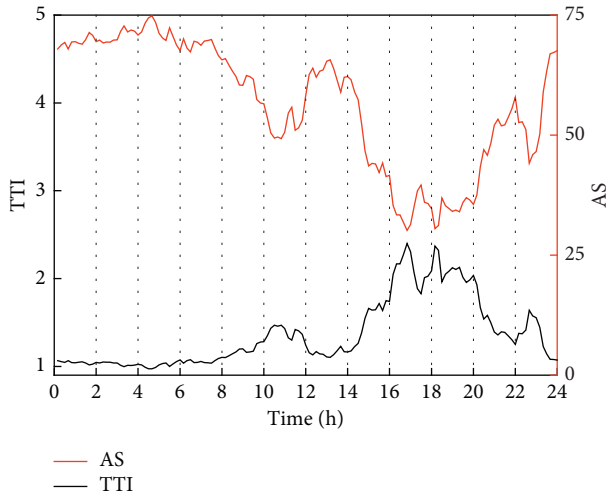

FIGURE 4: The patterns of traffic time indices (TTI) and average speed (AS) of a certain road section on 27 March 2019.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |x_i' - x_i|,$$

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|x_i' - x_i|}{x_i}, \qquad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} |x_i' - x_i|^2},$$

where x_i' and x_i denote the prediction result and the ground truth, respectively. n is the total number of test traffic flow segments. The prediction errors are calculated on every predicted time instance.

4.2. Comparative Results. Two widely used methods are involved in comparison experiments, including the LSTM network and ARIMA model. The optimal parameter configuration of the recurrent mixture density network is determined after several experiments, where the number LSTM layer is set to 4, the number of LSTM units in each layer is set to 256, and the number of mixture components in the mixture density layer is set to 15. In comparison experiments, the structure of the LSTM network is the same as the LSTM layers of the recurrent mixture density network. Three parameters of the ARIMA model (p, d, q) are determined as $(1, 1, 1)$ after several experiments. It must be noted that the recurrent mixture density network and LSTM network predict TTI and AS simultaneously, while the ARIMA model processes the two indices separately.

Figure 5 depicts the performance evaluation of the model in the training process. The model training process is completed in the vicinity of the 40th epoch according to the descending curve of the loss function.

The prediction performance comparison of the three methods is given in Table 1. The lowest errors of all metrics are highlighted in bold font, which evidently demonstrate the prominent performance of the recurrent mixture density network. In Table 1, the accuracy (1-MRE) of the recurrent mixture density network is over 0.92 at all the predicted timestamps for both TTI and AS prediction tasks. It is obvious that the prediction errors increase with the increase of prediction time steps. The performance of the recurrent mixture density network is promising in 10 min, 20 min, and 30 min prediction tasks. Compared to the results of the LSTM network, the better performance of the recurrent mixture density network proves the superiority of the fuzzy prediction strategy. The attempt of combining MDN and LSTM brings an accurate solution for the short-term traffic flow prediction issue. In 10 min prediction, the recurrent mixture density network improves the accuracy of the TTI prediction and AS prediction by 1.28% and 0.8%, respectively, against the ARIMA model, which verifies the prominent advantage of the nonparametric method.

The MAE is visualized in Figure 6, where we depict the cumulative distribution function (CDF) of mean absolute errors of 10 min, 20 min, and 30 min predicted results. It is clearly observed that the CDF curve of the recurrent mixture density network is above the curves of the other two methods. Figure 6 presents strong evidence that the overall performance of the recurrent mixture density network is superior to the LSTM network and ARIMA model. In the TTI prediction task, there are over 90% of the test data with prediction error less than 0.2, over 80% of the test data with prediction error less than 0.1. For the prediction of AS, the prediction error of more than 92% of the test data is under 6 km/h and more than 76% is under 3 km/h. It is notable that the proportion of test data with low prediction error in LSTM experiments and ARIMA experiments are close to that in the recurrent mixture density network experiments. While there are more test data with large errors in the LSTM experiments and ARIMA experiments, The CDF of error implies that, in most tests, all the three methods yield promising results, whereas, in some unexpected scenario such as confronting traffic congestion, the recurrent mixture density network is more reliable.

The prediction results of the recurrent mixture density network on two different road sections are demonstrated in Figure 7, which are the typical instances of heavy traffic congestion and low traffic volume, respectively. The observed TTI and AS are also depicted for comparison, which are clearly coherent in the periods of all the 21 days. It is
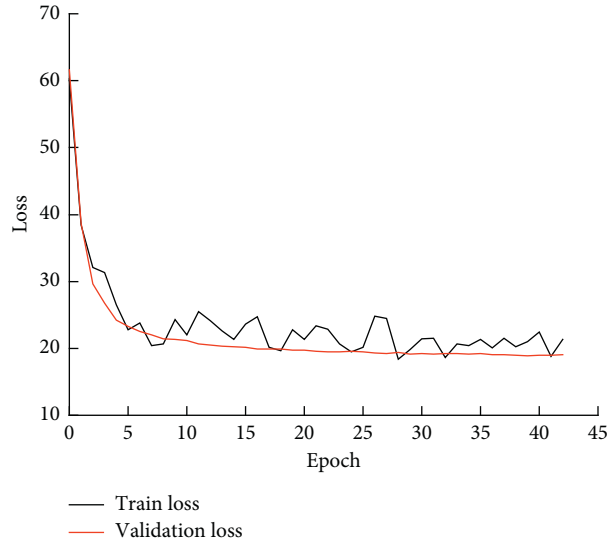
FIGURE 5: Loss function evaluation in the training process.

TABLE 1: Comparison of prediction performance.

| Time steps	MAE			MRE			RMSE		
	RMDN	LSTM	ARIMA	RMDN	LSTM	ARIMA	RMDN	LSTM	ARIMA
TTI									
1	0.0587	0.0705	0.0781	0.0376	0.0433	0.0504	0.1424	0.1630	0.1811
2	0.0923	0.1071	0.1115	0.0582	0.0650	0.0705	0.2180	0.2531	0.2587
3	0.1139	0.1261	0.1357	0.0718	0.0747	0.0857	0.2627	0.2998	0.3122
AS									
4	1.7141	1.8088	1.9774	0.0415	0.0463	0.0495	2.4898	2.8266	3.1717
5	2.5568	2.7001	2.7395	0.0644	0.0714	0.0708	3.8673	4.5994	4.5611
6	3.1055	3.2512	3.3361	0.0790	0.0880	0.0869	4.7613	5.6278	5.6286

(a)

(b)
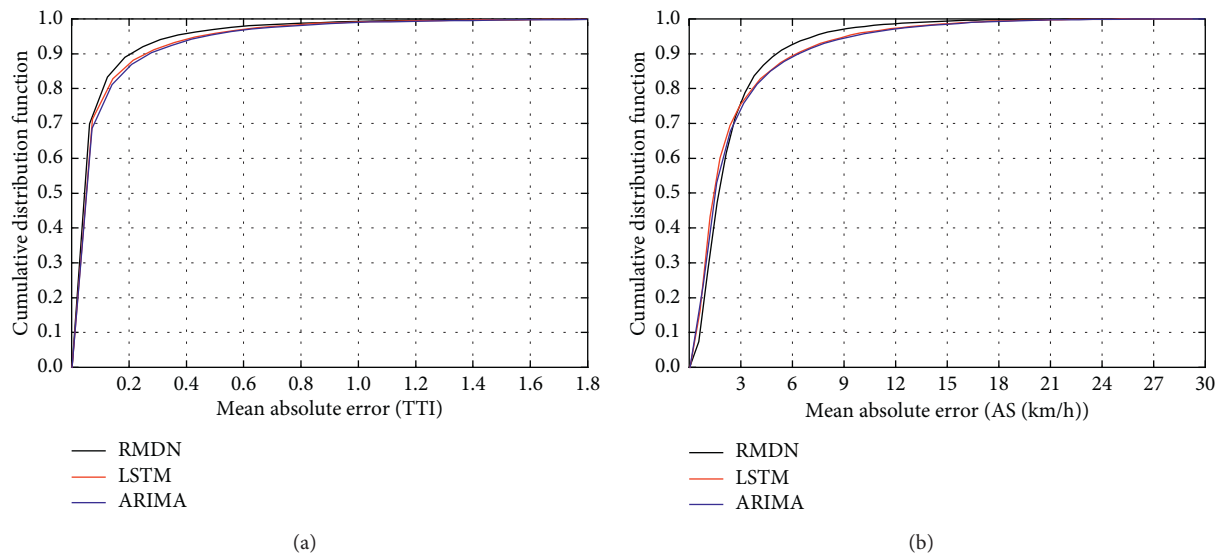
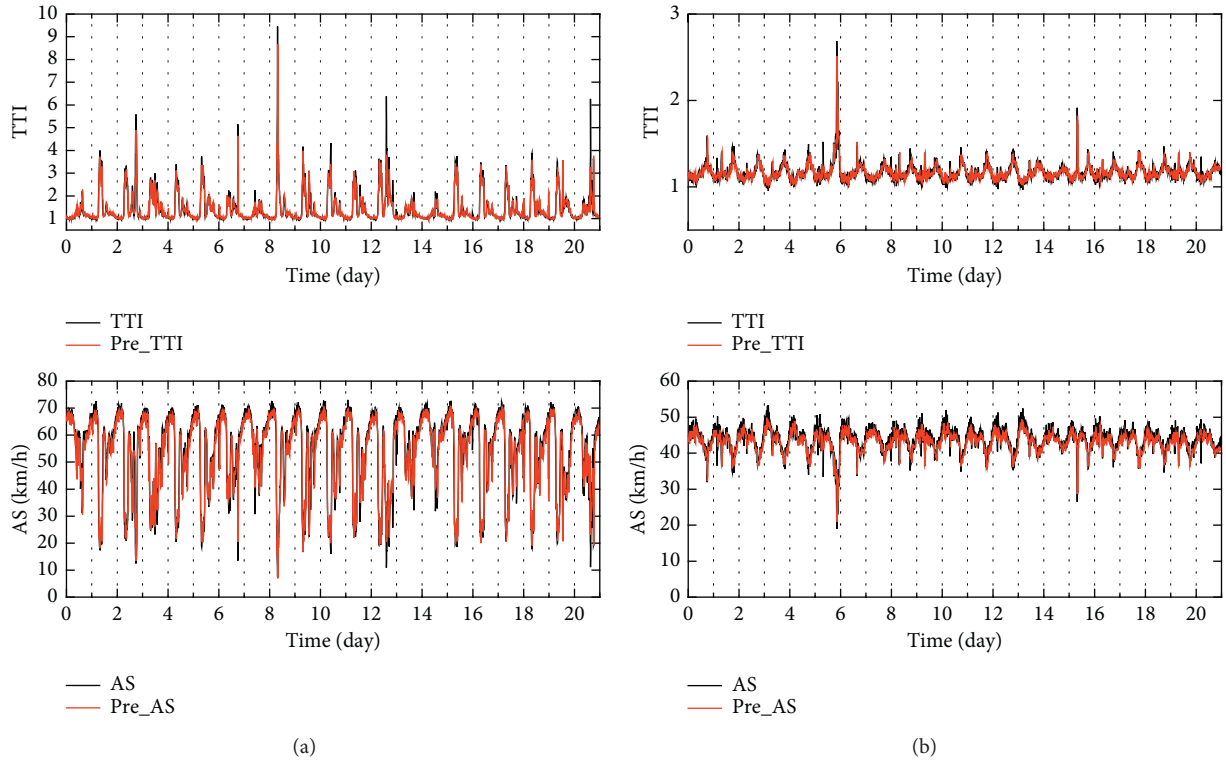FIGURE 6: Cumulative distribution function (CDF) of the mean absolute error. (a) TTI. (b) AS.

Figure 7: Short-term traffic flow prediction on different road sections. (a) Road section with heavy traffic congestion. (b) Road section with slight traffic congestion.
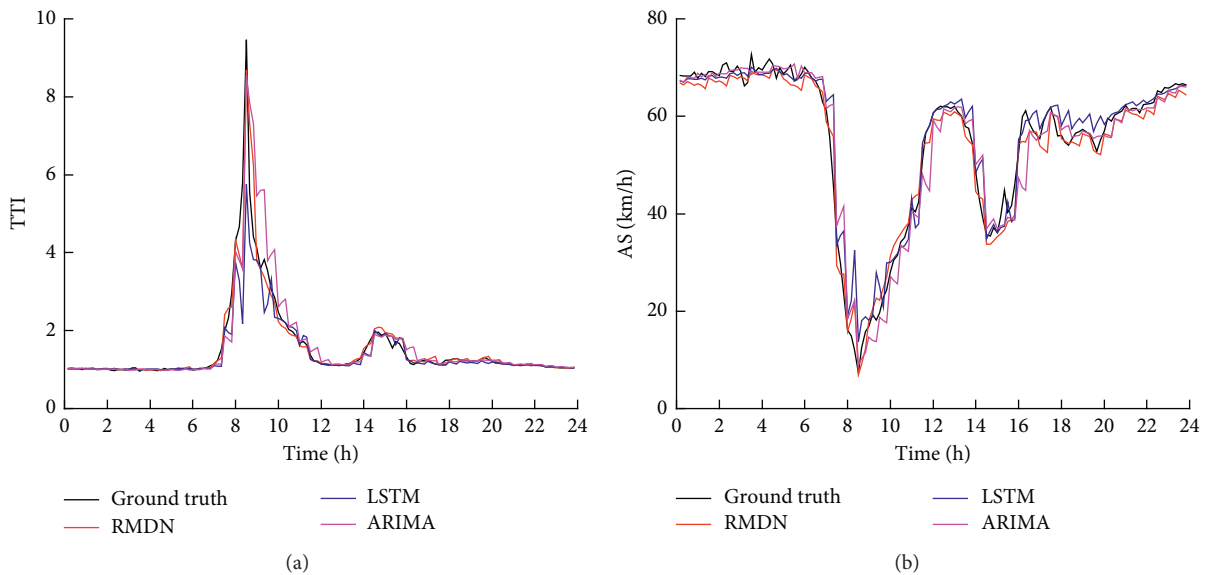

Figure 8: Comparison of prediction results of three methods on a certain road section on 9 December 2019.

shown that the predicted results match well in both two scenarios. However, the recurrent mixture density network performs better when the traffic condition is fluent. Although the periodical pattern of the traffic flow indices is evident, there are still some abnormal values, especially in the road section with heavy traffic congestion, where the

traffic condition could be extremely terrible in rush hour. It is challenging to predict the traffic congestion with a large TTI value and low AS value; even so, most of the mutations are matched well as shown in Figure 7.

Figure 8 presents the comparison of prediction results of three methods on a certain road section on 9 December

2019, where a typical traffic congestion phenomenon occurs. In the period from 8 o'clock to 9 o'clock, the traffic time index increased rapidly while the average speed became lower. In the rush hour, it is shown that the recurrent mixture density network and ARIMA model successfully capture this sudden change of the indices, while the performance of the LSTM network is not as good as the other two methods. Meanwhile, the overall prediction results of the recurrent mixture density network and the LSTM network match the ground truth better than that of the ARIMA model.

5. Discussion and Conclusion

In this paper, we are dedicated to the issue of short-term traffic flow prediction. The main contributions of this paper are as follows: (1) we constructed the recurrent mixture density network by integrating the LSTM network and mixture density network. To the best of our knowledge, it is the first time that the recurrent mixture density network is utilized in a short-term traffic flow prediction issue. (2) We attempted to implement these methods on a practical short-term traffic flow prediction task, including two comparison methods.

The advantage of the recurrent mixture density network consists in two aspects: (1) it has the ability to learn historical information and contextual dependency in traffic flow data series, which is useful for yielding promising prediction. (2) The strategy of fuzzy prediction generating probable distributions of the following traffic flow indices instead of concrete values is closer to the actual behavior of traffic flow evolution. Different from most of the existing work processing one certain traffic flow index at once, we take the coherent relationship between the traffic time indices and the average speed into consideration. The recurrent mixture density network treats the two indices as a two-dimensional vector and processes it simultaneously. Meanwhile, we predict the traffic flow at several following time steps at once, which could provide more valuable information for traffic situation awareness.

The recurrent mixture density model is easy to be applied to different application scenarios since it has strong generalization and learning ability. In practical fields, short-term traffic flow prediction is expected to play a crucial role in the development of intelligent transportation systems and the construction of a smart city.

Data Availability

The data can be accessed on the page https://www.sodic.com.cn/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Z. Lv, H. Fu, W. Tang, and X. Chen, "Traffic jam prediction based on analysis of residents spatial activities," in *Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, Guiyang, China, April 2020.

[2] Yu Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)-Special Section on Urban Computing*, vol. 5, no. 3, pp. 1–55, 2014.

[3] Z. Dongjie, D. Haiwen, S. Yundong, and C. Ning, "Research on path planning model based on short-term traffic flow prediction in intelligent transportation system," *Sensors*, vol. 18, no. 12, p. 4275, 2018.

[4] Y. Zheng, "Trajectory data mining," *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no. 3, pp. 1–41, 2015.

[5] C. Shifen, "A spatiotemporal multi-view-based learning method for short-term traffic forecasting," *International Journal of Geo-Information*, vol. 7, no. 6, p. 218, 2018.

[6] D. I. Tselentis, E. I. Vlahogianni, and M. G. Karlaftis, "Improving short-term traffic forecasts: to combine models or not to combine?" *IET Intelligent Transport Systems*, vol. 9, no. 2, pp. 193–201, 2015.

[7] 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.

[8] I. Okutani, "The Kalman filtering approaches in some transportation and traffic problems," *Transportation & Traffic Theory*, 1987.

[9] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: a survey," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[10] Y. Endo, H. Toda, K. Nishida, and J. Ikedo, "Classifying spatial trajectories using representation learning," *International Journal of Data Science and Analytics*, vol. 2, no. 3-4, pp. 107–117, 2016.

[11] R. Chen, M. Chen, W. Li, J. Wang, and X. Yao, "Mobility modes awareness from trajectories based on clustering and a convolutional neural network," *ISPRS International Journal of Geo-Information*, vol. 8, no. 5, p. 208, 2019.

[12] M. Chen, J. M. Davis, C. Liu, Z. Sun, M. M. Zempila, and W. Gao, "Using deep recurrent neural network for direct beam solar irradiance cloud screening," in *Proceedings of the Remote Sensing and Modeling of Ecosystems for Sustain ability XIV*, International Society for Optics and Photonics, San Diego, CA, USA, September 2017.

[13] R. Chen, M. Chen, W. Li, and N. Guo, "Predicting future locations of moving objects by recurrent mixture density network," *ISPRS International Journal of Geo-Information*, vol. 9, p. 2, 2020.

[14] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box–Jenkins techniques," *Transportation Research Record*, vol. 722, pp. 1–9, 1979.

[15] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 53–72, 2009.

[16] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, "The retrieval of intra-day trend and its influence on traffic prediction," *Transportation Research Part C: Emerging Technologies*, vol. 22, pp. 103–118, 2012.

[17] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.

[18] N. Polson and V. Sokolov, "Deep learning predictors for traffic flows," *Transportation Research Part C Emerging Technologies*, vol. 79, pp. 1–17, 2016.

[19] D. Kang, Y. Lv, and Y. Y. Chen, "Short-term traffic flow prediction with LSTM recurrent neural network," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, Japan, October 2017.

[20] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, Wuhan, China, November 2017.

[21] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "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] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proceedings of the IEEE International Conference on Smart City/socialcom/sustaincom*, Chengdu, China, December 2015.

[23] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, p. 4, 2017.

[24] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," 2016, https://arxiv.org/abs/1612.01022.

[25] Y. Lv, D. Yanjie, K. Wenwen, and L. Zhengxi, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 865–873, 2015.