2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) | 978-1-7281-9638-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/MLBDBI51377.2020.00032

An Online Retail Prediction Model Based on AGA-LSTM Neural Network

Keqiao Chen
Department of Computer Science and Engineering
Pennsylvania State University
University Park, the United States
kqc5533@psu.edu

***Abstract*—With the development of e-commerce, the scale of online retail is becoming larger and larger. The prediction of sales volume is facing the challenges of various categories and changeable customer demands. Based on this, this paper proposes an online retail prediction model based on AGA-LSTM neural network. In the process of building LSTM neural network, the adaptive genetic algorithm (AGA) is used to optimize the network parameters such as time step, hidden layer number and training times to improve the prediction accuracy of the model, and it is used to forecast the types of goods and the total sales volume. The results on the Online Retail II dataset in UCI show that the prediction accuracy of AGA-LSTM model is greatly enhanced compared with the traditional LSTM model, which verifies the effectiveness of this algorithm.**

Keywords- online retail prediction; adaptive genetic algorithm; parameter optimization; sales; AGA-LSTM

I. Introduction

In the information age, the rapid development of mobile Internet promotes the popularity of e-commerce. With the emergence of e-commerce industry, online retail has gradually become one of the main ways of people's consumption. On such well-known e-commerce platforms as Amazon and Taobao, there are a wide variety of goods, which brings great convenience to the society. Compared with the traditional sales industry, e-commerce has more products, shorter update period and more dynamic and complex. How to extract potentially useful information from massive e-commerce data is a hotpot. Finding suitable data mining methods to achieve high-accuracy prediction of sales volume and types of goods is the focus in the field of e-commerce, which is also conducive to the rapid development of economic society. The essence of online retail sales prediction lies in the accuracy. If the predicted value is too high, the inventory will be overstocked and the cost will rise; if the predicted value is too low, the online consumers will be dissatisfied and the phenomenon of "supply exceeding demand" will appear.

In recent years, machine learning has made great achievements in speech recognition[1], natural language processing[2,3], image processing[4] and market analysis[5]. In order to solve the prediction problem in e-commerce, various sales prediction methods based on machine learning have been proposed one after another. In 2015, Murray used data mining technology to identify online customers with similar demand, and predicted the commodity purchase volume of each type customer[6]. In 2016, Jiang X established a water resources demand forecasting model by using the BP neural network, which realized accurate demand prediction and guaranteed the advance supply of water resources[7]. In 2017, Salinas D proposed a LSTM neural network model based on generation probability, namely deep AR model[8], to forecast the online sales of Amazon. In 2018, Mukherjee S proposed an ARMDN neural network model[9], which can simulate the associated variables, time series and demand at the same time, so as to realize the prediction of e-commerce retail.

However, most of the above methods are based on the traditional machine learning or neural network model to predict the sales volume of goods, ignoring the shortcomings of the algorithm itself, and can no longer meet the prediction and analysis of complex online sales data. Therefore, this paper proposes an online retail prediction model based on AGA-LSTM neural network. The model uses adaptive genetic algorithm (AGA) to optimize the network parameters of LSTM neural network, such as the time step, the number of hidden layers and training times, to overcome the shortcomings of LSTM itself, such as the disappearance of gradient and high training cost. It can improve prediction accuracy of the model, which could be used to predict the total sales volume of online retail.

II. Adaptive Genetic Algorithm and Lstm Neural Network

Swarm intelligence algorithms have been widely used to optimize the parameters of neural network models. Among them, adaptive genetic algorithm(AGA) [10] is an optimized genetic algorithm, which was proposed in 1994. Different from the standard genetic algorithm, AGA algorithm can protect the excellent individuals in the population as much as possible, inhibit the occurrence of premature phenomenon, and can jump out of the local optimum, and the optimization performance is better. Therefore, we use the adaptive genetic algorithm to construct the optimization network, adjust and optimize the parameters of the LSTM neural network, such as the time step, the number of hidden layers and the training times, obtaining the LSTM neural network with the optimal parameters, and its prediction performance is better.

A. Adaptive Genetic Algorithm

The process of adaptive genetic algorithm is basically the same as that of genetic algorithm, including initialization of population space, selection and replication, adaptive crossover,

978-1-7281-9638-1/20/$31.00 ©2020 IEEE
DOI 10.1109/MLBDBI51377.2020.00032

adaptive mutation, iteration and so on. In the AGA algorithm, the crossover probability and mutation probability can be changed in real time according to the iteration of the population, to ensure the diversity and diversity of individuals and avoid falling into local optimum. The derivation formulas of crossover probability and mutation probability are shown in (1) and (2).

$$P_c = \begin{cases} \dfrac{k_1(f_{\max} - f^{'})}{f_{\max} - f_{avg}} & f^{'} \geq f_{avg} \\ k_2 & f^{'} < f_{avg} \end{cases} \quad (1)$$

$$P_m = \begin{cases} \dfrac{k_3(f_{\max} - f)}{f_{\max} - f_{avg}} & f \geq f_{avg} \\ k_4 & f < f_{avg} \end{cases} \quad (2)$$

Among them, P_c is the crossover probability, P_m is the mutation probability, f_{avg} is the average fitness of individual population, $f_{\max}$ is the maximum fitness, $f^{'}$ is the larger appropriate value in the crossover operation, f is the mutation individual, k_1, k_2, k_3 and k_4 are the parameters between 0 and 1.

By adjusting the parameters k_1, k_2, k_3 and k_4, in the early iteration process, the crossover probability and mutation probability should be appropriately reduced, so the individuals with larger fitness in the population could be retained, and the iteration should be completed as soon as possible to reduce the training cost; in the later iteration process, because the population tends to be stable, AGA algorithm can increase the crossover probability and mutation probability appropriately to avoid the local optimization and find the optimal individual in the global.

B. *The LSTM Neural Network*

The LSTM neural network is a kind of recurrent neural networks, which allows the continuous existence of time series information, can learn the information dependence between long time series, and is very sensitive to time series data. The LSTM model is a network structure composed of multiple hidden layers, input layers and output layers. Each hidden layer contains many LSTM neurons. And each LSTM neuron has three gated units, which are forget gate, input gate and output gate. It is the special internal structure that enables the LSTM model to learn and memorize time series data, predict the future trend with high precision. The LSTM model can be widely used in common time series scenarios such as marketing, financial analysis, etc.

The structure of LSTM neuron is shown in Fig. 1. $Input\_Gate$, $Output\_Gate$ and $Forget\_Gate$ are input gate, output gate and forgetting gate respectively. Let the input sequence data be (x_ 1, x_ 2, … X_ t), at time t, the derivation formula of each gated unit of LSTM neuron is as follows:

$$i_t = S(W_i * [h_{t-1}, x_t] + b_i) \quad (3)$$

$$f_t = S(W_f * [h_{t-1}, x_t] + b_f) \quad (4)$$

$$o_t = S(W_0 * [c_t, h_{t-1}, x_t] + b_0) \quad (5)$$

$$c_t = f_t * c_{t-1} + i_t * \tanh(W_c * [h_{t-1}, x_t]) \quad (6)$$

$$h_t = o_t * \tanh(c_t) \quad (7)$$

Among them, i_t, f_t and o_t are the inputs of input gate, forget gate and output gate at time t; x_t is the input at time t, h_{t-1} represent the output state of the hidden layer at time t-1; W_i, W_f and W_0 are the weights of the input and the three gated units respectively, b_i, b_f and b_o are the corresponding offset amount; W_c is the weight between the input and neurons; then h_t represents the output of the hidden layer at time t; S and tanh are activation functions.

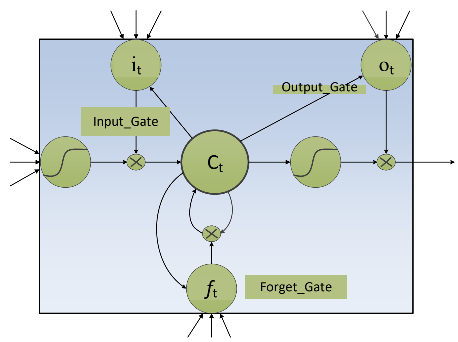

Figure 1. The structure of LSTM neuron.

III. ONLINE RETAIL PREDICTION MODEL BASED ON AGA-LSTM NEURAL NETWORK

This paper makes full use of the ability of adaptive genetic algorithm to optimize parameters of LSTM neural network, such as the time step, the number of hidden layers and training times, so as to make the parameters more optimal and effectively improve the time series prediction ability of the model. Then, an online retail prediction model based on AGA-LSTM neural network is constructed and applied to the daily online sales forecast.

A. *Chromosome Design*

In the AGA-LSTM model, the design of chromosomes is particularly important. A chromosome represents an individual in the population. We design the chromosome as a set of three parameters in the LSTM neural network, including time step, the number of hidden layers and training times, as shown in Fig. 2. Where, Ts is the time step parameter, n is the number of hidden layers, and epochs is the training times of data.

	Time steps	Number of hidden layers	Training times
Chromosome	Ts	N	Epochs

Figure 2. Chromosome design.

B. Chromosome Fitness

The fitness of a chromosome, that is, the fitness of the individual represented by the chromosome, can be used to evaluate the importance of each individual in the population. In the online retail prediction model based on AGA-LSTM neural network, the mean square error(MSE) between the predicted value and the actual value of LSTM neural network is designed as the fitness of chromosomes. The smaller the value, the more reasonable the parameter value of LSTM neural network is, and the better the prediction performance of the model is.

$$MSE_i = \frac{1}{num}\sum_{j=1}^{num}\left(P_{ij} - A_{ij}\right)^2 \tag{8}$$

Set MSE_i as the mean square error of chromosome i, the calculation formula is shown in (8), num is the number of test sets, P_{ij} is the prediction result of LSTM neural network corresponding to chromosome (the set of parameters), and A_{ij} is the actual value of sample.

C. The Process of AGA-LSTM Online Retail Prediction Model

In the online retail prediction model based on AGA-LSTM neural network, the adaptive genetic algorithm maps the three parameter values of LSTM neural network (time step parameter Ts, hidden layer number N, data training times Epochs), so that each chromosome becomes the parameters combination of the whole LSTM model. In the process of training the LSTM model, it iterates continuously to find the optimal parameters set of the model. The AGA-LSTM model is constructed to predict the total sales of online retail.

Fig. 3 is the flow chart of AGA-LSTM model. Let m be the current number of iterations and M is the maximum number of iterations. The specific process of the algorithm is as follows:

1) The online retail II dataset in UCI database is obtained and cleaned, normalized and partitioned.

2) The parameters of algorithms are initialized and the AGA optimization network is constructed. For example, the training times of LSTM neural network, the network structure and the maximum iteration times of AGA algorithm are set.

3) Training the LSTM neural network corresponding to each chromosome parameter combination, namely AGA-LSTM model.

4) AGA algorithm is used to optimize these parameters. The adaptive genetic algorithm can be used to optimize the parameters such as Ts, N and Epochs to generate new individuals.

5) The fitness of each chromosome was calculated to update the global optimal individual.

6) Determine whether the maximum number of iterations has been reached. Otherwise, the iteration is terminated and the global optimal individual is output, which represents the optimal parameter combination of LSTM model.

(7) The global optimal combination of parameters is output and brought into the LSTM model. The weights of the model are retrained and adjusted. The online retail prediction model based on AGA-LSTM neural network is successfully constructed.

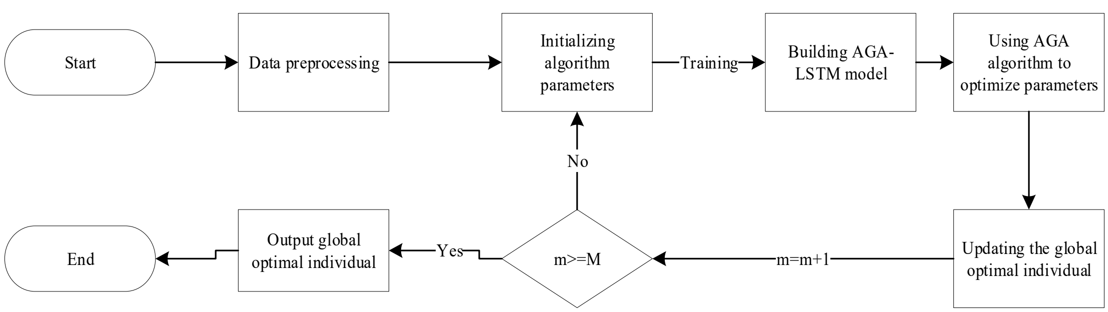

Figure 3. The process of AGA-LSTM model.

IV. Experiment and Analysis

A. Dataset Selection

The online retail II data set is selected as the experimental data set, which is from UCI database, which is a transnational data set. It records all online transaction order information of online retailers in the UK from December 1, 2010 to December 9, 2011, including customer number, order number, total type of goods, total sales volume and unit price of goods. The company mainly sells gifts and has many wholesaler customers. This paper mainly forecasts the daily sales volume and total types of goods in the company's online retail.

B. Modeling

In time series prediction, if the state of the next n times is predicted, it is n-step prediction. When n = 1, it is a single step prediction, when n > 1, it is a multi-step prediction. Based on deep BP, LSTM and AGA-LSTM neural networks, this experiment established online retail prediction models to predict the sales volume of future one-day, five-day and ten-day of the British online retailer. The prediction accuracy of each model is compared to verify the performance of AGA-LSTM model.

After data preprocessing, we take 70% of the dataset as the training set, 20% as the verification set, and 10% as the test set.

$$MSE = \sum_{j=1}^{n_t} \left(y_i^{'} - y_i \right)^2 / n_t \tag{9}$$

The evaluation index of model is set as mean square error MSE, seen from formula (9). Where, n_t is the number of test sets, $y_i^{'}$ is the prediction result, y_i is the true value of the test sample. The smaller the sum of error leveling is, the better the performance of the model is and the smaller the prediction error is. It should be noted that for the MSE value under multi-step prediction, both the real value and the predicted value of MSE are the average values in the future, and then calculated according to formula (9).

C. Results Analysis

TABLE I. THE MSE VALUE OF THREE MODELS UNDER DIFFERENT PREDICTION STEPS. (TOTAL DAILY SALES VOLUME)

Prediction step	The prediction model		
	BP	***LSTM***	***AGA-LSTM***
One-step	39.6	26.8	18.7
Five-step	50.7	32.3	25.6
Ten-step	62.3	48.5	37.3
Average MSE value	50.9	35.9	27.2

Table 1 shows the MSE values of three online retail prediction models based on BP, LSTM and AGA-LSTM under the single step, 5-step and 10-step prediction of the total daily sales volume.

It can be seen from table 1 that the MSE values of BP model in single-step, 5-step and 10-step prediction are the largest, followed by LSTM model. The MSE values of online retail prediction model based on AGA-LSTM neural network are the lowest, which shows that the prediction error of this model is the smallest, and the prediction of total sales volume is the closest to the actual value, which verifies the model.

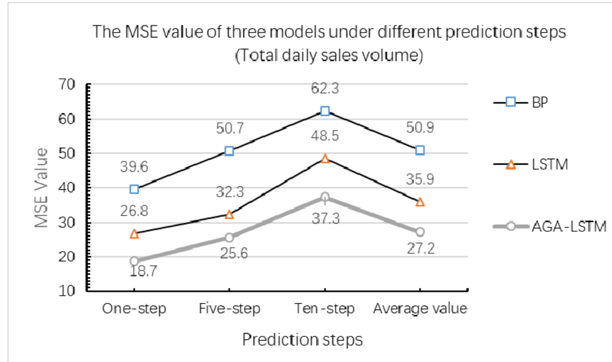

Figure 4. The MSE value of three models under different prediction steps. (Total daily sales volume)

Fig. 4 shows the change trend of MSE values and the average values of three models under single step, 5-step and 10-step prediction. In the case of three prediction steps, the average MSE of LSTM model is 29.5% lower than that of BP model, and that of AGA-LSTM model is 24.2% lower than that of traditional LSTM model. This proves that AGA algorithm can promote the parameter adjustment of LSTM neural network to a certain extent, and can significantly improve the performance of time series prediction, and realize the prediction of total daily sales volume.

Table 2 shows the MSE values of three models based on BP, LSTM and AGA-LSTM under the single step, 5-step and 10 step prediction of total types of goods. Figure 5 shows the trend of MSE value and average value of total goods types predicted by three models drawn in Table 2.

TABLE II. THE MSE VALUE OF THREE MODELS UNDER DIFFERENT PREDICTION STEPS.(TOTAL TYPES OF GOODS)

Prediction step	The prediction model		
	BP	***LSTM***	***AGA-LSTM***
One-step	13.5	9.6	5.7
Five-step	19.1	21.3	10.3
Ten-step	28.9	25.9	16.4
Average MSE value	20.5	18.9	10.8

It can be seen from table 2 and Fig. 5 that the MSE values of AGA-SLTM model proposed in this paper are the lowest under each prediction step. The average MSE value of AGA-LSTM model is 47.3% lower than that of BP model and 42.8% lower than LSTM model. It shows that AGA-LSTM model has the highest prediction accuracy for total types of goods, and the algorithm performance is better than the other two models. However, with the increase of prediction steps, the prediction accuracy of three models decreased to a certain extent.

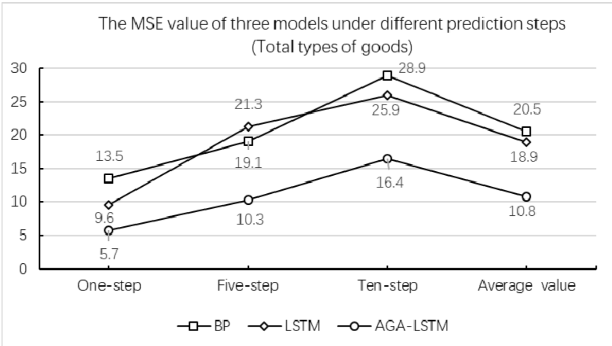

Figure 5. The MSE value of three models under different prediction steps. (Total types of goods)

V. Experiment and Analysis

Aiming at the difficulty in forecasting the sales volume of goods in e-commerce, this paper proposes an online retail prediction model based on AGS-LSTM neural network. The adaptive genetic algorithm is used to optimize the parameters of LSTM model, and the model can predict the types of goods and total sales volume of online retail. The results on the online retail II dataset show that the prediction error of AGA-LSTM model is greatly improved compared with the traditional BP and LSTM models, which verifies the effectiveness of this algorithm. However, with the increase of prediction step sizes, the accuracy of the model continues to decline. Therefore, we will focus on the research of multi-step prediction in the future.

Acknowledgment

For the successful completion of this paper, I would like to thank my parents for their great support to me, as well as the provider of experimental data: University of California Irvine.

References

[1] Lee M, Lee J, Chang J H. Ensemble of jointly trained deep neural network-based acoustic models for reverberant speech recognition[J]. Digital Signal Processing, 2019, 85: 1-9.

[2] Wei W , Wu J , Zhu C . Special issue on deep learning for natural language processing[J]. Computing, 2020, 102(3):601-603.

[3] Zhou G B, Wu J, Zhang C L, et al. Minimal gated unit for recurrent neural networks[J]. International Journal of Automation and Computing, 2016, 13(3): 226-234.

[4] Jung M, Lee H, Tani J. Adaptive detrending to accelerate convolutional gated recurrent unit training for contextual video recognition[J]. Neural Networks, 2018, 105: 356-370.

[5] Jiang X, Pan S, Jiang J, et al. Cross-domain deep learning approach for multiple financial market prediction[C]//2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018: 1-8.

[6] Murray P W, Agard B, Barajas M A. Forecasting supply chain demand by clustering customers[J]. IFAC-PapersOnLine, 2015, 48(3): 1834-1839.

[7] Jiang X, Chen W. Comparison between BP neural network and GA-BP crop water demand forecasting model J[J]. Journal of Drainage and Irrigation Machinery Engineering, 2018: 762-766.

[8] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: Probabilistic forecasting with autoregressive recurrent networks[J]. International Journal of Forecasting, 2020, 36(3): 1181-1191.

[9] Mukherjee S, Shankar D, Ghosh A, et al. Armdn: Associative and recurrent mixture density networks for eretail demand forecasting[J]. arXiv preprint arXiv:1803.03800, 2018.

[10] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1994, 24(4): 656-667.