2021 20th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Improved BiLSTM Model For Online Food Safety Risk Prediction

1st Zekai Zhang
School Of Artificial Intelligence and Computer Science,Jiangnan University
WuXi ,China
E-mail: 971988947@qq.com

3rd Jun Sun
School Of Artificial Intelligence and Computer Science,Jiangnan University
WuXi ,China
E-mail: junsun@jiangnan.edu.cn

2nd Mao Li
School Of Artificial Intelligence and Computer Science,Jiangnan University
WuXi ,China
E-mail: wxmaoli@163.com

3rd Qidong Chen
School Of Artificial Intelligence and Computer Science,Jiangnan University
WuXi ,China
E-mail: cqd_mu@hotmail.com

***Abstract*—The prediction of security risk is a new problem in the field of modeling. The value of food safety risk is dynamic, non-linear and non-stationary. In order to improve the prediction accuracy of non-stationary sequences, this paper proposes an improved time series prediction model of bidirectional long short-term memory network based on NS-ADAM optimization algorithm. By using NS-ADAM optimization algorithm, the non-stationary series can be approximately regarded as stationary series in a fixed window length. Combined with the advantages of two-way long-term and short-term network model to strengthen the correlation between current data and historical data, the combined network model significantly improves the prediction accuracy of non-stationary time series. The experimental results show that compared with the traditional integrated moving average autoregressive model and the conventional long-term and short-term memory network model, the improved model has higher prediction accuracy and better fitting effect for non-stationary series.**

Keywords-network food safety risk prediction;optimization algorithm;bidirectional long short-term memory network

I. INTRODUCTION

In recent years, food safety issues [1] have always maintained a high degree of hot spots at home and abroad. With the rapid development of the Internet economy, the proportion of online ordering methods in the consumer's dietary consumption patterns has shown a leap-forward growth, but the corresponding systems and laws and regulations of the online ordering industry have not been able to adapt to the rapid development of the industry. Online food safety issues continue to emerge, arousing increasing social concern. Most scholars at home and abroad have studied such issues mainly from the perspectives of social management and economics[2], and put forward their thoughts on strengthening the supervision of the online food ordering industry from the three aspects of the government, the food delivery platform[3], and the consumer. And suggestions. However, there is almost no research on the risk prediction of online food safety based on the modeling field. This article takes the long-term food safety risk value of a certain business in a city as an example, uses neural network methods to predict online food safety risks [4], and provides technical support for online food safety issues from an objective and scientific point of view.

II. BACKGROUND INTRODUCTION

A. LongShort-term Memory Network

LSTM[5-7] is extended on the basis of RNN[8], using "Gating" unit to keep the model and the long-term dependence of the learning input. The LSTM model consists of three gates, called Input Gate, Output Gate, and Forget Gate. The Forget gate is used to retain or remove the existing information, the Input Gate specifies the extent to which the new information enters the memory, and the Output Gate finally determines the value we will output based on all the information currently received. The gate operation is mainly through a Sigmoid neural layer and a point-by-point multiplication operation to achieve the screening of information.

B. Bidirectional LongShort-term Memory Network

BiLSTM[9-10] is an extended model of LSTMs, in which there are two LSTMs in the input data part. In the first round, LSTM is applied to the input sequence. In the second round, the inverse form of the input sequence is entered in the LSTM model. Applying LSTM twice can improve the long-term dependence performance of learning, thereby improving the accuracy of the model.

2473-3636/21/$31.00 ©2021 IEEE
DOI 10.1109/DCABES52998.2021.00052

III. IMPROVED BILSTM BASED ON NEW OPTIMIZATION ALGORITHM

A. *Adaptive Moment Estimation and Optimization Algorithm NS-ADAM for Non-stationary Sequence*

In recent years, many methods[11-12] based on stochastic gradient descent have been proposed for neural network models. The gradient formula is:

$$g_t = \nabla_\theta L(f(x_t;\theta), y_t) \qquad (1)$$

Among them, L is the loss function, and the most common one is the mean square error criterion. The model continuously updates the value of the parameter according to the given data sequence. The update formula of the basic stochastic gradient descent algorithm is as follows:

$$\theta_t = \theta_{t-1} - \alpha g_t \qquad (2)$$

Unlike the previous variable update based on the data of the current iteration step, Adam updates the variables θ based on the gradient estimation of all data points before the current time step t. Adam's update rules are as follows:

$$\theta_t = \theta_{t-1} - \alpha \hat{m}_t / \hat{v}_t \qquad (3)$$

The $\hat{m}_t = m_t(1-\beta_1^t)$ is the moment estimation version of m_t. It is the weighted sum of current and history variables g_t, which can be regarded as an estimate of expectations. The update rules are as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\cdot g_t \qquad (4)$$

Similarly, $\hat{v}_t = v_t(1-\beta_2^t)$ is the moment estimation version of v_t, which is the weighted sum of current and historical variables $g_t \odot g_t$. The update rules are as follows:

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\cdot g_t \odot g_t \qquad (5)$$

When the sequence is a non-stationary sequence, the statistical variables change with time t, and it is unreasonable to use the gradient far from the current step to determine the update direction of the current step. Yulai Zhang [13] studied this type of problem and added a window variable k. Assume that in sequence Y, any sub-sequence of length $k \ll T$ has stable statistical variables, so it can be approximately regarded as a stationary sequence[14]. The renewed update rules are as follows:

$t \le k$:

$$\hat{m}_t = m_t(1-\beta_1^t) \qquad (6)$$

$$\hat{v}_t = v_t(1-\beta_2^t) \qquad (7)$$

$t > k$:

$$\overline{m}_t = m_t - \beta_1^k m_{t-k} \qquad (8)$$

$$\overline{v}_t = v_t - \beta_2^k v_{t-k} \qquad (9)$$

$$\hat{m}_t = \overline{m}_t(1-\beta_1^k) \qquad (10)$$

$$\hat{v}_t = \overline{v}_t(1-\beta_2^k) \qquad (11)$$

Finally:

$$\theta_t = \theta_{t-1} - \alpha \hat{m}_t / (\sqrt{\hat{v}_t}) \quad (12)$$

B. *BiLSTM Optimized Based On NS-ADAM*

The static window length k value in the NS-ADAM algorithm is artificially specified, and the choice of the k value plays a key role in the realization of the NS-ADAM algorithm. However, no scholar has stated how to find the best fixed window length method. This variable has a clear correlation with the data type itself, and it also has a significant impact on the accuracy of the prediction. In this paper, the general evaluation index MSE of the model is used to judge the algorithm under different window length k, and k-MSE is drawn into a graph. The lowest point of the curve can be designated as the optimal window length of the data set. In Figure 1, the mean square error under different window length k is given, and the model is carried out under the number of training times of 1200, and other parameters remain unchanged. When k is around 20, the fitting effect of the model is the best.

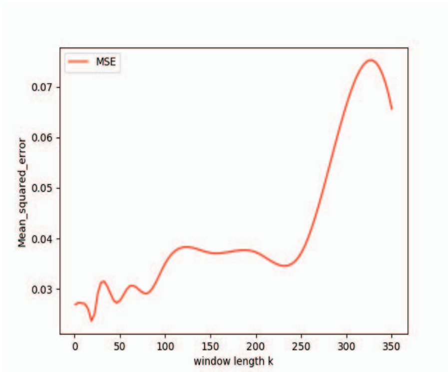

Figure 1 MSE of NS-ADAM algorithm

The NS-ADAM algorithm is based on the Adam algorithm and adds a static fixed window length k. At the same time, it proposes the assumption that items within k lengths can be approximately regarded as a stationary sequence, which enhances the correlation between the step direction of the gradient update and the previous gradient Sex. BiLSTM is an extended model of LSTM, which can be seen as a two-layer neural network. The first layer is from left to right as the input sequence of the model data sequence, and the second layer is from right to left as the sequence input sequence. This paper also uses the previously tested NS-ADAM optimization algorithm with a window length of 20 in the model to iteratively update the weights. The general structure is shown in Figure 2

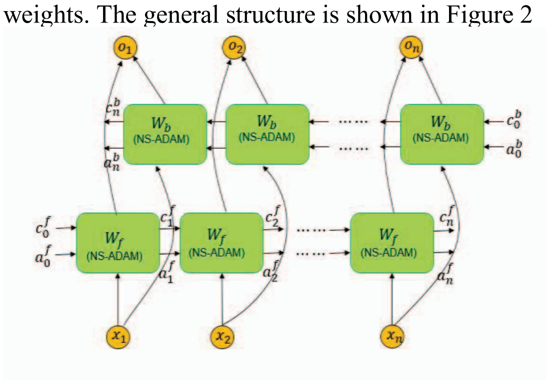

Figure 2 The new structure of BiLSTM

IV. COMPARATIVE ANALYSIS OF EXPERIMENTS

A. Data Introduction Preprocessing

This paper uses 46 key cities across the country as the data sampling area, and collects the order review data of some online catering merchants in the city's location, and the data is of the order of one million. After that, a risk assessment model was established with analytic hierarchy process and natural language methods as the main technical core, and the online food risk assessment value of each business was obtained. The author selects the risk value of a certain store from the final total data set as the data set for this article.

The original data is the food safety risk value of the merchant's network, and the data unit is generally between $10^{-2} \sim 10^{-4}$. In order to train the model better, the data is normalized and preprocessed. The normalization process maps the original data to the range [0,1]. The normalization formula[15] is shown in formula(13). x is the original data value, x' is the normalized data value.

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}) \tag{13}$$

B. Model Evaluation Criteria

The loss value in the model is usually reported by the deep learning algorithm. Objectively speaking, the loss value is a punishment for poor prediction. If the model's prediction is completely accurate, then the loss value will be zero. Therefore, the model minimizes the loss as much as possible by obtaining a set of weights and deviations that minimize the loss value. In addition to the loss of optimization algorithms in deep learning, researchers often use the root mean square error (RMSE) as an indicator to judge the predictive performance of the model. The root mean square error represents the deviation of the observed value from the correct value. The formula is as follows:

$$RMSE(X, h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(x_i) - y_i)^2} \tag{14}$$

In the formula, m is the total number of samples, h is the predicted value, and y is the actual value. The main benefit of using RMSE is that it penalizes larger mistakes. In addition, we also use the percentage of RMSE reduction as an indicator to evaluate improvement. The calculation formula is as follows:

$$\%Reduction = \frac{New\_Value - Origina\_Value}{Original\_Value} \tag{15}$$

C. Comparison With Traditional Optimization Algorithms

This paper uses the BiLSTM neural network model to predict and analyze non-stationary sequences based on the food safety risk value data of the merchant network. The experiment compares the new optimization algorithm NS-ADAM and the traditional optimization algorithm[16-20] Adam, RMSprop, SGD under different batch sample numbers. Mean square error.

TABLE I. RMSE UNDER DIFFERENT ALGORITHMS UNDER BILSTM MODEL

Batch_size	SGD	RMSprop	Adam	Ns-ADAM
128	185.841	166.435	**78.644**	99.587
256	269.261	171.889	170.568	**142.208**
350	151.838	128.369	130.744	**89.552**
512	239.220	185.524	264.929	**167.801**
Average	211.540	163.054	161.221	**124.787**

TABLE II. COMPARISON WITH TRADITIONAL ALGORITHMS

Algorithm Comparison	%Reduction
NS-ADAM OVER SGD	-41.01
NS-ADAM OVER RMSPROP	-23.47
NS-ADAM OVER ADAM	-22.60

It can be seen from Table I of the experimental results that NS-ADAM has the smallest average root mean square error under the condition of the number of batch training samples from 128 to 512, and the best results are also obtained in most cases. The model has the best training effect when the number of batch samples is about 350. Table II reports the model improvement degree of NS-ADAM compared with the traditional optimization algorithm. Compared with SGD, the evaluation index of NS-ADAM has dropped by as much as 41.01%, and for the RMSprop and Adam algorithms, it has also dropped by more than 20%. Experiments prove that the BiLSTM model based on the NS-ADAM optimization algorithm significantly improves the prediction accuracy than the traditional BiLSTM model.

D. Comparison with traditional models

For the rigor of the experiment, this paper compares the BiLSTM model with the traditional LSTM network model and the integrated moving average autoregressive model ARIMA model. In this paper, the original data set is normalized and preprocessed, and the processed data is input into the ARIMA, LSTM, and BiLSTM models as the only feature of the risk assessment sequence. The data set is divided into two parts: training and testing. 90% of the data set is used for training and 10% of the data set is used to test the accuracy of the model. Table III shows the model comparison of various optimization algorithms under the LSTM model and the comparison with the BiLSTM and ARIMA models.

TABLE III. RMSE OF ARIMA, LSTM AND BILSTM MODELS

Model	RMSE
ARIMA	345.262
LSTM_SGD	205.684
LSTM_RMSPROP	168.925
LSTM_ADAM	166.978
LSTM_NSADAM	153.852
BILSTM_NSADAM	89.5524

Table III reports the root mean square error achieved by each model that predicts the risk value of online food safety for merchants. It can be clearly seen that the LSTMs-based recurrent neural network model has a significantly lower RMSE value than the traditional autoregressive model ARIMA. In the one-way LSTM model, the RMSE value of the LSTM using the NS-ADAM optimization algorithm is also smaller than that of the one-way LSTM model using other optimization algorithms. And it can be seen that the performance of the BiLSTM model is better than the one-way LSTM model, and the gap is very large. In summary, the performance of the BiLSTM model is the best among the three types of models, and the BiLSTM model based on the NS-ADAM optimization algorithm performs best under various optimization algorithms. The experiment also compares the training time of BiLSTM and LSTM models. The average training time of the BiLSTM model is about 700s, and the average training time of the LSTM model is about 470s. Although the accuracy of the BiLSTM model is improved compared to the LSTM model, the time performance is reduced. Given that the internal structure of the BiLSTM model is more complex than that of the LSTM model, this experimental result is reasonable. Figures 3, 4, and 5 are the prediction diagrams of the food safety risk values of the three models.

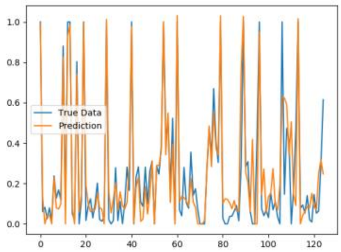

Figure 3 BiLSTM

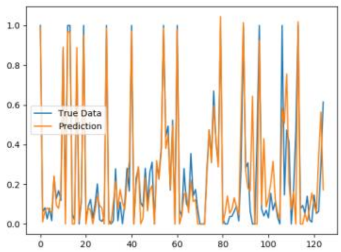

Figure 4 LSTM

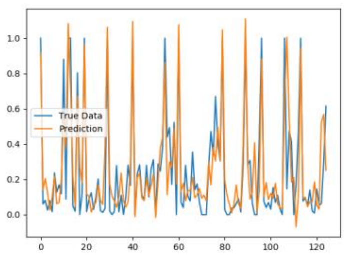

Figure 5 ARIMA

V. SUMMARY AND OUTLOOK

This article reports the results of an experiment in which the performance and accuracy of various optimization algorithms based on the BiLSTM model are compared and analyzed in the online food safety risk data prediction experiment. At the same time, analysis and comparison between different models of ARIMA, one-way LSTM and BiLSTM are carried out. The research question for this experiment focuses on the prediction of non-stationary series. Does the application of the NS-ADAM optimization algorithm in the LSTMs model have any positive impact on improving the accuracy of the time series? The results show that the application of the NS-ADAM algorithm helps the model improve the accuracy of prediction, and the effect is significant, which is conducive to modeling. We noticed that the NS-ADAM algorithm has significant advantages in either the one-way LSTM or the BiLSTM model. Compared with the LSTM model, the accuracy of the BiLSTM model is improved, but the training speed is relatively slow. When time conditions permit, this article recommends using the BiLSTM model combined with the NS-ADAM algorithm to predict non-stationary series. This research can be further extended to forecasting problems for multivariate and seasonal time series.

REFERENCES

[1] Zhang Shukai,Lei Xin. Internet Takeaway Food Safety Regulatory lssues[J].MODERN FOOD, 2016, 000(003):P.41-45.

[2] Wahyuni H C, Vanany I, Ciptomulyono U. Application of Bayesian Network for Food Safety Risk in Cattle Slaugtering Industry[C].2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE, 2019: 450-454.

[3] Ji Lihong, Chen hong. Problems in the supervision of online food delivery industry and their governance [J]. WESTERN SO, 2019(31)

[4] Cai Yuanzheng. The Dilemma and Outlet of Online Food Safety Supervision [J]. LEGALITY VISION, 2017 (3): 11.

[5] Kim J , Moon N . BiLSTM model based on multivariate time series data in multiple field for forecasting trading area[J]. Journal of Ambient Intelligence and Humanized Computing, 2019(5).

[6] Cui Z , Ke R , Pu Z , et al. Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction[J]. 2018.

[7] Lee S I , Yoo S J . Threshold-Based Portfolio: The Role of the Threshold and Its Applications[J]. Papers, 2018.

[8] Krauss, Christopher, Anh Do, Xuan, Huck, Nicolas. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500[J]. European Journal of Operational Research, 2016:S0377221716308657.

[9] De Baets L, Ruyssinck J, Peiffer T, et al. Positive blood culture detection in time series data using a BiLSTM network[J]. arXiv preprint arXiv:1612.00962, 2016.

[10] Sha S, Li J, Zhang K, et al. RNN-Based Subway Passenger Flow Rolling Prediction[J]. IEEE Access, 2020, 8: 15232-15240.

[11] Liu J, Qian L, Zhang Y, et al. Towards Safety-Risk Prediction of CBTC Systems With Deep Learning and Formal Methods[J]. IEEE Access, 2020, 8: 16618-16626.

[12] Siami-Namini S, Tavakoli N, Namin A S. The performance of LSTM and BiLSTM in forecasting time series[C]//2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019: 3285-3292.

[13] Bukhari A H, Raja M A Z, Sulaiman M, et al. Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting[J]. IEEE Access, 2020, 8: 71326-71338.

[14] Zhang Y, Wang Y, Luo G. A new optimization algorithm for non-stationary time series prediction based on recurrent neural networks[J]. Future Generation Computer Systems, 2020, 102: 738-745.

[15] Zhang Z. Improved Adam optimizer for deep neural networks[C].2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). IEEE, 2018: 1-2.

[16] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[17] Liu H, Xu H, Yan Y, et al. Bus Arrival Time Prediction Based on LSTM and Spatial-Temporal Feature Vector[J]. IEEE Access, 2020, 8: 11917-11929.

[18] Kalchbrenner N, Danihelka I, Graves A. Grid long short-term memory[J]. arXiv preprint arXiv:1507.01526, 2015.

[19] Zhao Y. Spatial-Temporal Correlation-Based LSTM Algorithm and Its Application in PM2. 5 Prediction Spatial-Temporal Correlation-Based LSTM Algorithm and Its Application in PM2. 5 Prediction[J].

[20] Li Y, Bao T, Gong J, et al. The Prediction of Dam Displacement Time Series Using STL, Extra-Trees, and Stacked LSTM Neural Network[J]. IEEE Access, 2020, 8: 94440-94452.