Neurocomputing 408 (2020) 285–291

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Application of LSTM for short term fog forecasting based on meteorological elements

Kai-chao Miao [b,1], Ting-ting Han [a,1], Ye-qing Yao [b], Hui Lu [a], Peng Chen [c,e,*], Bing Wang [d,*], Jun Zhang [a,*]

[a] Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Electronic Engineering and Automation, Anhui University, Hefei 230061, Anhui, China
[b] Anhui Meteorological Bureau, Hefei 230031, Anhui, China
[c] National Engineering Research Center For Agro-Ecological Big Data Analysis & Application, School of Internet, Anhui University, 230601, Hefei, Anhui, China
[d] School of Electrical and Information Engineering, Anhui University of Technology, Ma Ansha 243032, Anhui, China
[e] Institute of Health Sciences and Technology, Institutes of Physical Science and Information Technology, Anhui University, 230039 Hefei, Anhui, China

ARTICLE INFO

Article history:
Received 12 April 2019
Revised 20 November 2019
Accepted 8 December 2019
Available online 11 May 2020

Keywords:
Fog forecasting
Time series data
Long short-term memory(LSTM)

ABSTRACT

Fog is the main weather phenomenon that causes low visibility, which makes traffic and outdoor work extremely dangerous. In this paper, we propose a novel LSTM framework for short-term fog forecasting. The proposed network framework consists of an LSTM network and fully connected layer. In order to make the proposed LSTM framework work, the meteorological element observation data returned hourly is transferred into time series data. Based on these time series data, four datasets are created for short-term fog forecasting. In order to evaluate the proposed LSTM framework, we conduct comprehensive experiments with different machine learning algorithms. Compared with K-Nearest Neighbor (KNN), AdaBoost and convolutional neural network (CNN) algorithms, the experimental results show that the proposed LSTM framework achieves best prediction performance in four evaluation criteria. Especially in TS-Score, the proposed LSTM framework achieve 1.1%, 11%, 3%, and 11% higher performance than the best traditional machine learning algorithm.

1. Introduction

Fog is a kind of ground touching cloud that horizontal visibility less than 1000 m [1], the formation and dissipation are the results of complex interactions between various physical, thermal and dynamic processes [2,3]. Low visibility caused by fog influences human society in many ways. Fog affects transportation system like expressways, high-speed railways, airports, and waterways [4]. Due to fog combined with polluted air can form mist, which can cause various diseases and even endanger people's lives [5]. Research shows that the economic and human losses caused by fog are equivalent to those caused by extreme weather such as hurricanes and tornadoes [6]. Therefore, the fog forecasting becomes more important in many fields including agriculture, economics, public health, and transportation. However, since many factors affect fog formation, it is still difficult to predict fog effectively using existing methods.

Considering that the meteorological data is collected systematically, machine learning algorithms [7–9] are considered as one effective technique for meteorological application. In order to improve the accuracy of fog forecasting, many research works use a machine learning algorithm to create models for fog forecasting. The artificial neural networks (ANN) [10–12] is first considered to provide more accurate fog forecasting at Canberra International Airport(YSCB). Then the neural network is used to create the nonlinear model for fog forecasting on 39 terminal aerodrome forecast stations in the northwest United States [13,14]. The decision tree algorithm combined with ANN is employed to predict 3 hourly visibility over Kolkata airport in India [15]. A Bayesian decision network, support-vector regression and Gaussian process algorithm [16,17] have also been applied for low-visibility event forecasting. Since the classifier ensemble algorithm can achieve better performance than a single model, a local ensemble prediction system(LEPS) [18] is designed for fog forecasting in Charles de Gaulle International Airport in Paris.

* Corresponding authors.
E-mail address: 00568@ahu.edu.cn (P. Chen).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neucom.2019.12.129
0925-2312/

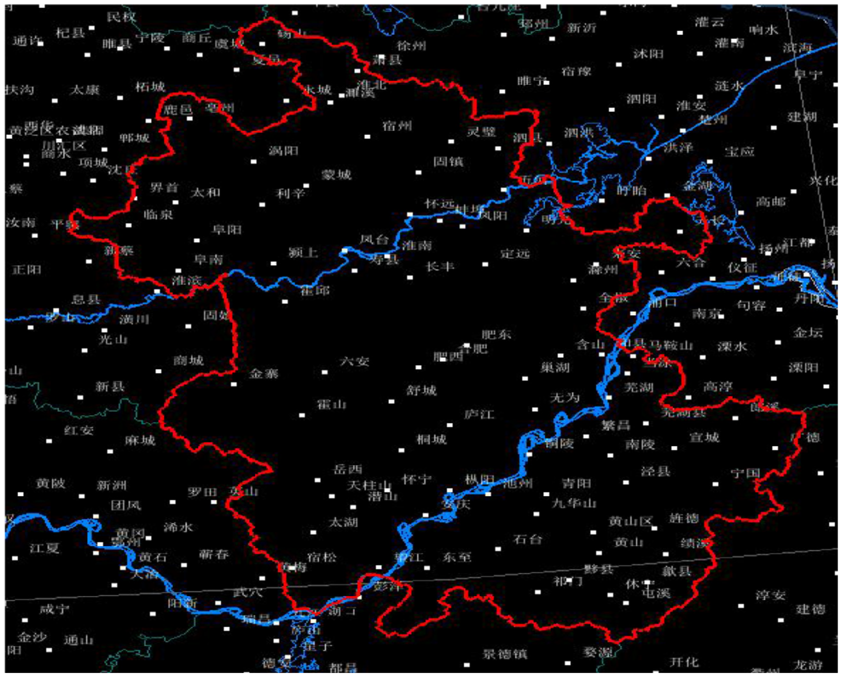

Fig. 1. Distribution map of national meteorological ground observation station in Anhui Province,China.

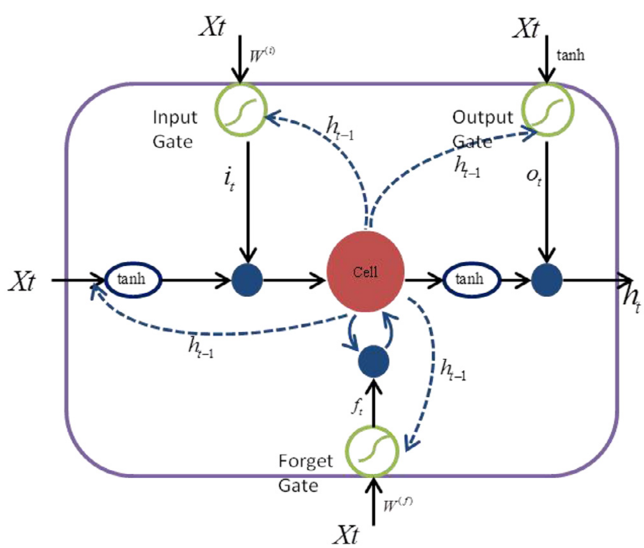

Fig. 2. Memory unit design of LSTM.

Although these research works show some good prediction results, the fog forecasting still faces a lot of challenges. In fact, the formation of fog is not only related to the meteorological elements of the current time but also related to the meteorological elements of the past time points. If the meteorological element data observed at a continuous time is employed to train the prediction model, the performance of the prediction model may be more effective. Therefore, instead of using the meteorological elements data in one time, the meteorological element data over a period of time is made into time series data. These data can be used to describe the dynamic change of meteorological element data over a period of time.

In recent years, deep learning especially the convolutional neural network (CNN) has achieved many successful applications in image classification and object detection [19–22]. Different from CNN, Recurrent Neural Network (RNN) has been applied in sequence data analysis [23]such as natural language processing (NLP) and speech recognition [24,25]. Long short term memory network (LSTM) is a special kind of RNN networks, which uses the out gate and forget gate for short-term memory. Compare ordinary RNN, LSTM model is a more effective tool to capture long time series information. Considering that the fog formation has a close relationship with previous meteorological status, it is very necessary to use a more effective approach to obtain the relation between time series data. Therefore, based on the time series data of meteorological elements, we use LSTM network for short-term fog forecasting. To our best knowledge, it is the first attempt to employ LSTM model in fog forecasting.

The remainder of this paper is organized as follows. In Section 2, the ground observation dataset and the proposed LSTM network framework is introduced. The extensive experiments and the discussions of the experimental results are provided in Section 3. Finally, we give conclusions in Section 4.

2. Data and methodology

2.1. Ground observation dataset

The target of this work is to predict the presence of fog, we have used hourly meteorological element observation data from Anhui Meteorological Bureau of China. Ground-based meteorological element observations are some measurements of near-surface atmospheric conditions through some meteorological instrument. In order to provide meteorological information for weather analysis and weather forecasting, China Meteorological Administration set up national ground station throughout the country and surface meteorological data can be obtained from the national ground station every hour. The available meteorological variables are summarized as below: temperature (TEMP), air pressure (AIP), wind speed (WS), wind direction (WD), and rainfall (RF), Dew Point Temperature (DPT), Relative Humidity (RH), Visibility per minutes(VISPM) and Visibility hourly(VISPH). In this paper, we have employed about 3 years of meteorological data to create a fog forecasting dataset, from October 1, 2015, to June 1, 2017. The Meteorologi-

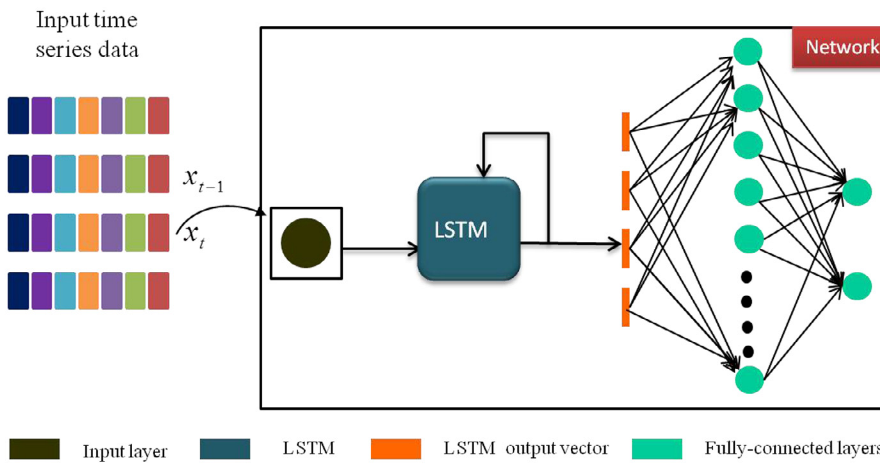

Fig. 3. LSTM network structure for fog forecast.

cal element observation data are provided by 81 national ground observation stations within Anhui Province, China. The distribution of the national weather stations is shown in the Fig. 1.

2.2. Methodology

LSTM [26] is a special kind of recurrent neural network (RNN) [27]. Compared with RNN, LSTM can learn long-term dependence information and avoid the gradient explosion and gradient disappearing during training period. Unlike other traditional machine learning algorithms, LSTM is a effective model to learn the features from sequence data automatically. The output of LSTM model can be a variable length sequences that can be used for fog forecasting. For example, for a time series data x_t ($t = 1,2,3,4,5\ldots$), at each step, x_t and the output of the previous LSTM cell are taken as input for current LSTM cell. At the same time, the current LSTM cell output a result. In LSTM network, the memory unit generates a state vector for the current time. In Fig. 2, a memory block of one unit of LSTM is shown, in which the red circle represents the memory unit. Different from ordinary RNN, LSTM use the memory block to store previous status information. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. In Fig. 3, blue solid circles represent multiplication and green circles represent sigmoid functions. In the equations below, the specific process of LSTM workflow is introduced:

$$i_t = \sigma(W^{(i)}H + b_i) \tag{1}$$

$$f_t = \sigma(W^{(f)}H_{(f)} + b_f) \tag{2}$$

$$o_t = \sigma(W^{(o)}H_{(o)} + b_o) \tag{3}$$

$$C_t{}' = \tanh(W^{(C)}H_c + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * C_t{}' \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

where i_t,f_t,o_t represent the output of the input gate, forget gate and output gate respectively.C_t represents the status of the current Cell, h_t is the output of the unit, a green circle is the sigmoid function, w_i, w_f, and w_o are the weights of the input gate, the forget gate and the output gate respectively. The weights are shared at each time step. H represents the superposition of the current input vector x_t and the output vector h_{t-1} of the unit at the previous moment H=[x_t, h_{t-1}]. The input gate is used to control which information is retained, preventing useless information from entering the memory unit. Forget gate is used to decide what information to discard from the status of the unit in the previous step. Forget gate and input gate will make the status of the memory unit to be updated.

2.3. LSTM network for fog-forecast

In order to train the LSTM model with the meteorological element data, the working mechanism of the proposed framework is shown in Fig. 3. The meteorological element data returned every hour is represented as a vector x_t and the multiple vectors into time series data. The fully-connected layers are connected to the LSTM network. The input of the network is the time series x_t and the output of the network is the prediction results. The process of LSTM framework can be summarized as the following steps:

(1) Firstly, the hourly meteorological element vector x_t is input into the input layer.
(2) LSTM receives input vector x_t combined with the output h_{t-1} of the unit at a time point when t=1, the hidden layer state is 0. The output of the input gate and forget gate of LSTM is a value between 0 and 1 through a sigmoid function. By using this mechanism, LSTM network can determine when and which information will enter the memory unit and which information should be forgotten by the memory unit. The state of the hidden layer is updated by the output of the input gate and the hidden gate. Cell outputs h_t based on cell state.
(3) Then, the meteorological element vector at the next time point and the output ht-1 of the unit at the previous time point are input into LSTM and the process is repeated.
(4) Finally, the output vector of the last time step of LSTM is input into the full-connected layer. The fully-connected layer extracts the features of the meteorological element time series data for further process, the classifiers are connected to the fully-connected layer to obtain the final label.

The number of LSTM units has a great influence on the performance of the classification. We conduct comprehensive experiments to get the optimal parameters set up by adjusting the number of units.

Table 1
The number of training sample and test sample.

Dataset	Sample	Train set	Validation set	Test set	Sequence length	Forecast time
DATA1	FOG	36,151	1500	10,246	2	1,2
&DATA2	CLEAR	764,341	1500	2,282,373	2	1,2
DATA3	FOG	39,640	1500	9399	4	3,4
&DATA4	CLEAR	761,605	1500	269,430	4	3,4

3. Experiment and discussion

3.1. Dataset and evaluation criterion

The meteorological element data is provided by Anhui Meteorological Bureau. In most situation, people only pay more attention to the status of fog. Therefore, we choose a categorical perspective to predict fog events. All the data is divided into two categories according to its possible visibility: FOG, the visibility is less than 1 km and CLEAR, the visibility is larger than 1 km. Since the formation of fog has a very close relationship with previous meteorological factors, the meteorological element data is processed into time series data according to the requirement of LSTM. In this work, we consider two different time-series length data to validate the performance of the proposed model. The original meteorological data is divided into four different datasets according to its sequence length and prediction time. The sequence length of Data1 and Data2 is set as 2, the label of Data1 and Data2 is the fog status in next 1,2 h respectively. Therefore, Data1 and Data2 can be used to train model to predict fog existing for next 1 and 2 h. In constrast, the sequence length of Data3 and Data4 set as 4, considering that sequence length is much longer, the label of Data3 and Data4 is the fog status in next 3,4 h respectively. To carry out the experiments, each dataset is divided into training sets and test sets. The training set is chosen from October 1, 2015 to December 30, 2016. The test set is available from January 1, 2017 to June 2017. In order to prevent the model over-fitting, we split a little data from the training sets to create the validation sets, the validation sets are used to monitor the training status. The number of samples of four datasets is described in Table 1.

Since the dataset is collected from real scence, the dataset is highly imbalanced because of the foggy event is rare in daily life. We calculate the ratio between FOG to CLEAR sample for all datasets, the ratio is approximately 1:20. To make the experimental results more effective, a simple oversample technique is used to deal with the imbalance problem. The FOG samples are repeated to select into the training set to make the ratio of FOG and CLEAR samples are 1:2. In order to verify the performance of the proposed LSTM model in a real scene, all test sets taken from the original data do not change the sample ratio. To assess the effectiveness of the proposed LSTM network model, four evaluation criteria are used. Among them are precision, accuracy, F1-score, and TS-score. The definitions are given as follows:

$$\Pr ecision = \frac{TP}{TP+FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{8}$$

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} \tag{9}$$

$$Accuracy = \frac{TP+TN}{TP+TF+NP+NF} \tag{10}$$

$$TS = \frac{TP}{TP+FP+FN} \tag{11}$$

Table 2
The experimental results of different LSTM cell number (n_c).

n_c	Prediction time	Precision	F1-Score	Accuracy	TS-Score
	1	0.595	0.73	0.97	0.55
30	4	0.33	0.47	0.935	0.31
	1	0.62	0.75	0.97	0.60
40	4	0.34	0.47	0.935	0.316
	1	0.63	0.756	0.974	0.60
50	4	0.33	0.47	0.93	0.31
	1	0.61	0.75	0.97	0.60
60	4	0.33	0.41	0.924	0.306
	1	0.62	0.75	0.98	0.6
70	4	0.344	0.482	0.938	0.316

Table 3
The experimental results of different LSTM layer number (n_l).

n_l	Precision time	Precision	F1-score	accuracy	TS-Score
	1	0.62	0.74	0.98	0.60
1	4	0.33	0.47	0.94	0.31
	1	0.55	0.70	0.97	0.54
2	4	0.33	0.47	0.93	0.31
	1	0.64	0.76	0.98	0.62
3	4	0.31	0.45	0.93	0.30

Table 4
The experimental results of different number of nodes (nfc) in fully connected layer.

n_{fc}	Prediction time	precision	F1-score	accuracy	TS-Score
	1	0.62	0.746	0.98	0.595
1(100)	4	0.338	0.48	0.9325	0.316
	1	0.618	0.744	0.922	0.596
1(200)	4	0.358	0.802	0.942	0.326
	1	0.634	0.758	0.98	0.61
2(100,100)	4	0.3175	0.46	0.9275	0.29
	1	0.64	0.758	0.697	0.611
2(200,100)	4	0.34	0.474	0.405	0.31

where TP (true positive) is the number of FOG samples are correctly classified. TN (true negative) is the number of CLEAR samples that are classified as CLEAR samples. FP (false positive) is the number of CLEAR samples is incorrectly classified as FOG samples. FN (false negative) refers to the number of FOG samples that are classified as CLEAR samples. Precision is the ratio of correctly predicted FOG samples to the total predicted FOG samples. Recall refers to the ratio of the correctly predicted FOG samples to all FOG samples, which is used to calculate F1 Score. F1 Score is the weighted average of Precision and Recall. Threat Score (TS-Score) or critical success index (CSI) is used extensively by the national weather service to indicate the value of weather forecaseting. TS-Score ranges from zero to one. In this work, TS-Score is also used as one criterion to validate the performance of the proposed LSTM model.

3.2. Experiment setup

In order to evaluate the effectiveness of the proposed LSTM network, the experiments are carried out on four datasets including DATA1, DATA2, DATA3, DATA4. The structure parameters of the neural network play a very important role for all experiments.

Table 5
The Comparison Experiments for next 1 to 2 h prediction.

	CNN	DATA1			CNN	DATA2		
		LSTM	AdaBoost	KNN	CNN	LSTM	AdaBoost	KNN
F1-score	0.508	0.758	0.74	0.60	0.485	0.66	0.61	0.50
Precision	0.378	0.640	0.63	0.48	0.38	0.51	0.48	0.38
TS-Score	0.34	0.611	0.60	0.43	0.32	0.550	0.44	0.33

Table 6
The Comparison Experiments for next 3 to 4 h prediction.

	CNN	DATA3				DATA4		
		LSTM	AdaBoost	KNN	CNN	LSTM	AdaBoost	KNN
F1-score	0.444	0.53	0.51	0.40	0.33	0.474	0.475	0.36
Precision	0.34	0.39	0.38	0.29	0.21	0.34	0.342	0.73
TS-Score	0.28	0.36	0.39	0.25	0.20	0.31	0.311	0.22

Three parameters n_l (the number of layers of LSTM), n_c (the number of cells of the LSTM) and n_{fc} (the number of layers of the fully connected layer) are tuned under different datasets carefully. Considering DATA1, DATA2, and DATA3, DATA4 use the same length of the time series data respectively, we only conduct experiments on DATA1 and DATA4 to determine the optimal parameters. When the optimal parameters are obtained, these parameters are used to determine the structure of the LSTM network for the rest experiments. Besides, we conduct many experiments to determine the importance of nine meteorological elements. In order to verify the performance of the proposed LSTM framework, many traditional machine learning algorithms including AdaBoost and K-Nearest Neighbor(KNN) are used for comparison experiments. All these machine learning algorithms are based on the Scikit-Learn library [28]. To make the comparison experiments fairer, convolution neural networks (CNN) is also used as a reference method. Some machine learning algorithms and their parameters setup are listed as follows:

AdaBoost: AdaBoost algorithm is an ensemble algorithm, the basic classifier is a decision tree. The number of base classifiers is set as 200.

CNN: Considering the meteorological data is not complicated, we design a simple network structure including an input layer, convolution layer, a pooled layer, and three fully connected layers. We convert the time series data of the meteorological elements into 2D images to fit the CNN network. 300 convolutional kernels are set for each convolution layer and the size of the convolution kernel is 3 × 3. The pooling layer selects max-pooling, the kernel size is 2 × 2, and the stride is 2. The two fully connected layers choose 200 filters respectively.

In this work, the used computational environment is Intel i7 3.4 Ghz*4 cores, 16 G RAM and NVIDIA GTX1080,ubuntu16.04 64 bits operating system, and python2.7. CNN and LSTM network is implemented on Tensorflow1.4.0.

3.3. Optimal parameters setup

In Table 2, the experiments are performed on DATA1 and DATA4 to determinen_c when the length of the time series is 2 and 4 respectively. We set n_l equals 1 and n_{fc} equals 2. The optimal n_c will be determined from some option values, among these values are 30, 40,50, 60 and 70. Usually, the Precision and F1 score are used to evaluate the performance of a model for imbalanced data. Considering these two evaluation criteria n_c is set as 50 for DATA1. The same experiment is performed on DATA4 and the setup is similar to the experiments on DATA1. When n_c is set as 70, the optimal performance is achieved. Therefore, in rest experiments, n_c is set as 50 for DATA1, DATA2 and 70 for DATA3, DATA4 respectively.

When the n_c is fixed, we need to use experiments to obtain the optimal n_l value. Some option n_l value is prepared for the experiment based on DATA1, DATA4. When n_l set as 1, the performance of DATA4 is slightly better than the others from Table 3. Therefore, n_l is set as 1 forDATA3, DATA4. The best performance is obtained based on DATA1 when n_l is set as 3. So,n_l is set as 3 for the experiments of DATA1, DATA2.

Based on the above two experiments, n_c and n_l are set with the optimal value obtained before. The different number of nodes for the fully-connected layer (n_{fc}) is used. In Table 4, it can be found that the performance of setting different nodes per layer is almost the same. Therefore, the number of nodes set as (100,100) for two layers. n_{fc} set as 200 for one full-connected layer based on DATA4. Therefore, in the following experiments, when the length of the time series is 2, n_{fc} set as 100, two full-connected layers are used. When the time series length is 4, n_{fc} set as 200 and only one fully-connected layer is used.

In order to verify the importance of nine meteorological factors, we delete one factor each time in purpose to create a subset for validation experiment based on DATA1. The validation experiments still use the same evaluation criteria as before. If one factor is removed from the original dataset, the performance of the experiments is decreased significantly, the meteorological factor is considered important. Otherwise, the meteorological factor may be considered as irrelevant to the fog prediction. The experimental results are shown in the Fig. 4. It is clear to find that hourly visibility and wind speed are two important meteorological factors for fog forecasting. In contrast, other factors do not produce a great impact on the prediction of fog.

3.4. Experimental results and discussion

To validate the performance of the proposed LSTM model, we compare the proposed method with some popular machine learning algorithms including AdaBoost, KNN and CNN. The comparison experiments are conducted on four datasets: DATA1, DATA2, DATA3, and DATA4 respectively. Tables 5 and 6 show the experimental results on DATA1, DATA2, and DATA3, DATA4 respectively. Based on four datasets, the proposed LSTM network achieved the best performance in the four evaluation criteria. Especially on TS_score, the best results obtained, which are 61.1%, 55%, 39%, and 34% for the next 1 to 4 h. Compared with the CNN method, it improves by 27.1%, 23%, 8%, and 11% respectively. In traditional machine learning algorithm, AdaBoost has achieved the best performance, but it still cannot compete with the proposed LSTM model. The proposed LSTM model achieve 1.1%, 11%, 3%, and 11% higher than AdaBoost in TS-Score. TS-Score of different methods versus different prediction time is drawn in Fig. 5. With the forecast

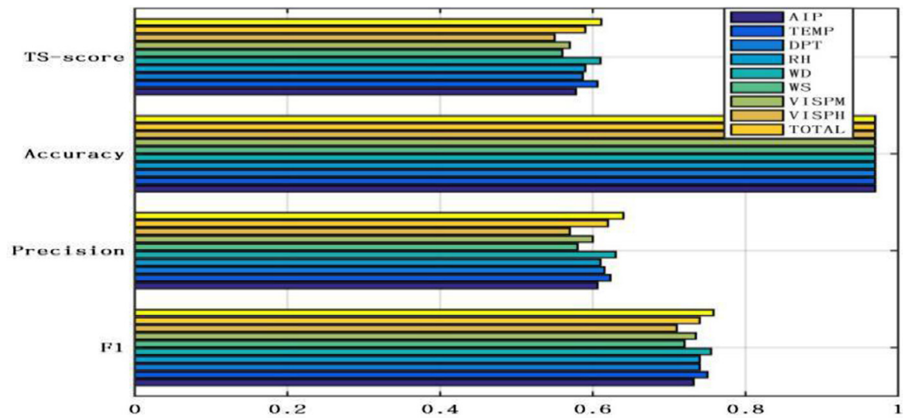

Fig. 4. Importance of meteorological element factors.

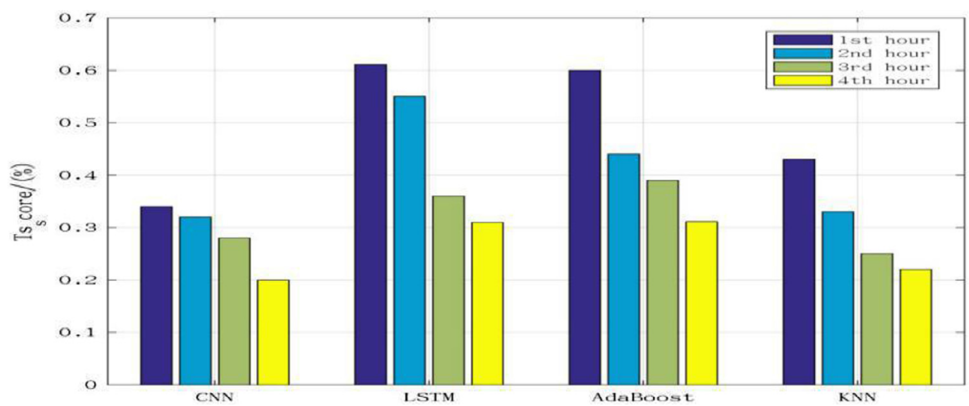

Fig. 5. TS-Score of CNN, LSTM, KNN, and AdaBoost.

time increasing, TS-Score is decreased significantly. It means that the lower prediction accuracy accompanies with long prediction time. In other words, it is more difficult to predict fog in a long time. This is consistent with current research on the formation of the fog.

4. Conclusion

In this paper, we proposed a novel LSTM framework for short-term fog forecasting. In order to evaluate the proposed LSTM framework, the original meteorological element observation data was transferred into time series data. To make the evaluation more effective, we created four different datasets based on time series data and used four evaluation criteria to verify different algorithms. Two datasets (DATA1, DATA4) were used to determine the optimal parameters for the proposed LSTM framework. After the optimal parameters determined, a large number of comparison experiments were conducted. Compared with KNN, AdaBoost and CNN algorithms, the experimental results show that the proposed LSTM framework can achieve better performance for short-term fog forecasting.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61872004 and 61672035, Jiangsu Province Meteorological Bureau BeiJiGe grant No. BJG201707, Anhui Province Meteorological Bureau meteorologist special grant No. KY201704, Anhui Provincial Natural Science Foundation (grant number 1608085MF136), Educational Commission of Anhui Province (No. KJ2019ZD05), Anhui Province Funds for Excellent Youth Scholars in Colleges (gxyqZD2016068), the fund of Co-Innovation Center for Information Supply & Assurance Technology in AHU (ADXXBZ201705), and Anhui Scientific Research Foundation for Returness.

References

[1] T.S. Glickman, 2000: Glossary of Meteorology, American Meteor Society, 1995.
[2] C. Román Cascón, et al., Observations and WRF simulations of fog events at the Spanish Northern Plateau, Adv. Sci. Res. (8) (2012) 11–18.
[3] C. Román-Cascón, et al., Forecasting radiation fog at climatologically contrasting sites: evaluation of statistical methods and WRF, Q. J. R. Meteorol. Soc. 142 (695) (2016) 1048–1063.
[4] J. Bendix, W. Eugster, O. Klemm, Fog—Boon or bane? Erdkunde (2011) 229–232.
[5] Z. Lina, Z. Chaolin, W. Bizheng, Evolution characteristics of atmospheric visibility in the Beijing expressway and the corresponding physical analysis, Chin. J. Atmos. Sci. 32 (6) (2008) 1229–1240.
[6] I. Gultepe, et al., Fog research: a review of past achievements and future perspectives, Pure Appl. Geophys. 164 (6-7) (2007) 1121–1159.
[7] B. Li, J. Du, X.-P. Zhang, Feature extraction using maximum nonparametric margin projection, Neurocomputing 188 (2016) 225–232.
[8] B. Li, L. Lei, X.-P. Zhang, Constrained discriminant neighborhood embedding for high dimensional data feature extraction, Neurocomputing 173 (2016) 137–144.

[9] B. Li, et al., Robust dimensionality reduction via feature space to feature space distance metric learning, IEEE Trans. Neural Netw. 112 (2019) 1–14.
[10] D. Fabbian, R. de Dear, S. Lellyett, Application of artificial neural network forecasts to predict fog at Canberra International Airport, Weather Forecast. 22 (2) (2007) 372–381.
[11] D.-S. Huang, A constructive approach for finding arbitrary roots of polynomials by neural networks, IEEE Trans. Neural Netw. 15 (2) (2004) 477–491.
[12] D.-S. Huang, H.H. Ip, Z. Chi, A neural root finder of polynomials based on root moments, Neural Comput. 16 (8) (2004) 1721–1762.
[13] C. Marzban, S. Leyton, B. Colman, Ceiling and visibility forecasts via neural networks, Weather Forecast. 22 (3) (2007) 466–479.
[14] W.W. Hsieh, B. Tang, Applying neural network models to prediction and data analysis in meteorology and oceanography, Bull. Am. Meteorol. Soc. 79 (9) (1998) 1855–1870.
[15] D. Dutta, S. Chaudhuri, Nowcasting visibility during wintertime fog over the airport of a metropolis of India: decision tree algorithm and artificial neural network approach, Nat. Hazards 75 (2) (2015) 1349–1368.
[16] T. Boneh, et al., Fog forecasting for Melbourne Airport using a Bayesian decision network, Weather Forecast. 30 (5) (2015) 1218–1233.
[17] L. Cornejo-Bueno, et al., Efficient prediction of low-visibility events at airports using machine-learning regression, Bound. Layer Meteorol. 165 (2) (2017) 349–370.
[18] S. Roquelaure, T. Bergot, A local ensemble prediction system for fog and low clouds: Construction, Bayesian model averaging calibration, and validation, J. Appl. Meteorol. Climatol. 47 (12) (2008) 3072–3088.
[19] O. Abdel-Hamid, et al., Convolutional neural networks for speech recognition, IEEE/ACM Trans. Audio Speech Lang. Process. 22 (10) (2014) 1533–1545.
[20] Z. Wang, T. Oates, Encoding time series as images for visual inspection and classification using tiled convolutional neural networks, in: Proceedings of the Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.
[21] I. Sutskever, G.E. Hinton, A. Krizhevsky, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105.
[22] Y. Yuan, L. Mou, X. Lu, Scene recognition by manifold regularized deep learning architecture, IEEE Trans. Neural Netw. Learn. Syst. 26 (10) (2015) 2222–2233.
[23] P. Lingras, S. Sharma, M. Zhong, Prediction of recreational travel using genetically designed regression and time-delay neural network models, Transp. Res. Rec. 1805 (1) (2002) 16–24.
[24] A. Graves, N. Jaitly, Towards end-to-end speech recognition with recurrent neural networks, in: Proceedings of the International Conference on Machine Learning, 2014.
[25] M. Sundermeyer, H. Ney, R. Schlüter, From feedforward to recurrent LSTM neural networks for language modeling, IEEE/ACM Trans. Audio Speech Lang. Process. 23 (3) (2015) 517–529.
[26] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[27] A. Graves, *Generating Sequences With Recurrent Neural Networks*. arXiv preprint 2013.
[28] F. Pedregosa, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (Oct) (2011) 2825–2830.

Kaichao Miao was born in Anhui Province, China, in 1973. He received the B.S. degree in agrometeorology from Nanjing Meteorological Institute (now Nanjing University of Information Engineering) in 1997. He received M.S. degree in Software Engineering from Hefei University of Technology in 2004. Now, He is a senior engineer and deputy director of the Public Meteorological Service Center of Anhui Meteorological Bureau. He is mainly engaged in meteorological application software design and development, professional meteorological services and other research.

Tingting Han was born in Shandong Province,China. She received M.S. degree in Electronical Information Science and Technology from Dezhou University, Dezhou, China in 2005. She is now a candidate of a Master's degree in Pattern Recognition and intelligent system with Anhui University, China. Her research interests focus on machine learning, deep learning and image processing.

Yao Yeqing was born in Anhui province, China, in 1970. She received M.S. degree in Atmospheric Physics and Atmospheric Environment from Nanjing University in 2006. She worked at Anhui Meteorological Bureau for more than 20 years on weather forecasting, in particular dense fog forecasting. Her research interests focus on different weather forecasting methods.

Hui Lu was born in Anhui Province, China, in 1993. She received M.S. degree in Automation from Huangshan University, Huangshan, China in 2006. She is now a candidate of a Master's degree in Pattern Recognition and intelligent system with Anhui University, China. Her research interests focus on machine learning, deep learning and weather forecasting.

Peng Chen received his Bachelor degree from Electronic Engineering Institute, PLA (1997), Master degree from Kunming University of Science and Technology (2003), and Ph.D degree from University of Science and Technology of China (2007). He is an Professor in the School of Computer Science & Technology and Institutes of Physical Science & Information Technology, Anhui University, Hefei, China. He has published about 80 high quality referred papers in international conferences and journals. He speciliases in machine learning and deep learning with applications to bioinformatics, drug discovery, computer vision, etc. Prior to joining Anhui University, he served in City University of Hong Kong (2006, as senior research associate), Howard University, USA (2008–2009, as Postdoc Fellow), Nanyang Technological University, Singapore (2009–2010, as Research fellow), and King Abdullah University of Science and Technology (KAUST), Saudi Arabia (2012–2014, as Postdoc Fellow). From 2011 to 2013, he was an Associate Professor in Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, China.

Dr. Bing Wang received the B.S. and M.S degree from Hefei University of Technology, Hefei, China in 1998 and 2004 respectively. He received the Ph.D. degree from University of Science and Technology of China, Hefei, China in 2006. Now Dr. Wang is serving as professor in the School of Electrical and Information Engineering, Anhui University of Technology, Maanshan, China. His research interests mainly focus on machine learning, information processing and bioinformatics. He published more than 100 papers, and is a senior member of IEEE.

Jun Zhang was born in Anhui Province, China, in 1971. He received M.S. degree in Pattern Recognition and Intelligent System in 2004, from Institute of Intelligent Machines, Chinese Academy of Sciences. He received the Ph.D. degree from University of Science and Technology of China, Hefei, China in 2007. Currently, He is professor in the School of Electrical Engineering and Automation, Anhui University, China. His research interests focus on deep learning, ensemble learning and bioinformatics. He is a member of International Neural Network Society.