Journal of Network Intelligence
Taiwan Ubiquitous Information

Volume **7**, Number **2**, May **2022**

A Prediction Method of Consumer Buying Behavior Based on Attention Mechanism and CNN-BiLSTM

Jian-Nan Wang

College of Computer and Information Engineering
Xiamen University of Technology
Xiamen,361024,China
wangjiannan98@163.com

Jian-Feng Cui*

College of Software Engineering
Xiamen University of Technology
Xiamen,361024,China
jfcui@xmut.edu.cn

Chin-Ling Chen*

1 School of Computer and Information Engineering,
Xiamen University of Technology, Xiamen 361024, China.
2 School of Information Engineering,
Changchun Sci-Tech University, Changchun 130600, China.
3 Department of Computer Science and Information Engineering,
Chaoyang University of Technology, Taichung 41349, Taiwan.
clc@mail.cyut.edu.tw

Corresponding Author: Jian-Feng Cui, Chin-Ling Chen

jfcui@xmut.edu.cn,clc@mail.cyut.edu.tw

Received November 2021; revised January 2022

Abstract. *A prediction method of consumer buying behavior based on attention mechanism and CNN-BiLSTM is proposed, to fully extract the effective features of the behavior and make its prediction more accurate and stable. In this paper, the CNN model is used to automatically extract local features of the consumer buying behavior to select and optimize important information. The BiLSTM model is used to collect sequence patterns of buying behavior to find the correlation among features. The attention mechanism is used to assign different weights to the state values of the hidden layers of each neural network so that the hidden layers can better express important information to obtain more accurate and stable predictions. The experiment results demonstrate that the proposed prediction method has advantages in not only accuracy but stability over other models.*
Keywords: Buying behavior, Deep learning, feature selection, Attention mechanism, CNN-BiLSTM

---

1. **Introduction.** With the continuous development of e-commerce platforms, the amount of information increases exponentially, which leads to serious information overload [1–3]. How to find Commoditys that meet our own needs efficiently has become a major problem. Consequently, recommendation systems came into being, which can improve consumer experience and marketing results by optimizing the consumer decision-making process [4]. Prediction is the premise of recommendation, so that accurate prediction of the consumer

purchasing tendency can improve the recommendation effect and optimize the system. In the early stage, due to the immaturity of e-commerce platforms and the lack of data related to consumer buying behavior, the research on its prediction has been stagnant for a long time. Recently, with the rapid development of information technology, plenty of data on consumer buying behavior have been gathered on the platform, which provides strong support for the prediction research [5].

Machine learning models are widely used to predict consumer buying behavior [6]. Early researchers introduced new methods in classical machine learning models to achieve better prediction results. For example, Tang [7] proposes a regression prediction method based on support vector machine, which improves the accuracy of user loyalty prediction. Although improving the model can enhance its performance, a single model's capacity to generalize across multiple datasets is restricted. To improve the generalization ability of the model and obtain better prediction effect, ensemble learning is widely used in consumer purchase behavior prediction. Qiu et al. [8] utilized XGBoost to predict consumer behavior and assess risk. XGBoost is quicker and more accurate than typical machine learning methods. Ge et al. [9] proposed a prediction method of consumer purchase behavior based on deep forest, which realizes more efficient prediction. Hux et al. [10] on this basis, the deep forest is applied to online purchase behavior prediction, which reduces the time cost and improves the prediction accuracy. Although the above methods improve the generalization ability of the model and achieve more accurate predictions, they can only extract shallow features from the dataset. Therefore, complex feature engineering tasks need to be completed before model training, which requires a lot of manpower and time. Because of the exponential growth of information on the e-commerce platform, it is challenging to apply XGBoost and other integrated learning models to analyze large-scale consumer behavior data. In order to fully extract the effective features of consumers purchasing behavior while saving manpower and time, the academic community has increasingly used the deep learning model to predict consumer behavior. Deep learning algorithms are composed of multiple-layer neural networks, which are self-learning mechanisms that learn from data using a more generic learning process. Scholars can focus more on improving the adaptability and classification accuracy of the model rather than spending a lot of effort on feature extraction [11]. In the past, the prediction of consumer purchase behavior based on deep learning mostly used a single model, which was affected by many uncertain factors, and the feature extraction and generalization ability of the model were weak. To solve the above problems, this paper presents a prediction method based on attention mechanism and CNN-BiLSTM for consumer purchasing behavior. The model combines the advantages of CNN extracting local features [12,13] and BiLSTM extracting sequence features and introduces attention mechanism to dynamically assess the significance of each feature in the sample input [14, 15]. The experimental results show that compared with other models, this prediction method has advantages in accuracy and stability.

2. **The framework of the proposed method.** The main structure of the customer buying behavior prediction method based on attention mechanism and CNN-BiLSTM is depicted in Fig.1, which consists of four essential blocks: feature extraction block, sequence learning block, attention mechanism (AM) block, and the prediction block.

In the feature extraction block, CNN is used to automatically extract and optimize local important features from the input data, and import these extracted local features into the BiLSTM network as input. Considering the correlation between behavior features, in the sequence learning block, BiLSTM uses the memory mechanism to extract the sequence features of purchase behavior, and the results are fed as input to the attention layer. In the

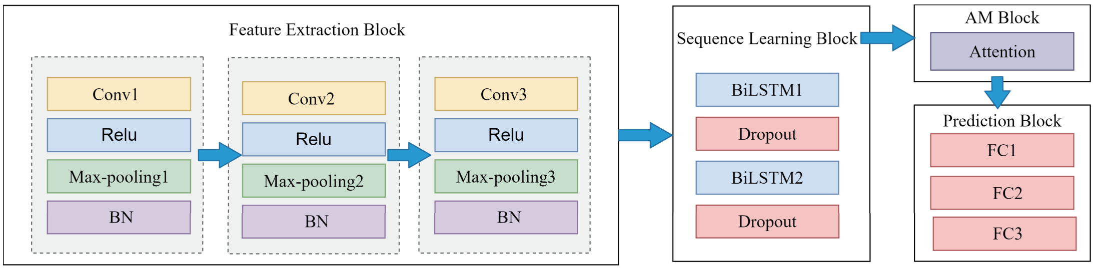

Figure 1. Block diagram of prediction model based on attention mechanism and CNN-BiLSTM.

AM block, AM assigns various weights to the input features to achieve the enhancement and optimization of important information, and assist the model to make more correct decisions. In the prediction block, fully connected layers realize the final prediction. Each layer of the network in these blocks contains trainable parameters, such as filter size, loss function, number of neurons and kernel. The optimal adjustment of these parameters can reduce the prediction error of the proposed method. The parameter settings in this paper are shown in Table 1. The model used by each block of the proposed method will be described in detail below:

Table 1. Model parameter settings

Block	Layer	Number/size/stride of kernels Or the number of neurons
Feature extraction block	Conv1	32*5*1
	Max-pooling1	32/2/1
	Conv1	64*3*1
	Max-pooling1	64/2/1
	Conv1	128*3*1
	Max-pooling1	128/2/1
Sequence learning block	BiLSTM1	128
	BiLSTM2	64
AM block	Attention	-
Prediction block	FC1	32
	FC2	10
	FC3	1

2.1. **Convolutional Neural Network(CNN).** CNN has been widely used in image processing, face recognition and time series analysis and other fields [16]. It consists of several hidden layers including convolution layers, pooling layers and fully connected (FC) layers, and its main function is to automatically extract and optimize the local features of the data [17, 18]. Its network structure model is shown in Fig.2.

After the purchasing features enter the convolution layer from the input layer, one-dimensional convolution kernels in this layer extract features by convoluting local areas of consumer buying behavior. Different convolution kernels can extract different feature signals from the input. For the n-th convolution layer, its output can be expressed as

$$x_j^n = f\left(\sum_{i=1}^{M} k_i^{n-1} * w_{ij}^n + b_j^i\right) \tag{1}$$

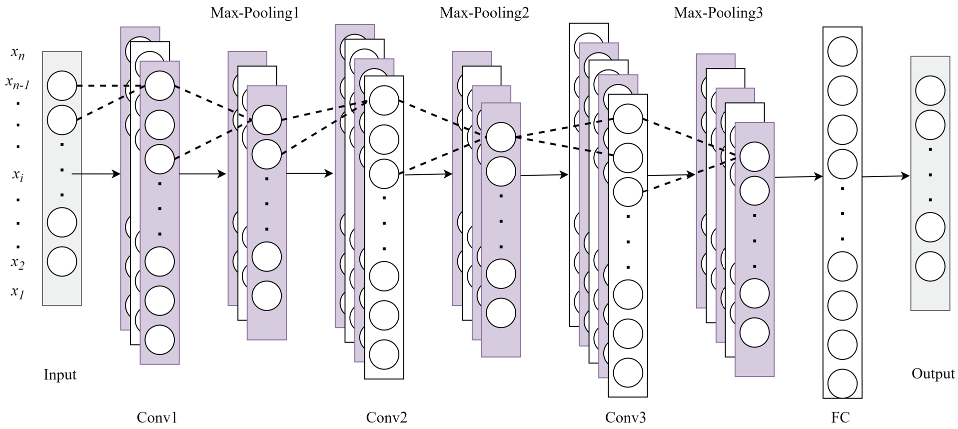

FIGURE 2. CNN model.

where k is convolution kernel, j is the number of convolution kernels, m represents the number of channels for input x^{n+1}, b represents the bias corresponding to the kernel, f denotes the activation function, and $*$ denotes convolution operator. The feature signals extracted by the convolution layer then enter the pooling layer to reduce the feature dimension, which simplifies the computational complexity of the network. If the last pooling layer is the m-th layer, its output acts as BiLSTM input. The output at the full connection layer is

$$y_i^{m+1} = f\left(w^m * x^m + b^m\right) \tag{2}$$

where w represents the weight and b represents the offset.

2.2. **Bidirectional LSTM(BiLSTM).** LSTM [19], as a special RNN, controls data by adding gate structure to the network nodes [20] so that it solves the problem of long-term dependence of sequence data. The unit structure of LSTM is shown in Fig. 3 (a). It is composed of an input gate (i^t), an output gate (o^t), and a forget gate (f^t), which can write, read and reset information, respectively. In each gate, controlling the state of memory cells is performed through point-wise multiplication and sigmoid function operations. The forget gate decides what information should be ignored or kept. The forget gate's output value is between zero and one. If the value is near zero, it indicates that the information will be discarded. Otherwise, the closer to one, the more information will be kept. The formula of the forget gate is calculated as follows:

$$f^t = \sigma\left(W_f * \left[h_{t-1,x_t}\right] + b_f\right) \tag{3}$$

Where σ is the sigmoid activation function, and W and b denote the weight and bias of any gate unit, respectively. The input gate is used to update the cell state by first feeding the current input x_t and the information h_{t-1} of the previous hidden state into the sigmoid function. The input gate decides which information to update. Its 0 means unimportant and 1 means important. The formula for calculating the input gate is as follows:

$$i_t = \sigma\left(W_i * \left[h_{t-1}, x_t\right] + b_i\right) \tag{4}$$

The *tanh* function is then fed the current input x_t and the hidden state h_{t-1}. The cell state $\hat{C}_t$ is again computed, and the cell state is updated with the new value.

$$\hat{C}_t = \tanh\left(W_c * \left[h_{t-1}, x_t\right] + b_c\right) \tag{5}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \hat{C}_t \tag{6}$$

Where *tanh* is the hyperbolic tangent activation function, C_t is the new memory cell. To transition to the next time step, the output gate selects the new memory cell C_t and the new hidden state h_t.

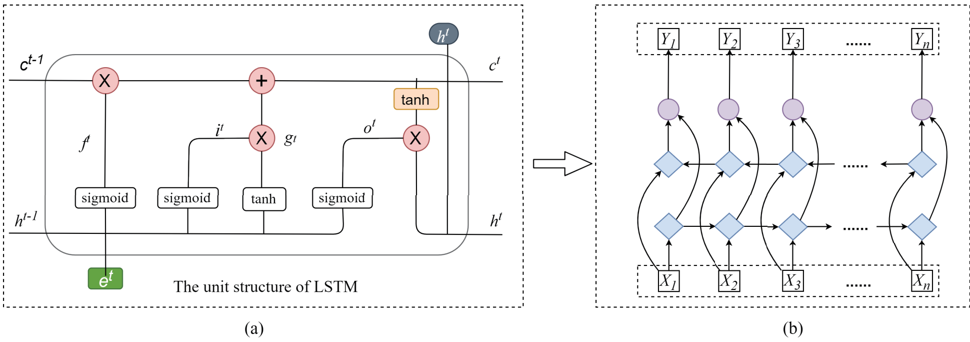

FIGURE 3. BiLSTM model.

The BiLSTM network structure consists of two LSTMs superimposed. The two-layer LSTM network structure is the same, but the weight parameters are different [21]. The BiLSTM network structure is shown in Fig. 3 (b). The two-layer LSTM learns the before and after behavior information of the current sequence according to the forward and backward passes respectively and then connects to the same output. The hidden state h_t of BiLSTM at current time t contains both of forward $\overrightarrow{h_t}$ and backward $\overleftarrow{h_t}$:

$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} \tag{7}$$

Compared with LSTM, BiLSTM can better capture the correlation among features of buying behavior. BiLSTM receives the important feature vector sequence of purchasing behavior extracted by convolutional layer, updates the previously hidden memory cells through i^t , o^t and f^t. The purpose is to keep the relevance of the purchasing features and extract the sequence features of buying behavior, to predict consumer behavior more accurately.

2.3. **Attention mechanism(AM).** Generally speaking, in neural network learning, the number of model parameters determines the expressiveness of the model to a certain extent, and the more the parameters, the more information it stores. However, a huge amount of information will lead to serious information overload. Attention mechanism simulates the human habit of observing the environment and pays more attention to important information [22, 23]. It can solve the problem of information overload by automatically adjusting the attention weight, screening vital information of the current task and ignoring the useless. AM is commonly used in various fields such as image captioning [24], machine translation [25], earthquake prediction [26]. In this paper, an attention mechanism is introduced to optimize features and automatically select important features to improve the efficiency and accuracy of task processing. The structure of the attention model is shown in Fig. 4.

The attention mechanism generates a vector with a probability distribution of attention by weighting the sequence feature extracted from the BiLSTM. The probability distribution is expressed as

$$a_n = \frac{\exp\left(h_n'\right)}{\sum_j^N h_j'} \tag{8}$$

where N is the number of input feature vectors, h_n' is a complete hidden state sequence, h_j' represents the j-th hidden state vector, and a_n represents the attention distribution of

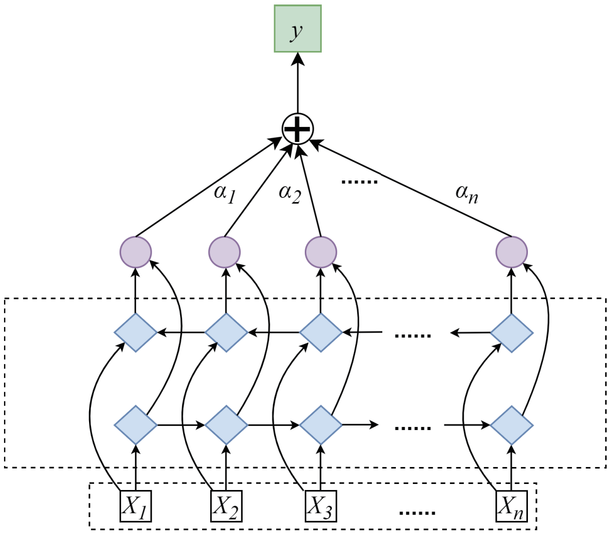

Figure 4. Attention model.

h_n' for h_j'. Then, matrix multiplication is performed between the original feature and the attention matrix to obtain the formula of the feature matrix with attention weight

$$F = \sum_{n=1}^{N} a_n * h_n \tag{9}$$

3. Experiment process and result analysis.

3.1. Experiment platform. The experiment uses Keras deep learning framework based on Python programming language and the operating system is 64-bit Microsoft Windows 10. The CPU is Intel (R) Core (TM) i5-7200u @ 2.50 GHz, the graphics card model is NVIDIA GTX 940MX, and the video memory is 2 GB.

3.2. Data preprocessing. The data used in this paper is the historical purchase data of consumers on an e-commerce platform from February 1, 2013 to August 31, 2013, including consumer information, commodity information and order information. Because the data is incomplete, multi-source heterogeneous, and susceptible to noise, the following data preprocessing work is done in this paper:

1) Historical purchase data includes purchase records of members and non members, and all records of non members are deleted.
2) The records with negative amount in the order data are return orders, and the records with 0 amount are gifts during promotion. In order to avoid the impact of these noise data on the prediction results, delete all records with 0 or negative order amount.
3) There are missing values and disordered order in the consumer's historical purchase data, and interpolation and sorting operations are carried out respectively.
4) In order to make heterogeneous data comparable, the data is standardized.
5) consumer information, commodity information and order information data are separated from each other, so they are integrated according to the format of consumers' corresponding commodities.
6) There are few features that can directly represent consumer behavior in the original data. The features of consumer purchase behavior are constructed by using the method of statistical analysis. The statistical features of consumer purchasing behavior are shown in Table 2.

Table 2. Statistical features of consumer purchasing behavior

Feature type	Feature name
Consumer features	Frequency features of purchases
	Number of purchases by consumers
	Consumer shopping time preferences
	Features of Consumers' Crazy Shopping
	Consumer repurchase rate
Commodity features	Commodity popularity features
	Repurchase rate of goods
Consumer-commodity cross-feature	The number of times the same consumer purchases a certain com
	Repurchase rate of a certain commodity by the same consur

3.3. **Model evaluation index.** In this paper, the prediction problem of consumer purchase behavior is a two-category problem, and the performance of the prediction model is evaluated by three indexes: precision rate, recall rate and F1 score. The sample classes with purchase behavior are positive and the others are negative. The classifier's prediction results, either correct or incorrect, can be one of the four below: predicting positive class as positive (TP), predicting negative class as positive (FP), predicting negative class as negative (TN), and predicting positive class as negative (FN). The evaluation indicators are defined as follows:

1) The equation below expresses the precision rate (P), which describes how many positive samples anticipated by computation are successfully categorized.

$$P = \frac{TP}{TP + FP} \tag{10}$$

2) The recall rate (R) is linked to the categorization of minority samples, which measures the accuracy with which minority samples are classified,which calculated as follows

$$R = \frac{TP}{TP + FN} \tag{11}$$

3) The $F1$ value is a harmonic average of precision rate and recall rate.

$$F1 = \frac{2 * P * R}{P + R} \tag{12}$$

3.4. **Comparison method.** Common shallow machine learning models used to predict consumer purchase behavior include SVM [7], XGBoost [8] and Random Forest (RF) [10]. In the deep learning model, CNN, LSTM and CNN-LSTM have advantages in time series prediction. Therefore, we choose SVM, RF, XGBoost, CNN, LSTM and CNN-LSTM as the comparison model of CNN-BiLSTM-AM proposed in this paper. And the comparison models are divided into the following two categories.

3.4.1. *Shallow machine learning.*

1) SVMs are machine learning algorithms that are based on statistical learning theory. The main hyperparameters of the SVM model are set as follows: the type of kernel function is chosen Radial basis function (RBF) because it has fewer hyperparameters, which decreases the complexity of the model. The regulation factor (C), epsilon (e), and gamma (c) hyper-parameters are set to 1, 0.01, and 0.1, respectively.

2) Random Forest (RF) is an efficient ensemble learning technique that employs averaging to boost prediction performance and avoid overfitting. It is extensively utilized in a variety of regression problems. In this experiment, the maximum depth of the trees and the number of trees in the forest is set to 6 and 100, respectively.
3) XGBoost is one of the boosting algorithms. XGBoost integrates many CART regression tree models to form a strong classifier. In this experiment, the maximum depth of the trees and the number of trees in the forest is set to 6 and 100, respectively.

3.4.2. *Deep learning.* Deep learning categories include CNN, LSTM and CNN-LSTM models. CNN model is related to the feature extraction block of the proposed model. The LSTM model is also similar to the sequence learning block of the proposed model but uses LSTM instead of BiLSTM. CNN-LSTM model consists of feature extraction block and sequence learning block.

3.5. **Analysis of experiment results.** To verify the effectiveness of the model proposed in this paper, it is compared with SVM, RF, XGBoost, CNN, LSTM and CNN-LSTM on the same dataset. The results are shown in Table 3. It can be seen that compared with other models, the accuracy, recall and F1 score of the model proposed in this paper are significantly improved, which shows that the model proposed in this paper combines the advantages of CNN, BiLSTM and attention mechanism in feature extraction and improves the accuracy of prediction.

Table 3. Comparison of prediction results of different models

Method	Evaluating indicators		
	Precision rate	Recall rate	F1 score
SVM	0.575	0.633	0.603
Random Forest	0.631	0.687	0.658
XGBoost	0.718	0.734	0.723
CNN	0.685	0.646	0.664
LSTM	0.728	0.754	0.741
CNN-LSTM	0.773	0.742	0.757
CNN-BiLSTM-AM	0.791	0.743	0.766

3.5.1. *Comparative analysis with shallow machine learning.* Experiments show that the model proposed in this paper can generate better prediction results than the shallow machine learning models, as shown in Table 3. It can be seen that the F1 score of the model proposed in this paper is higher than all compared shallow machine learning models, 0.43% higher than the XGBoost model and 1.63% higher than the SVM model. This is because the shallow machine learning model can only extract the shallow features of consumer purchase behavior, but can not extract the deeper features. The model proposed in this paper combines the advantages of CNN, BiLSTM and attention mechanism in feature extraction, and can more fully extract the effective features of consumer purchase behavior. To further verify the stability of the model, draw a broken line diagram of the F1 score of each model trained 20 times on the same dataset to analyze the fluctuation amplitude. The results are shown in Fig.5. It can be seen that the model proposed in this paper has higher prediction accuracy and better stability than SVM and RF. The prediction accuracy of the XGBoost model is sometimes higher than that of the model proposed in this paper. The main reason is that XGBoost is a strong classifier integrating many weak classifiers, which is more suitable for binary classification problems. However, the ability of the XGBoost model in dealing with large-scale datasets is weak, so the prediction results of the XGBoost model on the datasets used in this paper have high volatility and weak stability. The above experimental data show that the CNN-BiLSTM-AM model proposed in this paper has higher prediction accuracy and better stability than the shallow machine learning model.

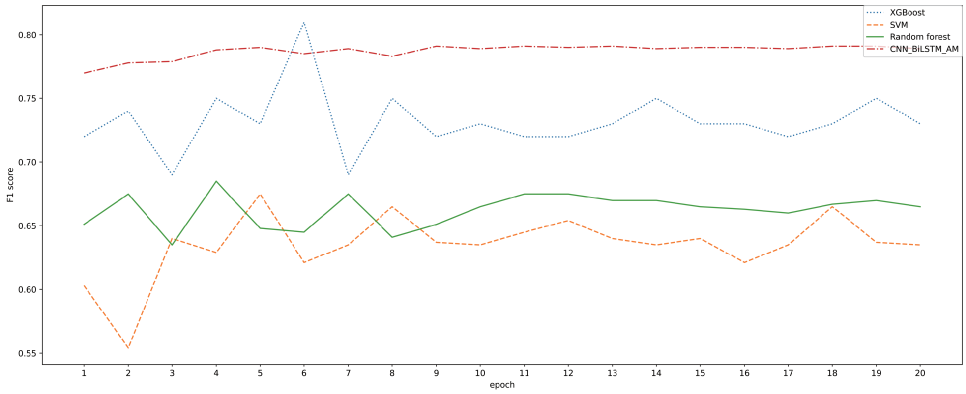

FIGURE 5. Comparison of F1 score change curve between shallow machine learning and the model proposed in this paper.

3.5.2. *Comparative analysis with deep learning.* Because shallow machine learning models can only extract shallow features from a dataset, complex feature engineering operations must be done before the model can be trained. Deep learning models are increasingly being used by researchers to predict consumer behavior. Through experiments, the experimental results we obtained prove that the model proposed in this paper can achieve better prediction results than other deep learning models compared, and the results are shown in Table 3. It can be seen that the F1 score of the model proposed in this paper is higher than the CNN, LSTM and CNN-LSTM models, 0.09% higher than the CNN-LSTM model, and 1.02% higher than the CNN model. To further verify the stability of the model, the F1 score of each model trained 20 times on the same dataset is also drawn as a line graph to analyze the fluctuation range, and the results are shown in Fig.6. The CNN-LSTM model has higher prediction accuracy and stability than the CNN and LSTM models, as can be seen. The model proposed in this paper improves accuracy while maintaining the stability of CNN-LSTM. In addition, the main reason why the LSTM model outperforms CNN is closely related to the selection of experimental data. The dataset used this time is the historical data of consumer purchase behavior, which belongs to standard time-series data. LSTM models are more suitable for the analysis and forecasting of time series data. Although the CNN model has higher accuracy in learning nonlinear sequence data, it is difficult for a single CNN model to learn the change law of sequence data well when the data volatility and instability are high. CNN-LSTM combines the benefits of CNN and LSTM to improve prediction accuracy while maintaining prediction stability. Based on the CNN-LSTM model, the model proposed in this paper introduces an attention mechanism to optimize the features and realize Enhanced optimization of important information to achieve higher prediction accuracy than CNN-LSTM.

4. **Conclusions.** In this paper, we propose a prediction method of consumer buying behavior, which combines attention mechanism and CNN-BiLSTM. By integrating feature selection with model selection, this method fully extracts important local features and behavior sequence features with combined advantages of the two models and uses

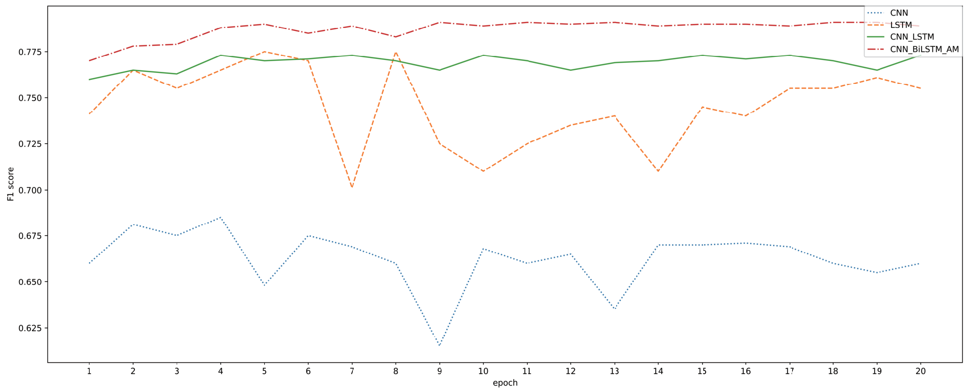

FIGURE 6. Comparison of F1 score change curve between deep learning and the model proposed in this paper.

an attention mechanism for optimization. Experiment results show that the prediction performance of the method proposed in this paper is better, and it is suitable for consumer buying behavior prediction. However, there are still some shortcomings in research methods and data selection. Due to the low conversion rate of buying behavior in the interaction process, there is an imbalance between the samples with purchase and those without purchase, which will affect the training effect of the model to some extent. In addition, Commodity reviews and promotion features such as “Double Eleven” and “June 18th” (both are shopping carnivals in China) are lacking in the sample, which will also affect consumer purchase behavior. In the follow-up research, we will try to balance the samples of the historical purchase data of consumers and incorporate more feature information that affects consumer purchase behavior. The model proposed in this paper also has some limitations. Compared with other models, it has no obvious advantage in training efficiency. In the later research, we will focus on improving the training efficiency of the model by adding new hidden layers and improving the network structure.

Acknowledgment. This research was supported by the industry-university research innovation fund of Chinese universities(2020ITA03003).

REFERENCES

[1] Q. S. Zhang, L. J. Weng, Overview of social recommendation system, *Computer Engineering and Applications*, vol. 56, no. 1, pp. 1-10, 2020.

[2] L. Yu, Q. H. Du, B. Y. Yan, J. Y. Xiang, G. Y. Xu, Y. F. Leng, Survey of reinforcement learning based recommender systems, *Computer Science*, vol. 48, no. 10, pp. 1-18, 2021.

[3] S. Benkessirat, N. Boustia, N. Rezoug, Overview of recommendation systems, *Smart Education and e-Learning 2019*, pp. 357-372, 2019.

[4] Y. Q. Gao, Research on personalized recommendation based on deep learning, *Industrial Control Computer*, vol. 34, no. 10, pp. 90-92, 2021.

[5] J. B. Zhang, J. Z. Huo, Stacking Fusion Model for Customer Purchase Behavior Prediction, *Shanghai Management Science*, vol. 43, no. 1, pp. 12-19, 2021.

[6] J. Xu, J. Wang, Y. Tian, J. P. Yan, X. Li, X. Gao, SE-stacking: Improving user purchase behavior prediction by information fusion and ensemble learning, *Plos one*,vol. 15, no. 11, e0242629, 2020.

[7] H. X. Tang, Research on the evaluation method of e-commerce user loyalty, *Computer Simulation* , vol. 34, no. 1, pp. 348-351, 2017.

[8] Y. Qiu, G. W. Yang, User behavior prediction and risk analysis based on xgboost algorithm, *Industrial Control Computer* , vol. 31, no. 9, pp. 2, 2018.

[9] S. L. Ge, J. Ye, M. X. He, Prediction model of user purchase behavior based on deep forest, *Computer Science*, vol. 46, no. 9, pp. 190-1944, 2019.

[10] X. Hu, Y. F. Yang, L. H. Chen, S. R. Zhu, Research on a prediction model of online shopping behavior based on deep forest algorithm, *2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 137-141, 2020.

[11] K. K. Tseng, C. Wang, T. J. Xiao, C. M. Chen, M. M. Hassan, V. H. C. de Albuquerque, Sliding large kernel of deep learning algorithm for mobile electrocardiogram diagnosis, *Computers & Electrical Engineering*, vol. 96, 107521, 2021.

[12] J. M. T. Wu, Z. C. Li, N. Herencsar, B. Vo, J. C. W. Lin, A graph-based CNN-LSTM stock price prediction algorithm with leading indicators, *Multimedia Systems*, pp. 1-20, 2021.

[13] J. M. T. Wu, Z. C. Li, G. Srivastava, M. H. Tasi, J. C. W. Lin, A graph-based convolutional neural network stock price prediction with leading indicators, *Software: Practice and Experience*, vol. 51, no. 3, pp. 628-644, 2021.

[14] W. J. Lu, J. Z. Li, J. Y. Wang, L. L. Qin, A CNN-BiLSTM-AM method for stock price prediction, *Neural Computing and Applications*, vol. 33, no. 10, pp. 4741-4753, 2021.

[15] P. Kavianpour, M. Kavianpour, E. Jahani, A. Ramezani, A CNN-BiLSTM model with attention mechanism for earthquake prediction, *arXiv preprint arXiv*, 2112.13444, 2021.

[16] Y. Li, Y. He, M. Zhang, Prediction of chinese energy structure based on convolutional neural network-long short-term memory (CNN-LSTM), *Energy Science & Engineering*, vol. 8, no. 8, pp. 2680-2689, 2020.

[17] L. Liu, J. C. Zhu, G. J. Han, Y. G. Bi, Bearing health monitoring and fault diagnosis based on joint feature extraction in 1D-CNN, *Journal of Software*, vol. 32, no. 8, pp. 12, 2021.

[18] Y. Zhang, L. Guo, C. N. Du, Y. Wang, D. G. Huang, Extraction of english drug names based on Bert-CNN mode, *J. Inf. Hiding Multim. Signal Process*, vol. 11, no. 2, pp. 70-78, 2020.

[19] J. M. T. Wu, L. Y. Sun, G. Srivastava, J. C. W. Lin, A long short-term memory network stock price prediction with leading indicators, *Big Data*, vol. 9, no. 5, pp. 343-357, 2021.

[20] K. Yang, S. D. Fan, Trend prediction and application of network time series data based on long-term and short-term memory, *Journal of Propulsion Technology*, vol. 42, no. 3, pp. 675-682, 2021.

[21] R. Mahajan, V. Mansotra, Predicting geolocation of tweets: using combination of CNN and BiLSTM, *Data Science and Engineering*, vol. 6, no. 4, pp. 402-410, 2021.

[22] B. F. Fu, Y. D. Yang, Y. Ma, J. Y. Hao, S. Q. Chen, S. Liu, T. G. Li, Z. Y. Liao, X. L. Zhu, Attention-based recurrent multi-channel neural network for influenza epidemic prediction, *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1245-1248, 2018.

[23] Z. Zhang, B. Wu, B. Schuller, Attention-augmented end-to-end multi-task learning for emotion prediction from speech, *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6705-6709, 2019.

[24] Y. Chu, X. Yue, L. Yu, M. Sergei, Z. K. Wang, Automatic image captioning based on ResNet50 and LSTM with soft attention, *Wireless Communications and Mobile Computing*, 2020.

[25] B. Zhang, D. Y. Xiong, J. Xie, J. S. Su, Neural machine translation with GRU-gated attention model, *IEEE transactions on neural networks and learning systems*, vol. 31, no. 11, pp. 4688-4698, 2020.

[26] M. H. Al Banna, T. Ghosh, M. J. Al Nahian, Attention-based bi-directional long-short term memory network for earthquake prediction, *IEEE Access*, vol. 9, pp. 56589-56603, 2021.