2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)

Long Short-term Memory Network Prediction Model Based on Fuzzy Time Series

Hua Qu
School of Software Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
qh@mail.xjtu.edu.cn

Jiaqi Li
School of Software Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
15594956230@163.com

Yanpeng Zhang
School of Software Engineering
Xi'an Jiaotong University
Xi'an, Shaanxi, China
hnypzhang@gmail.com

***Abstract*—This paper proposes a long short-term memory network (FTS-LSTM) prediction model based on fuzzy time series to improve the prediction accuracy of time series. First, the fuzzy C-means clustering FCM algorithm is used to classify the time series to form a fuzzy time series and obtain the membership matrix. Second, the LSTM net-work prediction model is constructed, and the FTS-LSTM network prediction model is proposed. The previously obtained membership is used as the full connection. The weight of the layer and its membership as the weight remain unchanged. This FTS-LSTM network prediction model not only considers the non-linearity and non-stationarity of the time series, but also resolves the inherent uncertainty and ambiguity of the data. Simulation results show that the FTS-LSTM network-based prediction model has faster training speed, higher prediction accuracy, and better prediction effect on time series with large ambiguities.**

Keywords*—*fuzzy C-means, fuzzy time series, membership matrix, LSTM, FTS-LSTM

I. INTRODUCTION

Time series refers to a series of certain statistical indicators arranged in chronological order. It has the characteristics of non-linearity and noise. There are many prediction methods for time series. Autoregressive moving average model (ARI-MA)[1] is a traditional linear time series prediction model, but ARIMA has greater limitations. With the development of artificial intelligence, Zhemin Li[2] and others used dynamic chaotic neural networks for agricultural product price prediction. Yong Long[3] used back propagation neural network (BPNN) for power load forecasting. However, due to their topological characteristics, these networks cannot capture long-term sequence dependencies, which reduces the accuracy of predictions.

The fuzzy C-means (FCM) algorithm was proposed by Bezdek in 1973. Specifically, it is a clustering algorithm that uses the degree of membership to determine the degree to which each data point belongs to a certain cluster. The basic idea of the algorithm is to iteratively optimize the objective function that represents the similarity between the sample point and the class C center, and obtain the local maximum value to obtain the optimal clustering.

Recurrent neural network (RNN) is a kind of recurrent neural network that can sense historical information[4],but RNN will eventually cause gradients to disappear or explode over time. Sak H[5] proposed a long-term short-term memory neural network (LSTM), which not only has the ability of long-term memory through the use of a "gate" structure[6], but also solves the problem of gradient disappearance or explosion. Therefore, LSTM has been widely used in various fields in recent years[7].

Aiming at the highly non-linear neural network may ignore the influence of linear factors, this paper proposes an FTS-LSTM prediction network model with membership as the weight of the fully connected layer, which improves the generalization ability and prediction accuracy of the network.

II. RELATED TECHNOLOGY RESEARC

A. Fuzzy C-means Clustering Algorithm

Clustering is to separate and classify things with the same attributes. It distinguishes the degree of similarity between things according to certain requirements.

Let $x_j = \{x_{j1}, x_{j2}, \ldots x_{jn}\}$,where x_{jn} represents the nth attribute of the sample x_j .The membership matrix U of the data set X is a matrix of $c \times n$, c is the number of clusters, u_{ij} represents the degree of membership of the j-th sample of the data set X to the i-th class, The membership u_{ij} meets the condition:

$$\sum_{i=1}^{c} u_{ij} = 1; \forall j = 1, 2, \ldots\ldots N \tag{1}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{ki}})^{2/(m-1)}} \tag{2}$$

Where m is the fuzzy index and $d_{ij} = \|x_j - v_i\|$ is the Euclidean distance between the data object x_j and the cluster center v_i .

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^m * x_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{3}$$

v_i represents the i-th cluster center.

The iterative process of the FCM algorithm is to find the membership matrix and cluster center that minimize the objective function. The calculation formula of the objective function value is:

978-1-7281-6590-5/20/$31.00 ©2020 IEEE

Dalian, China
March 20-22, 2020

$$J(u,v)=\sum_{j=1}^{n}\sum_{i=1}^{c}u_{ij}^{m}d^{2}(x_j,v_i) \quad (4)$$

The FCM clustering algorithm is an iterative process to obtain the cluster center and membership degree. The specific steps are:

1) *Initial*: Set the initial parameters of the FCM algorithm;

2) *Initial:* Random initial clustering center;

3) *Calculation:* The cluster center is calculated according to (3);

4) *Iteration:* Obtain the objective function value according to (4), and then determine whether the difference between the objective function values is less than a given threshold. If it is smaller, stop the algorithm iteration, otherwise skip to the calculation step 3) and continue the iterative calculation.

B. Long Short-Term Memory Network (LSTM)

LSTM network not only has the characteristics of long-term memory of historical information by RNN, but also avoids the problem of long-term dependence by adding forgetting gate. The LSTM cell structure includes forget gate f_t ,input gate i_t ,and output gate o_t ,The internal structure is shown in Fig. 1.

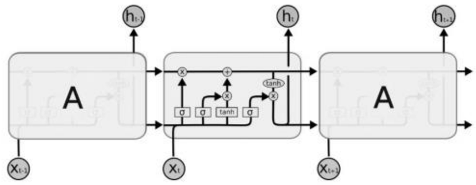

Fig. 1. Cell structure of long-term and short-term memory networks

The forgetting gate determines how much of the cell state C_{t-1} information at the previous moment is retained in the current state C_t , The calculation formula is:

$$f_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_f) \quad (5)$$

Where W_f and b_f are parameters, h_{t-1} is the output of the previous cell, x_t is the input of the current cell, σ is the sigmoid function and f_t is the output value of the forget gate;

The input gate controls the input of new information. The calculation formula is:

$$i_t=\sigma(W_i\cdot[h_{t-1},x_t]+b_i) \quad (6)$$

$$C_t^{'}=\tanh(W_c\cdot[h_{t-1},x_t]+b_c) \quad (7)$$

Where W_i and b_i are parameters, i_t and $C_t^{'}$ are the two output values in the input gate;

Update the memory unit status as:

$$C_t=f_t*C_{t-1}+i_t*C_t^{'} \quad (8)$$

Where c is the cell state of the previous cell;

The output gate calculation formula is:

$$O_t=\sigma(W_o\cdot[h_{t-1},x_{t-1}]+b_o) \quad (9)$$

Where W_o and b_o are parameters, and O_t is the sigmoid layer output value of the output gate;

Next, the cell state is processed through tanh and it is multiplied by the output of the sigmoid gate. In the end, the cell will only output the part that determines the output.

$$h_t=O_t*\tanh(C_t) \quad (10)$$

Although LSTM has achieved some results in the study of time series, when the target system contains a linear relationship and contains noise, a highly nonlinear LSTM network will cause the training model to over fit and reduce the accuracy of time series prediction.

III. Prediction Model Of Lstm Network Based On Fuzzy Time Series(Fts-Lstm)

The prediction model method of FTS-LSTM network first blurs the time series, and uses the FCM algorithm to cluster the time series set to obtain the fuzzy time series and membership matrix. Then, the FTS-LSTM model is constructed and the membership obtained by the previous module The degree is used as the weight of the fully connected layer of the model, and the weight of the membership degree of this layer is kept unchanged, so that other layers of the network participate in the training.

A. Fuzzy Time Series Module

This module will use the FCM algorithm to cluster the time series set to obtain a membership matrix, and integrate the time series into a fuzzy time series. The algorithm steps are as follows:

1) *Normalize:* Normalize the data to the number between -1 and 1;

2) *Serialization:* The data is converted into a time series set with a time interval of k: assuming the current time is t, $(x_t,x_{t+1},x_{t+2},......x_{t+k})$ is a time series, and the data is normalized into a time series set S with a time interval of k;

3) *Fuzzy clustering:* The fuzzy C-means algorithm is used to cluster the time series set to obtain the fuzzy time series and membership matrix U.

B. LSTM Model of Long-term and Short-term Memory Neural Network

The topology of the LSTM network structure is shown in Fig. 2:

In Fig.2, $\{x_0,x_1,x_2,......,x_{t-1}\}$ is the input value of t time steps, the first layer is the input layer, the second layer is the hidden layer of LSTM as the neuron, and the last layer is the fully connected layer connection as the final output layer.

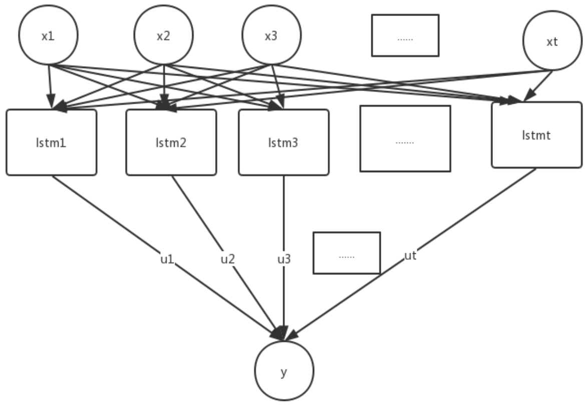

Fig. 2. LSTM network structure topology

C. LSTM Prediction Model Based on Fuzzy Time Series (FTS-LSTM)

This paper proposes the FTS-LSTM prediction model, so that the weight of the fully connected layer in the LSTM network model is assumed by the membership matrix obtained by the FCM-like algorithm to reflect the linear relationship in the time data, and also describes the uncertainty and ambiguity of the data.

The output after FTS-LSTM model is:

$$\hat{y}_t = u_1 h_t^1 + u_2 h_t^2 + u_3 h_t^3 + ... + u_c h_t^c \tag{11}$$

Therefore, the complete steps of the FTS-LSTM prediction model method are:

1) *Data pre-processing:* because FCM algorithm and LSTM network are more sensitive to data, the data needs to be normalized;

2) *Conversion:* Convert the data set normalized in step 1) to a time series set with a certain time interval;

3) *Fuzzy clustering:* Use FCM algorithm to cluster to obtain fuzzy time series and membership matrix;

4) *Freeze and traning:* Set the weight of the output layer in the prediction model of the prediction network to membership, and then use the divided time series training set to train the network model, and finally get the prediction value.

IV. Experimental Results And Analysis

A. Experimental Environment and Data

The experimental environment is python3, tensorflow1.4, kreas 2.2.

This experiment selected two sets of time-series data from different fields to test the usability and performance of the model. The first set came from the closing prices from January 2000 to September 2018 and the daily trading volume of the SP500 stock index. The second group is base station traffic data in a certain area of a city.

TABLE I. Data Set Information

No	Name	Amount	Train	Test
1	Stock closing price	4697	3992	705
2	Base station traffic	432	367	65

B. Experimental Evaluation Criteria

This article mainly evaluates the performance of FTS-LSTM through the following two indicators: (1) the comparison of the fitting effect between the prediction sequence of different prediction models and the actual sequence. The definition is as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i^{'} - y_i)^2 \tag{12}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i^{'} - y_i}{y_i}\right| \tag{13}$$

Among them, $y_i^{'}$ is the predicted value and y_i is the actual value.

C. Results and Analysis

Experiment 1: Comparison of the effect of fitting the prediction curve and the actual curve of different time-series models.

This paper uses LSTM network model and FTS-LSTM network model to design prediction experiments. The two models are respectively applied to the time series data of two groups in different fields, and the pros and cons of the different models are evaluated by fitting a line graph.

Fig. 3 to Fig. 6 shows the fitted and predicted values of the two prediction models on the two sets of data. It can be seen from the figure that the fitting effect of the FTS-LSTM network model is better than that of the LSTM network model. The LSTM model has a poor fitting effect in prediction. The FTS-LSTM network model uses membership as a weight, which overcomes the uncertainty and ambiguity of the time series and makes it fit well in the prediction. It shows that the FTS-LSTM prediction model has a stronger global optimization ability.

LSTM network fitting diagram:

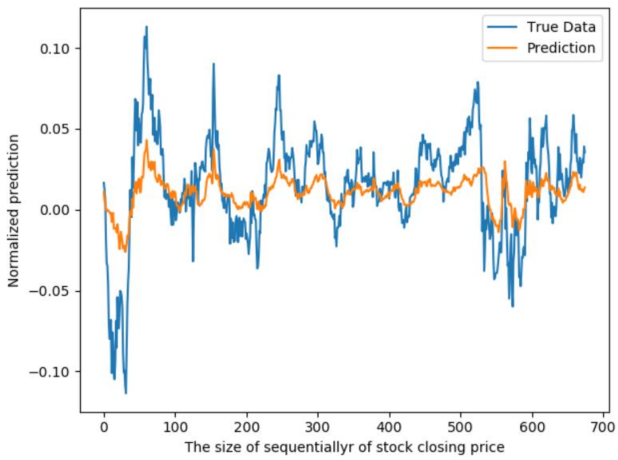

Fig. 3. Stock closing price

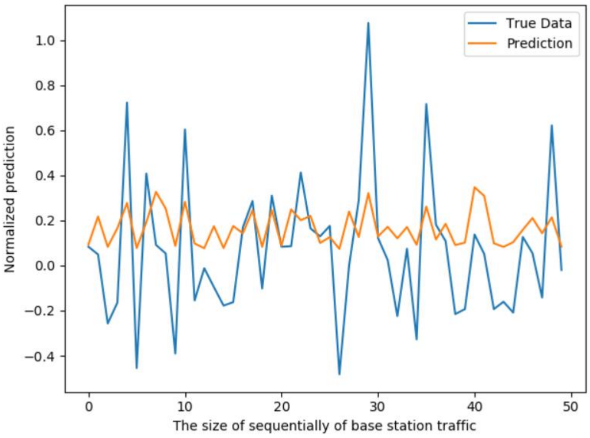

Fig. 4. Base station traffic data

FTS-LSTM network fitting diagram:

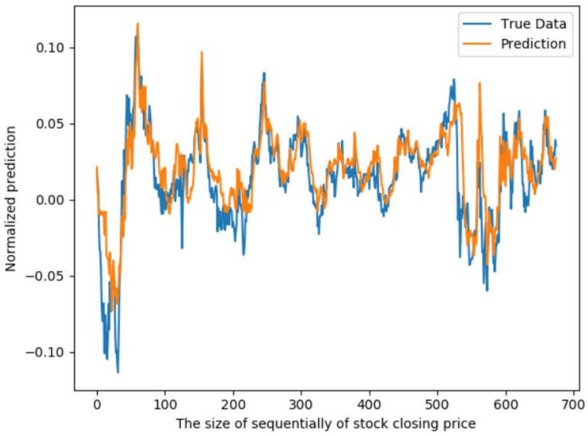

Fig. 5. Stock closing price

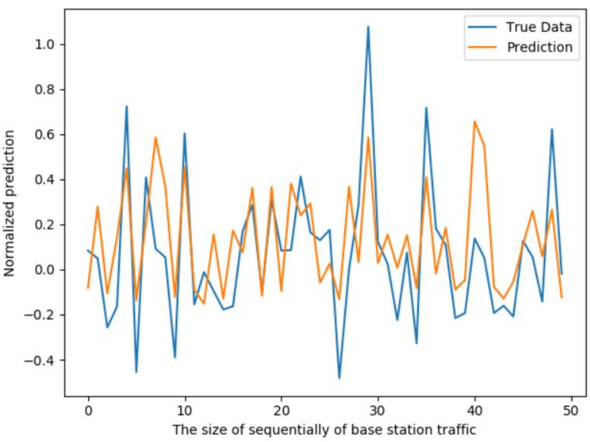

Fig. 6. Base station traffic data

Experiment 2: The normalized mean square error, average absolute error and model training time comparison results of each model under different data sets.

TABLE II. EXPERIMENTAL RESULTS OF TWO MODELS UNDER DIFFERENT DATA SETS

No	Name	MSE	MAE	Time
1	LSTM	1.7733e-04	0.0094	39.234187s
	FTS-LSTM	1.7445e-04	0.0091	27.473849s
2	LSTM	0.1840	0.3150	07.648123s
	FTS-LSTM	0.1647	0.2952	06.315752s

It can be seen from the table that the MSE and MAE of the FTS-LSTM model are significantly lower than the LSTM model, indicating that the FTS-LSTM prediction model has better prediction accuracy in terms of timing prediction than the traditional LSTM model. It is shown that using membership as the weight of a fully connected layer and freezing this layer improves the network's optimization ability, and then improves the prediction accuracy of the model. At the same time, the FTS-LSTM prediction model can shorten the training time at the "frozen" fully connected layer.

According to the above experiments, it can be concluded that the FTS-LSTM model proposed in this paper has stronger generalization ability on different data sets than the LSTM model, and is an efficient time series prediction model.

V. CONCLUSION

In order to enhance the effect on time series and improve the prediction accuracy of time series, this paper proposes a LSTM network prediction model based on fuzzy time series. First, the fuzzy C-means clustering FCM algorithm is used to classify the time series set to obtain a membership matrix; Secondly, an LSTM network prediction model is constructed, which includes an input layer, a LSTM hidden layer, and a fully connected layer as the output layer. Finally, an FTS-LSTM network prediction model is proposed. The membership of the weights remains unchanged, so that the fully connected layer becomes a "frozen layer" and does not participate in global training. This FTS-LSTM network prediction model considers both the time series non-linearity and the inherent uncertainty and ambiguity of the data. At the same time, the FTS-LSTM model proposed in this paper has high prediction accuracy in time series prediction, so this model can play an important role in practical applications. The next step is to study whether the network model proposed in this paper has the effect of further improving the prediction accuracy on the problem of fuzzy cluster number optimization.

ACKNOWLEDGMENT

Here I would like to thank Teacher Hua Qu and Ji Hong Zhao for their guidance and Brother Yan Peng Zhang for their help. This is with the guidance of my teachers and brothers, so that I can work hard and gain something on the road of scientific research. At the same time, I would like to thank the friends in the laboratory, each discussion will benefit me a lot. Finally, thank me My family has been silently supporting me behind my back, encouraging me, thank you all.

This work was financially supported by the National Natural Science Foundation of China (61531013) and National Major Project (2018ZX03001016).

REFERENCES

[1] Wang Y, Wang C, Shi C, et al. Short-term cloud coverage prediction using the ARIMA time series model[J]. Remote Sensing Letters, 2018, 9(3):275-284.

[2] Zhemin Li, Shiwei Xu, Liguo Cui. Prediction research based on dynamic chaotic neural network——Taking potato time series price as an example [J]. Systems Engineering Theory & Practice, 2015, 35 (8): 2083-2091.

[3] Yong Long, Zhenyu Su, Yu Wang. Monthly load forecasting based on seasonal adjustment and BP neural network [J]. Systems Engineering Theory and Practice, 2018, 38 (4): 1052-1060.

[4] Assaad M, Boné R, Cardot H. A new boosting algorithm for improved time-series forecasting with recurrent neural networks[J]. Information

Fusion, 2008, 9(1):41-55.

[5] Sak H, Senior A, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling[C]//Fifteenth annual conference of the international speech communication association. 2014.

[6] Ren Y, Suganthan P N, Srikanth N, et al. Random vector functional link network for short-term electricity load demand forecasting[J]. Information Sciences, 2016, 367(C):1078-1093.

[7] Srivastava N, Mansimov E, Salakhutdinov R. Unsupervised learning of video representations using LSTMs[C]//International Conference on International Conference on Machine Learning. JMLR.org, 2015:843-852.

[8] Furquim G, Pessin G, Faiçal B S, et al. Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory[J]. Neural Computing & Applications, 2016, 27(5):1129-1141.

[9] Alkhasawneh M S, Tay L T, Ngah U K, et al. Intelligent landslide system based on discriminant analysis and cascade-forward back-propagation network[J]. Arabian Journal for Science and Engineering, 2014, 39(7): 5575-5584.

[10] Arroyo J, Maté C. Forecasting histogram time series with k-nearest neighbours methods[J]. International Journal of Forecasting, 2009, 25(1):192-207.