

v1: 12 September 2023

Research Article

Attention Mechanism Model Combined with Adversarial Learning for E-commerce User Behavior Classification and Personality Recommendation

Peer-approved: 12 September 2023

© The Author(s) 2023. This is an Open Access article under the CC BY 4.0 license.

Qeios, Vol. 5 (2023)
ISSN: 2632-3834

Sharif Uddin Ahmed Rana¹

1. Limkokwing University of Creative Technology, Putrajaya, Malaysia

In traditional e-commerce websites, consumers' evaluation of products will affect new customers' decisions on whether to buy the products. Some fraudulent merchants manipulate consumers' online comments for their interests, and multitudes of fake comments abuse consumers' rights and interests and the development of traditional e-commerce. The purpose of the present work is to detect and identify fake comments through user behavior classification. A series of innovative research works are carried out around the user behavior recognition task from four aspects: extraction and description of low-level behavior features, spatial representation of high-level user behavior, design of behavior classification model, and user behavior detection in unsegmented text. A feature extraction model based on the super-complete independent component analysis algorithm and a behavior classification model via attention mechanism are proposed. Moreover, a feature source discriminator is designed, and adversarial learning is used to optimize discriminator loss and generator loss. Finally, an experiment is implemented to test the effects of attentional mechanism and adversarial learning on the text retrieval model and visualize the results. In this experiment, the text retrieval algorithm based on a stacked cross-attention mechanism and adversarial learning retrieves the Microsoft Common Objects in Context (MS-COCO) and Flickr30K data sets on mainstream transmedia. The experimental results demonstrate that the stacked cross-attention mechanism has an excellent matching ability of fine-grained hierarchical features; the average accuracy of the algorithm after improvement increases from 81.23% to 83.11%. Besides, the prediction accuracy coverage is above 95%, which can significantly improve the predicted effect of text characteristics and image features, thus enhancing the accuracy of text retrieval and classification. The research has a certain experimental reference value for the classification and discrimination of business users' behavior.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

With the deepening of global informatization, the rapid popularization of computer and Internet technology has created favorable conditions for the development of e-

commerce. Meanwhile, the huge demand for e-commerce has promoted the growth of related industries in the field of the Internet [1][2][3]. In the scenario of e-commerce on the Internet, because the construction of electronic websites does not need realistic sites and equipment, the cost of store expansion is relatively low. Consequently, the number of e-commerce stores began to show an

exponential growth trend. With the explosive increase in the number of users, each business platform has formulated the corresponding user classification and product personalized recommendation mechanism. Besides, to optimize the user experience, each platform has formulated the corresponding user online comment and feedback mechanism. The rapidly increasing user feedback comments imply consumer consumption intention, which has enormous commercial application value.

However, with the proliferation of e-commerce stores, some businesses take unfair measures to expand their competitiveness and product sales. On the one hand, some unscrupulous merchants hire online supporters to publish fake praise for their products and stimulate the purchase desire of potential users through fake comments [4][5][6]. On the other hand, to suppress their competitors, some businesses deliberately hire professional writers to fabricate negative reviews for their competitors and reduce users' desire to buy. All kinds of fake comments can be found on e-commerce platforms. Ordinary users have a low level of recognition of fake comments [7], which brings immense difficulties to user identification and is not conducive to the sound and stable development of e-commerce platforms.

The development of the Internet of Things [8] and adversarial learning algorithms [9] provide a new solution to this problem. Based on the attention mechanism [10], the semi-supervised learning method combined with the machine learning model supplements the comment corpus. In view of the text information related to comments, an in-depth study is conducted here from four aspects, namely the user behavior recognition task from four aspects: extraction and description of low-level behavior features, spatial representation of high-level user behavior, design of behavior classification model, and user behavior detection in unsegmented text. In addition, independent component analysis (ICA) is utilized to build a feature extraction model, and the attention mechanism is adopted to construct a behavior classification model. These models can carry out natural language processing and recognition for repeated and unrelated texts and perform supervised learning of fake comment corpus via the behavior detection algorithm. The present work has practical application value for the classification of e-commerce users' behavior and the purification and development of the e-commerce website environment.

2. Related works

2.1. Recent studies on the attention mechanism model

Many scholars have conducted studies on the attention mechanism model. Wang et al. (2018) [11] studied the credit scoring mechanism of Internet lending by using the attention mechanism and Long Short-Term Memory (LSTM) algorithm. They proposed a consumer credit scoring method based on an attention mechanism by using the online operation behavior data of borrowers. They also tested their scheme and found that the proposed consumer credit scoring method based on attention mechanism and LSTM had an obvious improvement effect compared with the traditional artificial feature extraction method. Zhu et al. (2019) [12] explored depth sensors with an attention mechanism and convolution neural network (CNN) model and analyzed action recognition based on the skeleton as a case. They adopted the attentional mechanism for depth sensors to extract the most relevant features. These sensors could collect human bone data, providing a wealth of information for motion recognition. The authors extracted the most relevant features to analyze deep learning, and the final results verified the effectiveness of the method. Fang et al. (2019) [13] used the improved model with a multi-level vector and attention mechanism to detect and study phishing emails. They confirmed that the overall accuracy of the proposed model reached 99.848%; meanwhile, the false positive rate was 0.043%. The high accuracy and low error rate ensured that the filter could identify phishing emails with high probability and filter out legitimate emails as little as possible, which verified the effectiveness of the model in detecting phishing emails. Chauhan et al. (2020) [14] studied the two-step hybrid unsupervised model of attention mechanism used for aspect extraction. They used rules-based methods to extract the single-word and multi-word aspects and used these aspects as label data to train the attention-based deep learning (DL) model for aspect-term extraction, reducing the cost of manual annotation of text. Finally, they verified the effectiveness and reliability of the proposed method through experimental evaluation.

Meanwhile, the research on the attention mechanism model has been extended to many fields, such as public transportation, in recent years. For example, Zhou et al. (2020) [15] built a prediction model based on the spatio-temporal attention mechanism for the city's passenger boarding and boarding demand. They finally proved that the spatio-temporal attention model was superior to the traditional method in terms of prediction performance. Li (2020) [16] adopted the attention mechanism to construct a click-through rate prediction model based on user interest

to study the characteristics of the automatic encoder based on stack height nonlinear interaction. Their experimental results showed that the proposed model achieved higher performance than the most advanced models available. Gao et al. (2021) ^[17] carried out DL modeling of building energy consumption prediction based on attention mechanism. They proposed three interpretable decoders and self-attention models by using the attention mechanism based on hidden layer state and feature-based attention mechanism. They reported that visualization of model attention weight strengthened the interpretability of the model at the hidden state and feature levels. Kang et al. (2021) ^[18] obtained high-resolution root images through scanners and introduced an attention mechanism to refine the segmentation of root edges. The authors employed the Deeplabv3+ model for end-to-end pixel segmentation of cotton root images. Through simulation, they found that the precision value, recall value, and F1-score of the proposed model were 0.9971, 0.9984 and 0.9937, respectively, which could accurately segment different growth states in the cotton growth cycle. Ren et al. (2021) ^[19] introduced the explicable mixed attention mechanism into the concrete dam displacement technology for the applied research on the new DL prediction model. Specifically, they put forward an interpretable mixed attention mechanism model based on encoder-decoder architecture. The comparison results showed that the model was superior to other models in most cases. To sum up, introducing the temporal attention mechanism model into the decoder stage can correctly extract critical periods by identifying the relevant hidden states of all time steps, which has practical reference value for quantification and visualization of temporal attention weight.

2.2. Research progress of adversarial learning and e-commerce

Many scholars have discussed adversarial learning and e-commerce. For example, Hershcovis et al. (2018) ^[20] explored the influencing factors of uncivilized behaviors in the workplace by means of confrontation and avoidance. They found that confrontation and avoidance were ineffective in preventing the recurrence of uncivilized behaviors, and avoidance also led to increased emotional exhaustion. Brighi et al. (2019) ^[21] investigated the impact of direct confrontation and toughness on the happiness promotion of adolescents. They conducted a questionnaire survey of youth online victimization and built a structural equation model. The survey results indicated that the network victims and confrontational tactics on mood symptoms were mediated by the elasticity, the influence of network victims showed a positive role, and direct confrontation showed negative

effects. Ashburn-Nardo et al. (2020) ^[22] predicted whether people would face other people's discriminatory behaviors through the applied research on fighting prejudice and perception. They assessed whether perceived responsibility for bias for people in supervisory roles at work increased as the number of employees they supervised increased. The results suggested that confronting prejudice and learning were key factors in predicting whether bystanders face discrimination. Case et al. (2020) ^[23] studied the cross-mode of fighting prejudice, and they concluded that gender socialization and the stereotype that equated femininity with male homosexuality might reduce the alliance behavior between men and increase the resistance against anti-gay prejudice between women. Zaragas et al. (2021) ^[24] discussed psychodrama and its contribution to children's competition, made an in-depth study of the phenomenon in the psychodrama circle, and then highlighted the results of applying psychodrama technology to young athletes participating in competitions. The findings suggested that psychodrama could be an innovative alternative to schooling.

Different from the study of adversarial learning, e-commerce has always been the focus of scholars' study and attention. Bhattacharyya et al. (2020) ^[25] explored the factors influencing the purchase and recommendation of goods on e-commerce websites and found that Facebook-driven social commerce benefited from a large number of likes. At the same time, suggestions on Facebook influenced customers' purchases and recommendations on linked e-commerce sites. Vasić et al. (2021) ^[26] studied the satisfaction function of logistics service users in e-commerce and developed a methodology and an original measurement tool with eight dimensions to explore different dimensions of logistics service. The results demonstrated that the proposed measurement function could improve customer satisfaction with e-commerce. In the context of the computer digital marketing era, Wang et al. (2021) ^[27] studied the immersive interactive experience of users of content e-commerce live broadcasting. They explored the characteristics of a panoramic video transmission scheme based on a simple mail transmission protocol that solved the synchronization problem of heterogeneous networks. Liu et al. (2021) ^[28] proposed a prudent iterative naive Bayes algorithm by analyzing and studying diseases in pharmaceutical e-commerce. Experimental test results showed that the accuracy of the algorithm reached 98.64% on six diseases, and the recall rate reached 90.90%, which was better than most benchmark algorithms.

3. Mixed attention mechanism model based on CNN and LSTM

3.1. Analysis of practical functions of the attention mechanism and adversarial learning algorithm

In the development of e-commerce, attention mechanisms and adversarial learning algorithms are the core elements, and their performance is directly related to the application of network and information components. In the human visual system, there are two different attention mechanisms: top-down and bottom-up. In terms of the way of applying attention ^[29], the spatial attention model can complete the preprocessing of adaptive tasks by learning the deformation of input but fails to take into account the changes in the number of semantic matches between images and sentence descriptions. Therefore, the

present work adopts the spatial-temporal stacked cross-attention mechanism and takes multitudes of network devices as the nodes of information exchange through the combination of wired and wireless networks. If network devices such as switches, gateways, and firewalls are attacked, their information and connected devices will be adversely affected. In addition, the operating system primarily uses Windows, VxWorks, and other commercial systems that can timely repair and protect the vulnerability of the e-commerce system. If the system is rarely upgraded or patched, the system may age and have vulnerabilities that cannot be remedied in time, posing great information security risks. Fig. 1 displays the framework diagram of user behavior classification and comments discrimination system of e-commerce platforms based on attention mechanism and adversarial learning algorithm.

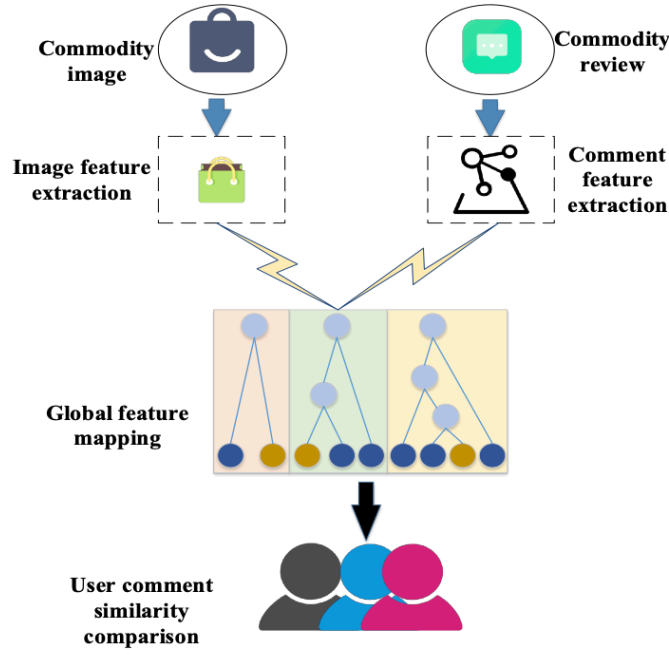


Fig. 1. Framework of the user behavior classification and comment discrimination system of e-commerce platforms

Based on the above analysis of the e-commerce system, as a complex network, the system has security risks in network equipment, operating system, communication protocol, access equipment, and other aspects. Therefore, it is extremely important to use the attention mechanism and adversarial learning algorithm to detect and classify user behaviors.

3.2. Spatio-temporal mixed attention mechanism model based on super-complete ICA algorithm

In the spatio-temporal mixed attention model [30] based on the adversarial learning algorithm, the generative adversarial network (GAN) needs to be constructed and optimized. GAN, composed of the generator G and discriminator D, adopts the same calculation method as the gates in the LSTM network. The updating gate information z_t and the resetting gate information \tilde{h}_t are calculated according to:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (1)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (2)$$

where r_t represents the hidden layer information calculated by multiple gates, and h_{t-1} refers to the amount of information forgotten by the hidden layer at the previous moment. Meanwhile, x_t signifies the demand for hidden layer information \tilde{h}_t at the current moment, and $\sigma()$ denotes the memory hiding function of the network

updating gate. The text discrimination model reported here discriminates the comment content through target iteration, to carry out the antagonistic learning and game between text feature recognition generation and feature source discrimination. The objective function of the comment recognition model can be expressed as Equation (3).

$$\max_D E_{x \sim P_r} [\log D(x)] + E_{x \sim P_g} [\log(1 - D(x))] \quad ()$$

In Equation (3), P_r and P_g respectively represent the distribution of the real text content and the distribution of the text generated by the generator. In the model, the input z of the generator G is random noise, and the generator converts this random input into data type and output examples.

For text block description, first, each word in the text is represented by an independent coding vector. Then, each independent coding vector is multiplied by an embedding matrix for vector dimension adjustment, and the word embedding vector of the word is adjusted through the bidirectional network. The word embedding vector is averaged from the forward and reverse outputs of each stage, as presented in Equations (4) ~ (6):

$$\overrightarrow{h_l} = \overrightarrow{GRU}(x_i), i \in [1, n] \quad (4)$$

$$\overleftarrow{h_l} = \overleftarrow{GRU}(x_i), i \in [1, n] \quad (5)$$

$$e_i = \frac{\overrightarrow{h_l} + \overleftarrow{h_l}}{2}, i \in [1, n] \quad (6)$$

where $\overrightarrow{GRU}(x_i)$ represents the forward input of each layer of the network, and $\overrightarrow{h_l}$ denotes the forward output of this layer of the network. Besides, $\overleftarrow{GRU}(x_i)$ signifies the reverse input of each layer of the network, $\overleftarrow{h_l}$ refers to the reverse output of this layer of the network, and e_i stands for the unit word embedding vector obtained after averaging.

In the text feature similarity measurement module, after obtaining the predicted global feature V'' of the image through text prediction, its dimension is compared with the global feature V' of the original image. Moreover, the mean square error of the two is used as the loss function of the two, as shown in Equation (7).

$$l_{mse} = (V'' - V')^2 \quad (7)$$

Since the mean square error function treats all data sets equally in training, it cannot solve the problem that the retrieval accuracy of difficult negative samples decreases. Therefore, triplet loss maximization is adopted to calculate the backpropagation gradient based on the most difficult negative samples. By emphasizing the negative sample optimization problem in the loss function, the loss function $l_{MH}(i, c)$ of the matching text (i, c) is defined as:

$$l_{MH}(i, c) = \max_c [\alpha - S(i, c') + S(i, c)] + \max_{i'} (\alpha - S(i', c) + S(i, c)) \quad (8)$$

where c' and i' represent the two sets of the most difficult negative samples, α refers to the batch size of the training set, and $S(i, c)$ denotes the cosine similarity function adopted by coarse-grained hierarchical similarity measurement. Fig. 2 reveals the flow of the super-complete ICA algorithm.

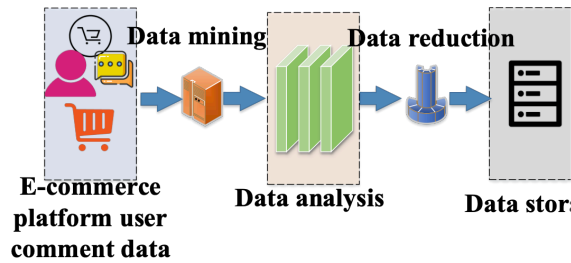


Fig. 2. Flow of the super-complete ICA algorithm

The original attention mechanism is improved as an image-text stacked cross-attention mechanism by combining images and text. Moreover, the aforementioned fine-grained hierarchical feature recognition model of images and statements is used to calculate the image-text similarity between each region in the image and each word in the sentence synthetically. Then, the attention mechanism is used for more accurate recognition and alignment. For the construction of the stacked cross-attention mechanism model, the similarity between image fine-grained hierarchical features V and between text fine-grained hierarchical features E is measured and calculated, respectively. Firstly, each local region after image segmentation is used to add the attention mechanism to the experimental text. Denote S_{ij} as the judgment of word similarity between the i -th text region and the j -th text region, which is determined by Equation (9).

$$S_{ij} = \frac{v_i^T e_j}{\|v_i\| \|e_j\|}, i \in [i, k], j \in [1, n] \quad (9)$$

In Equation (9), v_i^T represents the vector representation of text description under the image region, e_j denotes the weighted comprehensive value of text represented by the embedded vector. The weighted vector representation can be written as Equation (10).

$$a_i^t = \sum_{j=1}^n \alpha_{ij} e_j \quad (10)$$

In Equation (10), $\alpha_{ij} = \frac{\exp(\lambda_1 \bar{s}_{ij})}{\sum_{j=1}^n \exp(\lambda_1 \bar{s}_{ij})}$ represents the word vector after text partition, λ_1 denotes the hyperparameter, and \bar{s}_{ij} refers to the result after regularization.

The text description vector defined under the region v_i in the i -th image is calculated by the cosine similarity function of which the vector representation is a_i^t , as shown in Equation (11).

$$R(v_i, a_i^t) = \frac{v_i^T a_i^t}{\|v_i\| \|a_i^t\|} \quad (11)$$

Finally, the image-text similarity model after average pooling is obtained by integrating the similarity functions between vector representations under each image region and the corresponding attention mechanism, which are calculated according to:

$$S_{LSE}(I, T) = \log \left(\sum_{i=1}^k \exp(\lambda_2 R(v_i, a_i^t)) \right)^{(1/\lambda_2)} \quad (12)$$

$$S_{AVG}(I, T) = \frac{\sum_{i=1}^k R(v_i, a_i^t)}{k} \quad (13)$$

where λ_2 is the hyperparameter, $S_{LSE}(I, T)$ represents the image-text similarity error of adding attention to the

image in the way of text starting, and $S_{AVG}(I, T)$ denotes the average image-text similarity of adding attention to the image in the way of text starting.

3.3. Model construction of the mixed attention mechanism based on CNN and LSTM

To re-add attention mechanism to each word in the text, it is necessary to build a mixed attention mechanism model based on CNN and LSTM algorithm and carry out weighted synthesis by adding image vector features of the word attention mechanism. The specific calculation methods are as follows:

$$a_j^v = \sum_{i=1}^k \alpha'_{ij} v_i \quad (14)$$

$$\alpha'_{ij} = \frac{\exp(\lambda_1 \bar{s}_{ij})}{\sum_{i=1}^k \exp(\lambda_1 \bar{s}_{ij})} \quad (15)$$

where a_j^v represents the weighted value of word text in comments, and α'_{ij} refers to the weighted value of similarity between text statement and image. For each word vector in image matching, the similarity between the vector e_j of the j -th word embedding vector and function value of the image T is also defined as cosine, as shown in Equation (16).

$$R'(e_j, a_j^v) = \frac{e_j^T a_j^v}{\|e_j\| \|a_j^v\|} \quad (16)$$

The overall similarity between images and texts is calculated according to:

$$S'_{LSE}(I, T) = \log \left(\sum_{j=1}^n \exp(\lambda_2 R'(e_j, a_j^v)) \right)^{(1/\lambda_2)} \quad (17)$$

$$S'_{AVG}(I, T) = \frac{\sum_{j=1}^n Rr(e_j, a_j^v)}{n} \quad (18)$$

where λ_2 denotes the hyperparameter, $S'_{LSE}(I, T)$ represents the similarity error between images and texts, and $S_{AVG}(I, T)$ refers to the average similarity between images and texts. The form of emphasizing the most difficult negative sample is adopted for the judgment of similarity of fine-grained level. For matched image-text (I, T) , the most difficult negative sample is determined by:

$$\hat{I}_h = \operatorname{argmax}_{m \neq I} S(m, T) \quad (19)$$

$$\hat{T}_h = \operatorname{argmax}_{d \neq T} S(I, d) \quad (20)$$

where \hat{I}_h represents the most difficult negative sample of commodity image data, and \hat{T}_h refers to the most difficult negative sample of comment text data. The loss function $l_{\text{hard}}(I, T)$ of the overall stacked cross-attention mechanism model is defined as follows:

$$l_{\text{hard}}(I, T) = \left[\alpha - S(I, T) + S(I, \hat{T}_h) \right] + \left[\alpha - S(I, T) + S(\hat{I}_h, T) \right] \quad (21)$$

$$l_{\text{adv}}(\theta_D) = -\frac{1}{n} \sum_{i=1}^n (m_i \cdot (\log D(v_i; \theta_D) + \log(1 - D(t_i; \theta_D))) \quad (22)$$

where $l_{\text{adv}}(\theta_D)$ denotes the cross-entropy loss of modal classification used in each training iteration; meanwhile, m_i stands for the real modal label of each comment instance, which is represented in the form of an independent coding vector. Besides, $D(t_i; \theta_D)$ represents the probability that the discriminator of each instance outputs a similar feature. Fig. 3 illustrates the framework of the mixed attention mechanism model based on CNN and LSTM.

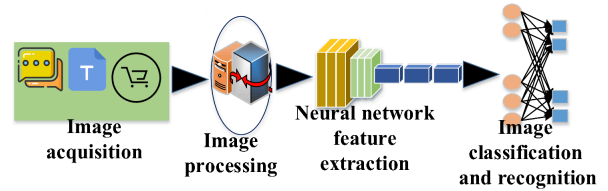


Fig. 3. Framework of the mixed attention mechanism model based on CNN and LSTM

For the overall image feature recognition and prediction network, the loss function optimization process of its generator is shown in Equations (23) and (24).

$$(\hat{\theta}_v, \hat{\theta}_t, \hat{\theta}_{imd}) = \underset{\theta_v, \theta_t, \theta_{imd}}{\operatorname{argmin}} (L_{emb}(\theta_v, \theta_t, \theta_{imd}) - L_{adv}(\hat{\theta}_D)) \quad (23)$$

$$\hat{\theta}_D = \underset{\theta_D}{\operatorname{argmax}} (L_{emb}(\hat{\theta}_v, \hat{\theta}_t, \hat{\theta}_{imd}) - L_{adv}(\theta_D)) \quad (24)$$

In Equations (23) and (24), $\hat{\theta}$ represents the sample characteristics of text comments, $\hat{\theta}_t$ refers to the minimum sample parameter of text comments, $\hat{\theta}_{imd}$ signifies the maximum sample parameter of text comments, and $\hat{\theta}_D$ denotes the maximum loss function.

The annotated text corpus is limited, and the performance of trained classifier models varies greatly in the initial stage. Therefore, the data accuracy of each initial data set is evaluated by referring to the annotation results of the classifier, as shown in Equation (25).

$$C(x) = \underset{y \in \text{Labe}}{\operatorname{argmax}} \frac{\sum_{i=1}^3 E(y, C_i(x)) \times P_i(x)}{\sum_{i=1}^3 P_i(x)} \quad (25)$$

In Equation (25), $C(x)$ refers to the final label of the corpus x , $P_i(x)$ represents the accuracy of the classification model in the initial corpus, and y signifies the label of the corpus. Besides, $E(y, C_i(x))$ represents the text classification according to different labels, and the classification method is shown in Equation (26).

$$E(y, C_i(x)) = \begin{cases} 1, & C_i(x) = y \\ 0, & C_i(x) \neq y \end{cases} \quad (26)$$

For the global text vector, dimension reduction decomposition is carried out through the global word contribution matrix. Then, the final word vector's target loss function J can be presented as:

$$J = \sum_{i,j=1}^V f(X_{ij}) \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2 \quad (27)$$

where X_{ij} refers to the word contribution matrix, indicating the frequency of occurrence of the word in the statement, b_i denotes the bias term, and $f(x)$ represents the set weight function, which can be described as Equation (28).

$$f(x) = \begin{cases} (x/x_{\max})^\alpha & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases} \quad (28)$$

The model is optimized by finding the minimum of the objective loss function to obtain the final word vector. The feedforward attention model after semantic coding can be expressed as:

$$e_t = a(h_t) \quad (29)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^T \exp(e_k)} \quad (30)$$

$$C = \sum_{t=1}^T \alpha_t h_t \quad (31)$$

where h_t represents the output result of the encoder, $a(h_t)$ refers to the function with learning ability, α_t stands for the probability of attention distribution, and C represents the semantic encoding format after the probability calculation of attention distribution. In the working process of the feedforward attention mechanism, the input into the hidden layer is sent to $a(h_t)$ function to calculate the probability distribution vector a of attention. Then, the probability distribution vector is multiplied and accumulated with the hidden layer output of history nodes to get the final semantic coding. Fig. 4 illustrates the process of electronic commerce data compression and processing based on the product image and the comment text.

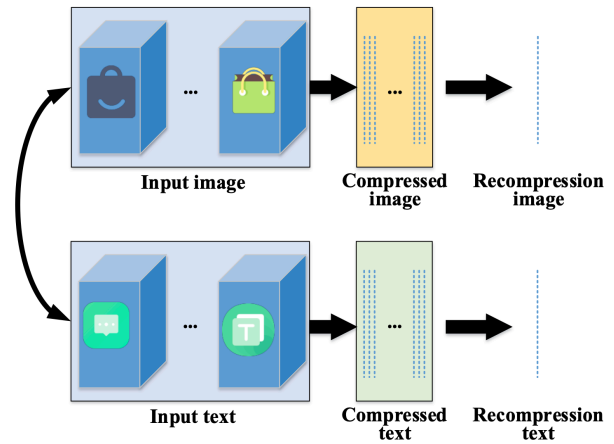


Fig. 4. Compression and processing process of commodity image and user comment data in the e-commerce system

3.4. Case analysis

To verify the performance of the mixed attention mechanism model based on CNN and LSTM constructed here, the application scenarios of an online e-commerce platform are selected for case analysis. MATLAB is adopted to analyze the user comment data. The user comment data of each commodity on the platform are collected, and the collected data in the data set is divided into a training data set and a test data set by a ratio of 7:3. The proportion of each type of data in the two data sets is consistent. The

effects of attentional mechanism and adversarial learning on the text retrieval model are tested and visualized by using mainstream cross-media retrieval data sets Microsoft Common Objects in Context (mS-COCO) and Flickr30K. This test can help the mixed attention mechanism model based on CNN and LSTM achieve the expected prediction results. Besides, the performance of the model reported here is compared with that of traditional algorithms in terms of accuracy, precision, recall, and F1 value. The Alex Network (AlexNet) [31], Dense Network (DenseNet) [32], Interleaved Group Network (IGCNet) [33], Visual Geometry Group Network (VGGNet) [34], and Residual Network (ResNet) [35] are selected for the comparative analysis. Meanwhile, the

training time and test time of these models in mS-COCO and Flickr30K datasets are evaluated, respectively.

4. Results and discussion

4.1. Comparison of experimental performance of different models under the MS-COCO dataset

With the MS-COCO data set, the performance of the mixed attention mechanism model based on CNN and LSTM is compared with that of the traditional AlexNet, DenseNet, IGCNet, VGGNet, and ResNet models. The accuracy, precision, recall, and F1 values of these models are shown in Fig. 5 to Fig. 8. Meanwhile, the algorithm training time and testing time under this data set of these models are compared, and the results are shown in Fig. 9 and Fig. 10.

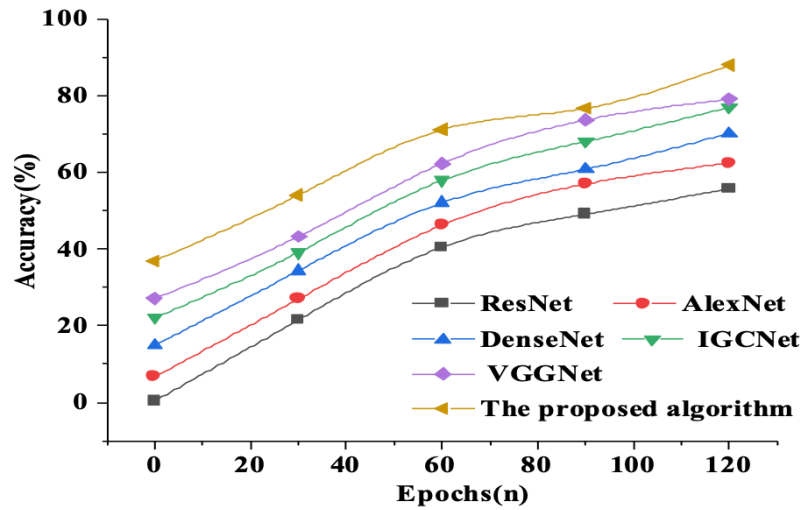


Fig. 5. Comparison of accuracy results of different models

From the comparison of accuracy results of different algorithm models in Fig. 5, compared with other models, the model proposed here has the highest accuracy in the test data set. When the number of model iterations is 90, the accuracy of the mixed attention mechanism model based on CNN and LSTM can reach 80%. Subsequently,

with the increase in the number of model iterations, the accuracy of each algorithm increases slowly. However, the accuracy performance of other models is not better than that of the mixed attention mechanism reported here, indicating that this model has the highest accuracy performance.

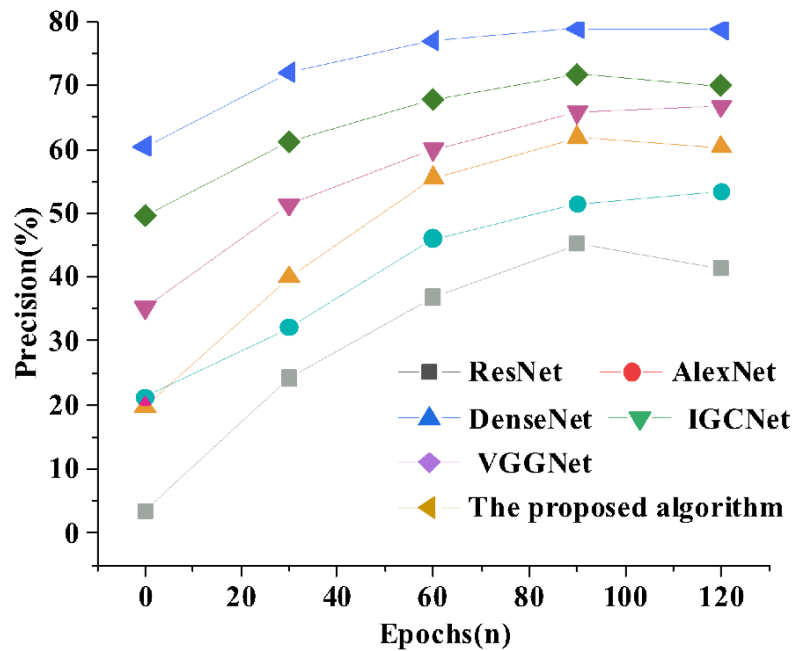


Fig. 6. Comparison of precision results of different models

In Fig. 6, with the growth of the model number of iterations, the classification precision of each model improves slowly. Among them, the mixed attention mechanism model based on CNN and LSTM always has the optimal precision. After 60 iterations, the classification

precision of this attains 75%, which is at least 7.8% higher than other traditional models. Besides, the classification precision of this model reaches 80% after 120 iterations, which significantly improves the classification accuracy of e-commerce user behavior.

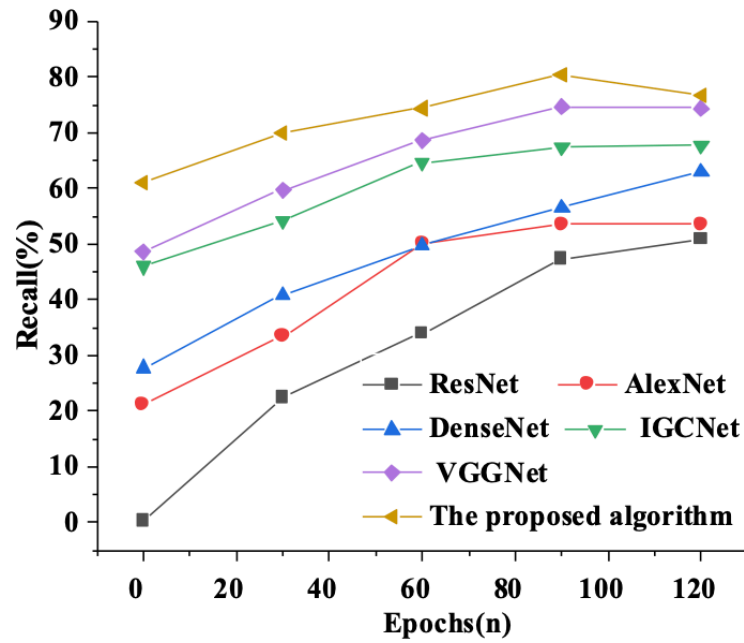


Fig. 7. Comparison of recall results of different models

According to the results of the recall of different models in Fig. 7, the recall of the mixed attention mechanism model based on CNN and LSTM always remains above 60% during the experiment. Compared with other models, although the recall of the VGGNet model is close to that of

this model after 120 iterations, there is at least a 6.8% difference in recall between the two in 30 iterations. This indicates that the model reported here can maintain optimal recall performance.

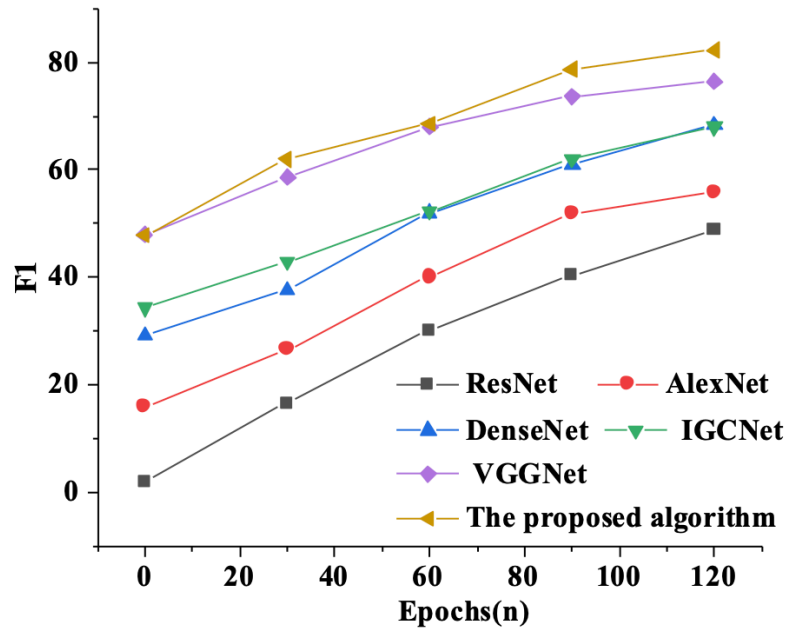


Fig. 8. Comparison of F1 values of different models

According to the comparison of the F1 value results of different models in Fig. 8, among the six models, the F1 value curve of the ResNet model always stays at the bottom. Besides, the F1 value performance of the DenseNet model and IGCNet model are relatively close throughout the experiment. In the first 60 iterations of the model, the performance of the VGGNet model is not significantly

different from that of the model proposed here in the F1 value. However, with the increase in the number of model iterations, the F1 value of the mixed attention mechanism model becomes increasingly prominent. After 120 iterations, the F1 performance of the model proposed here reaches more than 80%, which is at least 3.2% higher than that of other models

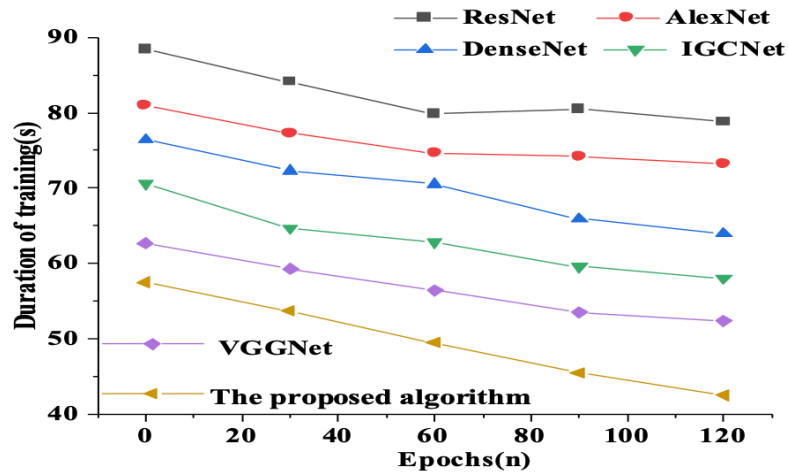


Fig. 9. Comparison of the training time of different models

From Fig. 9, except for the mixed attention mechanism model, the other comparative models need a long time for training. As the iteration continues, although the training time of each model is shortened, the model proposed here takes the shortest time. At the beginning of the

experiment, this model spends about 60 seconds for training. After 120 iterations, the training time is reduced to about 42 seconds, which can save 18 seconds of training time compared with the beginning of the experiment, greatly improving the training efficiency of the model.

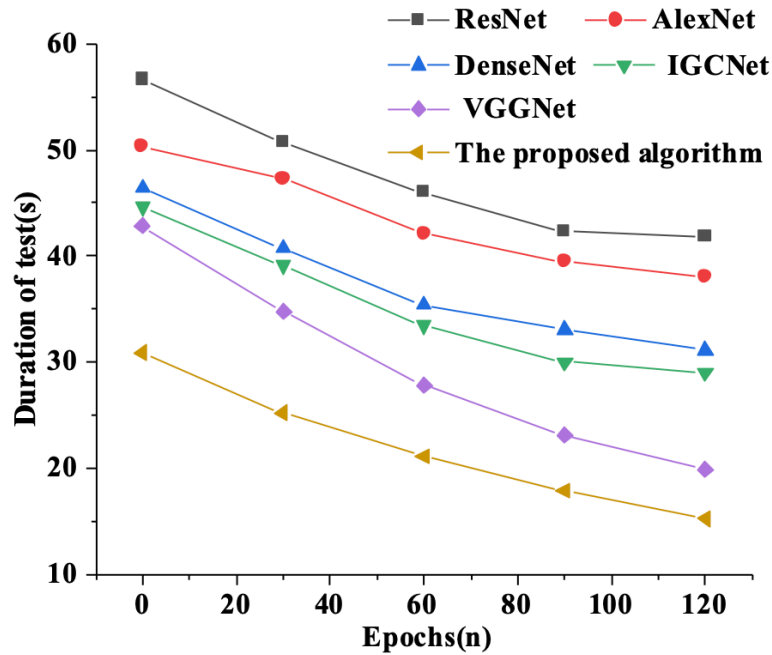


Fig. 10. Comparison of the testing time of different models

According to Fig. 10, the traditional models all take a long testing time. With the increase in the number of iterations, the testing time can be shortened, but the testing time of the model proposed here is the shortest. At the beginning of the experiment, the mixed attention mechanism model requires more than 30 seconds for testing. As the number of model iterations increases, the testing time of this model can be reduced to about 16 seconds after 120 iterations, which decreases by 14 seconds compared with the testing time at the beginning of the experiment. Therefore, the mixed attention mechanism model can achieve outstanding performance in test efficiency.

4.2. Comparison of experimental performance of different models under the Flickr30K data set

The accuracy, precision, recall, and F1 value of the mixed attention mechanism model based on CNN and LSTM under the Flickr30K data set are compared with the traditional AlexNet, DenseNet, IGCNet, VGGNet, and ResNet models, as shown in Fig. 11 to Fig. 14. Meanwhile, the algorithm training time and testing time under this data set of these models are compared, and the results are shown in Fig. 15 and Fig. 16.

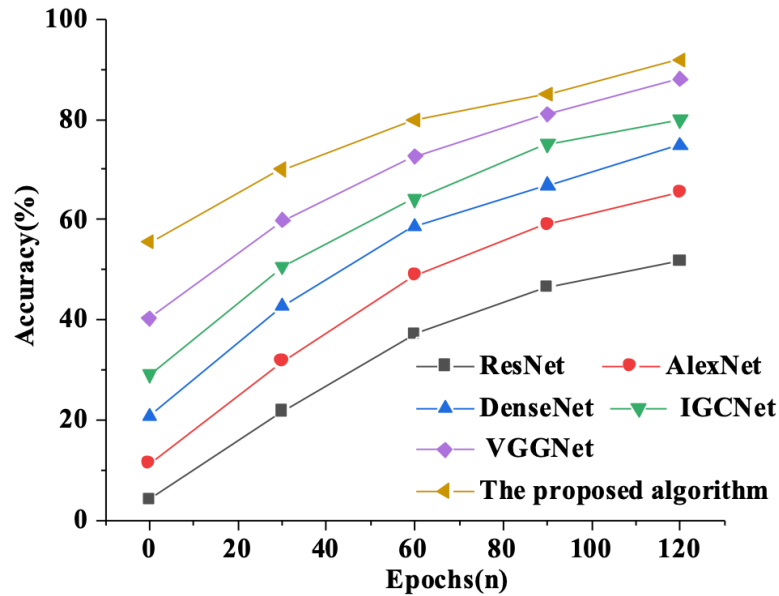


Fig. 11. Comparison of accuracy results of different models

From the comparison of the accuracy results of different models in Fig. 11, compared with other models, the mixed attention mechanism model based on CNN and LSTM has the highest accuracy on the Flickr30K test data set. After 60 iterations, the accuracy of the mixed attention mechanism model can reach 80%. Subsequently, with the

increase in the number of iterations, the accuracy of each model increases slowly, but the accuracy performance of other traditional models is still poorer than that of the mixed attention mechanism model, indicating that this model has the highest accuracy performance.

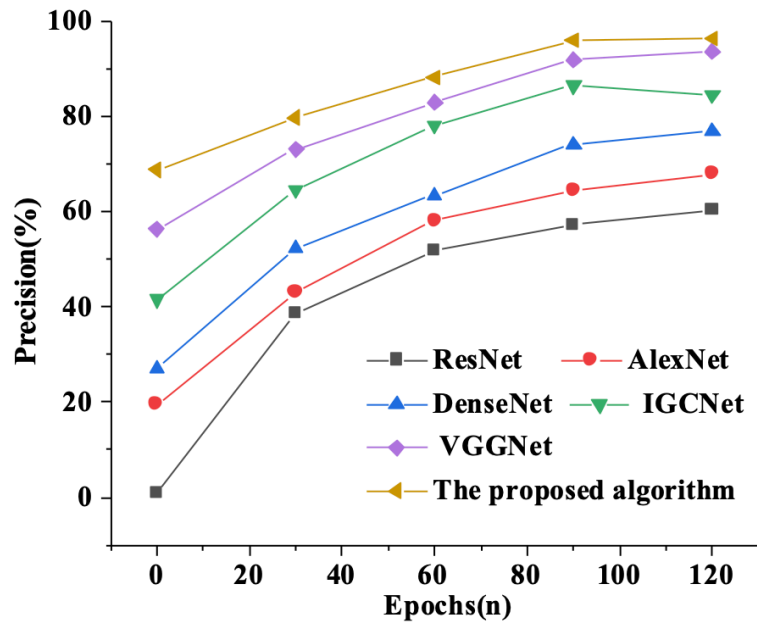


Fig. 12. Comparison of accuracy results of different models

From the comparison of the accuracy results of different models in Fig. 12, the precision of the mixed attention mechanism model based on CNN and LSTM always

maintains above 70%. Compared with other algorithms, this model has significant advantages in precision performance.

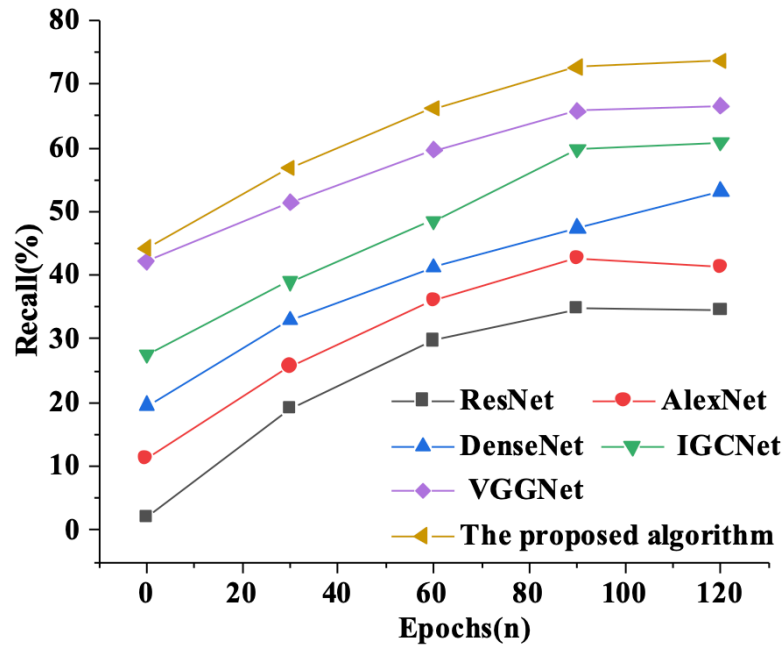


Fig. 13. Comparison of recall results of different models

Through the comparison of recall of different models in Fig. 13, the recall of the mixed attention mechanism model based on CNN and LSTM always remains above 45% during the experiment. Compared with other models, although the recall of the VGGNet algorithm is close to this

model at the beginning of the model iteration, there is a difference of at least 7.2% between the recall rates of the two at 120 iterations. This demonstrates that the model reported here can maintain the optimal recall.

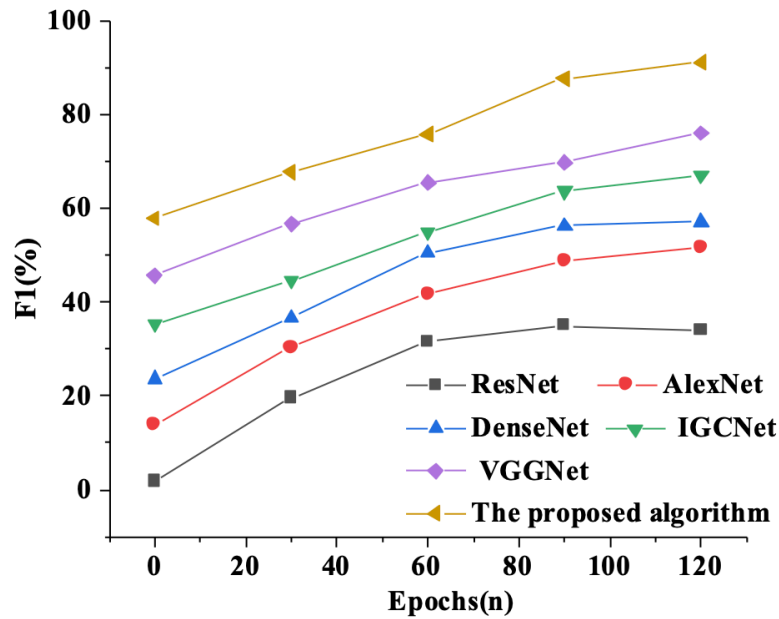


Fig. 14. Comparison of F1 values of different models

According to the comparison of the F1 value results of different models in Fig. 14, among the six algorithm models, the F1 value curve of the ResNet model always stays at the bottom. On the contrary, the F1 value performance of the DenseNet model and IGCNet model is relatively close throughout the experiment. With the

increase in iterations, the F1 value advantage of the mixed attention mechanism model based on CNN and LSTM becomes increasingly prominent. After 90 iterations, the F1 value of this model can reach more than 90%, which is at least 9.6% higher than that of other models.

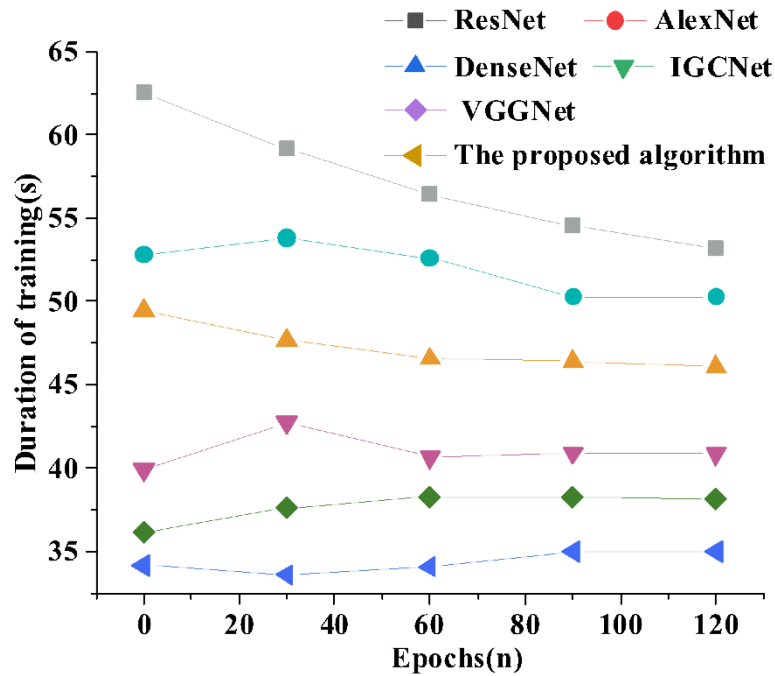


Fig. 15. Comparison of the training time of different models

In Fig. 15, except for the mixed attention mechanism model based on CNN and LSTM, the rest of the traditional models take a long time to train. With the increase in the number of iterations, the training time of each model is shortened, but the mixed attention mechanism model

takes the shortest time. The overall training time of the model reported here is always maintained under 35 seconds, which can greatly improve the efficiency of model training.

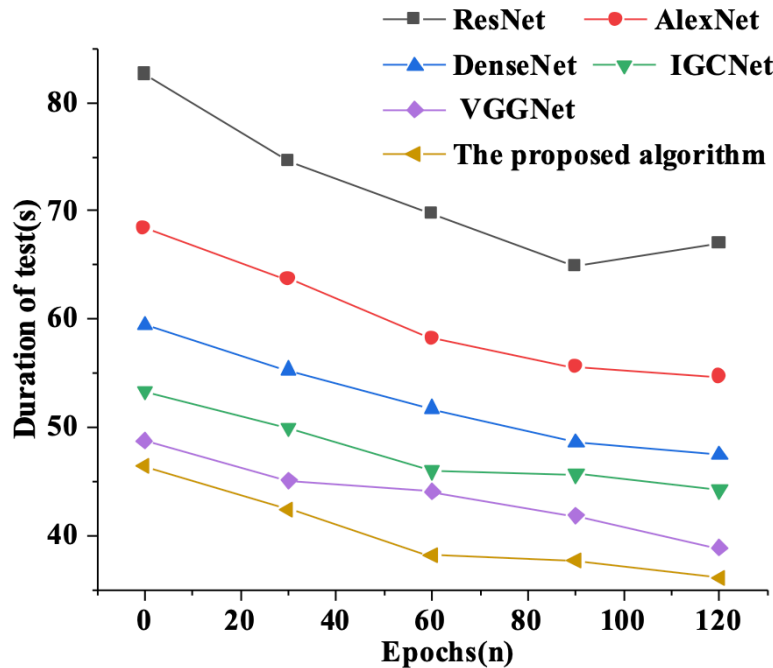


Fig. 16. Comparison of the testing time of different models

According to Fig. 16, the traditional models all take a long testing time. With the increase in the number of iterations, the testing time can be shortened, but the testing time of the model proposed here is the shortest. At the beginning of the experiment, the mixed attention mechanism model requires more than 45 seconds for testing. As the number of model iterations increases, the testing time of this model can be reduced to about 35 seconds after 120 iterations, which decreases by 10 seconds compared with the testing time at the beginning of the experiment. Therefore, the mixed attention mechanism model has a practical application value to improve the efficiency of the user behavior classification model.

5. Conclusion

With the rapid development of e-commerce platforms, some dishonest merchants manipulate consumers' online comments for their own interests, and a large number of fake comments seriously affect consumers' interests and the development of traditional e-commerce. It is urgent to purify the network environment of e-commerce, promote the healthy development of the platform, and safeguard the rights and interests of consumers. Therefore, an in-depth study is conducted on the user comment data from the extraction and description of the underlying behavior characteristics through case analysis of the application

scenario of an online e-commerce platform. Moreover, a spatio-temporal mixed attention mechanism model based on super-complete ICA is proposed. The test results of different models show that stacked cross-attention has a good matching ability of fine-grained hierarchical features. In addition, the recognition accuracy of the mixed attention mechanism algorithm based on CNN and LSTM reported here is above 80% in different test data sets, and the recognition accuracy can be guaranteed at about 95%. The research has certain practical application value for the classification of business users' behavior and the discrimination of true and false comments. However, some disadvantages are inevitable. For example, in the experiment, the optimization effect of adversarial learning on the stacked cross-attention retrieval model is not significant. To strengthen the ability to represent image and text features of the model, the future study will improve the word embedding algorithm for comments to extract more effective features.

References

1. [△]Kwilinski A, Volynets R, Berdnik I, et al. E-Commerce: Concept and Legal Regulation in Modern Economic Conditions. *Journal of Legal, Ethical and Regulatory Issues*, 2019, 22: 1-6.

2. [△]Sadeeq M, Abdulla A I, Abdulraheem A S, et al. Impact of electronic commerce on enterprise business. *Technol. Rep. Kansai Univ*, 2020, 62(5): 2365–2378.
3. [△]Leonard L N, Jones K. Trust in C2C electronic commerce: Ten years later. *Journal of Computer Information Systems*, 2021, 61(3): 240–246.
4. [△]Lawton R. It Ain't What You Do (But the Way That You Do It): Will Safety II Transform the Way We Do Patient Safety?: Comment on" False Dawns and New Horizons in Patient Safety Research and Practice". *International journal of health policy and management*, 2018, 7(7): 659.
5. [△]Sujan M. A Safety-II Perspective on Organisational Learning in Healthcare Organisations: Comment on" False Dawns and New Horizons in Patient Safety Research and Practice". *International journal of health policy and management*, 2018, 7(7): 662.
6. [△]Kuroda N. Editorial Comment to False-positive 123 I-metastadiobenzylguanidine scan in a patient with renal cell carcinoma: A case of chromophobe renal cell carcinoma on cocytic variant with a complicated clinical course. *IJU Case Reports*, 2021, 4 (1): 42–43.
7. [△]Chang D, Gui H Y, Fan R, et al. Application of improved collaborative filtering in the recommendation of e-commerce commodities. *International Journal of Computers Communications & Control*, 2019, 14(4): 489–502.
8. [△]Nauman A, Qadri Y A, Amjad M, et al. Multimedia Internet of Things: A comprehensive survey. *IEEE Access*, 2020, 8: 8202–8250.
9. [△]Qi N, Wang W, Xiao M, et al. A learning-based spectrum access Stackelberg game: Friendly jammer-assisted communication confrontation. *IEEE Transactions on Vehicular Technology*, 2021, 70(1): 700–713.
10. [△]Yu X, Feng W, Wang H, et al. An attention mechanism and multi-granularity-based Bi-LSTM model for Chinese QA system. *Soft Computing*, 2020, 24 (8): 5831–5845.
11. [△]Wang C, Han D, Liu Q, et al. A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM. *IEEE Access*, 2018, 7:2161–2168.
12. [△]Zhu K, Wang R, Zhao Q, et al. A cuboid CNN model with an attention mechanism for skeleton-based action recognition. *IEEE Transactions on Multimedia*, 2019, 22(11): 2977–2989.
13. [△]Fang Y, Zhang C, Huang C, et al. Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. *IEEE Access*, 2019, 7:56329–56340.
14. [△]Chauhan G S, Meena Y K, Gopalani D, et al. A two-step hybrid unsupervised model with attention mechanism for aspect extraction. *Expert Systems with Applications*, 2020, 161: 113673.
15. [△]Zhou Y, Li J, Chen H, et al. A spatiotemporal attention mechanism-based model for multi-step citywide passenger demand prediction. *Information Sciences*, 2020, 513: 372–385.
16. [△]Li H, Duan H, Zheng Y, et al. A CTR prediction model based on user interest via attention mechanism. *Applied Intelligence*, 2020, 50(4): 1192–1203.
17. [△]Gao Y, Ruan Y. Interpretable deep learning model for building energy consumption prediction based on attention mechanism. *Energy and Buildings*, 2021, 252: 111379.
18. [△]Kang J, Liu L, Zhang F, et al. Semantic segmentation model of cotton roots in-situ image based on attention mechanism. *Computers and Electronics in Agriculture*, 2021, 189: 106370.
19. [△]Ren Q, Li M, Li H, et al. A novel deep learning prediction model for concrete dam displacements using interpretable mixed attention mechanism. *Advanced Engineering Informatics*, 2021, 50: 101407.
20. [△]Hershcovis M S, Cameron A F, Gervais L, et al. The effects of confrontation and avoidance coping in response to workplace incivility. *Journal of occupational health psychology*, 2018, 23(2): 163.
21. [△]Brighi A, Mameli C, Menin D, et al. Coping with cyber victimization: the role of direct confrontation and resilience on adolescent wellbeing. *International journal of environmental research and public health*, 2019, 16(24): 4893.
22. [△]Ashburn-Nardo L, Lindsey A, Morris K A, et al. Who is responsible for confronting prejudice? The role of perceived and conferred authority. *Journal of Business and Psychology*, 2020, 35(6): 799–811.
23. [△]Case K A, Rios D, Lucas A, et al. Intersectional patterns of prejudice confrontation by White, heterosexual, and cisgender allies. *Journal of Social Issues*, 2020, 76(4): 899–920.
24. [△]Zaragas C K. The Psychodrama and its Contribution to The Children's Competitive manifestation. *Case Study. Cultural-Historical Psychology*, 2021, 17(3): 143–151.
25. [△]Bhattacharyya S, Bose I. S-commerce: Influence of Facebook likes on purchases and recommendations on a linked e-commerce site. *Decision Support Systems*, 2020, 138: 13383.
26. [△]Vasić N, Kilibarda M, Andrejić M, et al. Satisfaction is a function of users of logistics services in e-commerce. *Technology Analysis & Strategic Management*, 2021, 33(7): 813–828.
27. [△]Wang K, Shawl R Q, Neware R, et al. Research on immersive interactive experience of content e-commerce live users in the era of computer digital marketing. *International Journal of System Assurance Engineering and Management*, 2021: 1–11.
28. [△]Liu X, Xu Y C, Yang X. Disease profiling in pharmaceutical E-commerce. *Expert Systems with Applications*, 2021, 178: 115015.
29. [△]Fang Y, Zhang C, Huang C, et al. Phishing email detection using improved RCNN model with multilevel vectors and

- d attention mechanism. *IEEE Access*, 2019, 7:56329–56340.
30. [△]Li J, Liu X, Zhang W, et al. Spatio-temporal attention networks for action recognition and detection. *IEEE Transactions on Multimedia*, 2020, 22 (11): 2990–3001.
 31. [△]Lu S, Lu Z, Zhang Y D. Pathological brain detection based on AlexNet and transfer learning. *Journal of computational science*, 2019, 30:41–47.
 32. [△]Zhang K, Guo Y, Wang X, et al. Multiple feature reweight DenseNet for image classification. *IEEE Access*, 2019, 7: 9872–9880.
 33. [△]Zhang X, Yang F, Hu Y, et al. RANet: Network intrusion detection with group-gating convolutional neural network. *Journal of Network and Computer Applications*, 2021: 103266.
 34. [△]Wang Z, Zheng X, Li D, et al. A VGGNet-like approach for classifying and segmenting coal dust particles with overlapping regions. *Computers in Industry*, 2021, 132: 103506.
 35. [△]Hammad M, Pławiak P, Wang K, et al. ResNet-Attention model for human authentication using ECG signals. *Expert Systems*, 2021, 38(6): e12547.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.