Electronic Commerce Research (2020) 20:321–342
https://doi.org/10.1007/s10660-020-09411-6

Product advertising recommendation in e-commerce based on deep learning and distributed expression

Lichun Zhou[1]

Published online: 24 April 2020

Abstract

With the advent of Internet big data era, recommendation system has become a hot research topic of information selection. This paper studies the application of deep learning and distributed expression technology in e-commerce product advertising recommendation. In this paper, firstly, from the semantic level of advertising, we build a similarity network based on the theme distribution of advertising, and then build a deep learning model framework for advertising click through rate prediction. Finally, we propose an improved recommendation algorithm based on recurrent neural network and distributed expression. Aiming at the particularity of the recommendation algorithm, this paper improves the traditional recurrent neural network, and introduces a time window to control the hidden layer data transfer of the recurrent neural network. The experimental results show that the improved recurrent neural network model based on time window is superior to the traditional recurrent neural network model in the accuracy of recommendation system. The complexity of calculation is reduced and the accuracy of recommendation system is improved.

Keywords Deep learning · E-commerce · Distributed expression · Advertising recommendation

1 Introduction

Recommendation has always been an important traffic entry for e-commerce platforms. In the past, on the e-commerce platform, the recommended scenarios more covered all aspects of the transaction, such as details page, shopping cart, order and payment, etc. In recent years, the development of recommendation has gradually diversified, and the scenarios have gradually covered all traffic entrances, and the recommended entities have also expanded to activities, categories, operation

✉ Lichun Zhou
zhoulc666@163.com

[1] School of Media and Communication, Shangqiu Normal University, Shangqiu 476000, Henan, China

bits, etc. [1]. In the e-commerce website, product recommendation can improve the effective conversion rate of the whole website product sales and increase the product sales. Through the records that users have browsed, collected and purchased, we can understand users' needs more accurately, cluster and label users, recommend products that users are interested in, help users quickly find the products they need, enlarge their needs in time, and sell more diversified products [2, 3]. Product recommendation is divided into regular recommendation and personalized recommendation. Regular recommendation refers to the selection of some fixed commodities in the recommendation position, or based on the relevance between commodities, to recommend the relevant commodities. For example: recommend milk powder after the user buys a bottle. Personalized recommendation refers to recommendation based on user's shopping habits and product characteristics [4]. Recently, deep learning has become a popular word. Deep learning can improve learning results by increasing the size of data sets. Compared with traditional machine learning tools, deep learning can tap the potential of neural networks. Based on powerful feature extraction, it is more suitable for pattern recognition (image, text, audio) than other tools. Deep learning is really used in the retrieval system of search recommendation advertisement, and it will encounter matching problem or retrieval problem. Generally speaking, such a traffic end business can be divided into several modules. After a traffic comes, a traffic back usually represents a user's browsing behavior under a certain scenario. First match, and the later prediction model estimates the degree of interest of a given product, click rate and conversion rate. After the prediction, there are some sorting displays. At present, artificial intelligence technology, especially the model represented by deep learning, has developed rapidly in practical applications, mainly benefiting from massive big data and large-scale computing power. Through the digital abstract and stylized learning of the physical world, artificial intelligence has a strong ability to limit the acquisition of knowledge, but it is difficult to obtain knowledge beyond data, let alone knowledge analogy, migration and reasoning. Taobao Search is the best scenario of big data intelligent application. The cognitive intelligence of machines, such as autonomous learning and discovery, and even creativity, is the higher level of artificial intelligence. Of course, there is still a lot of work to be done in general artificial intelligence, but in this process, how to combine human knowledge and machine intelligence first to achieve preliminary cognitive intelligence, so that Taobao Search and recommendation have an intelligent experience is the direction we are currently exploring.

The major advertisers are trying their best to reduce the cost of advertising, and at the same time, they want to maximize the effect and influence of advertising; and the advertisers are also in the reform of "small alliance becomes big, big alliance becomes strong" [5]. Due to the increase of the number of mobile intelligent terminal users, the proportion of mobile Internet advertising in the overall advertising is also increasing year by year. Personalized advertising is to display different contents and forms of advertising for different users, so as to achieve the precise placement of advertising. With the development of Internet and network marketing, how to identify the interests of users to achieve different advertising for different users, and the advertising is still needed by users, which has become the main problem faced by Internet marketers and advertisers [6]. With the development of the Internet, all kinds of advertising recommendation platforms have been launched one after another. At the same time, the major companies have also developed a better advertising recommendation system. Since the first article on collaborative filtering algorithm was published, the recommendation system has been one of the hot topics in various fields, and in a short period of more than ten years, different fields have achieved good results in the research of recommendation system. Jonghun [7] et al. Push video by using related technologies of text processing. After entering the twenty-first century, the Internet industry has achieved an unprecedented high-speed growth, and the amount of data available to people also shows an exponential growth trend. Human beings have entered the era of information overload, which leads to the recommendation system in the academic and industrial circles have been widely concerned. At present, many foreign enterprises have successfully applied the recommendation system in the business field and achieved good results [8, 9]. Among them, Amazon's book recommendation system, eBay's commodity recommendation system and Netlix's video recommendation system are more concerned.

Nowadays, recommendation system has become an important means for e-commerce website to improve user stickiness and website revenue, and has become one of the core technologies of e-commerce website. In the past, many classic recommendation algorithms, such as content-based recommendation algorithm and collaborative filtering based recommendation algorithm, can not model the sequential behavior data: each item is independent of each other, and can not model the user's preference information based on the influence of time and item sequence [10, 11]. Because of the shortcomings of the traditional recommendation algorithm in this respect, this paper uses the recursive neural network model of deep learning technology to recommend based on the implicit feedback data of consumers, and uses the distributed expression technology of word vector in natural language processing for reference to further improve the calculation method. Recommendation system, especially in-depth recommendation system, has been widely used in industry, especially in e-commerce scenarios (such as Taobao and Jingdong commodity recommendation). A good industrial recommendation system can promote business growth and bring a lot of economic benefits. In this paper, the time sensitivity of consumers' personal preferences is taken into account, and a recommendation model based on improved recurrent neural network is proposed, which has a good recommendation effect. In this paper, firstly, from the semantic level of advertising, we build a similarity network based on the distribution of advertising themes, and then build a deep learning model framework for advertising click through rate prediction. Finally, an improved recommendation algorithm based on recurrent neural network and distributed expression is proposed. In view of the particularity of the recommendation algorithm, the traditional recurrent neural network is improved, and the time window is introduced to control the hidden layer data transmission of the recurrent neural network. The experimental results show that the improved recurrent neural network model based on time window is better than the traditional recurrent neural network model in the accuracy of recommendation system.

2 Recommendation system status and design path

2.1 Intelligent demand of e-commerce

There is no doubt about the importance of customer service to the development of e-commerce In recent years, the rapid development of e-commerce activities such as "double 11", "double 12" and "anniversary" is not a group, and a large number of customer service teams are also difficult to cope with such activities Each e-commerce company began to plan to establish an intelligent customer service team, and through some machine algorithms to simulate people's thinking, to achieve the effect of customer service and user communication With the popularity of deep learning technology, e-commerce has deepened the development of intelligent customer service team [12, 13]. In the past two years, through the research and application of neural network, knowledge level, heterogeneous computing and other emerging fields to ensure the leading edge of intelligent customer service team technology, greatly improving its intelligence and the universality of its application.

E-commerce enterprises train intelligent service teams based on hundreds of millions of data generated by their human customer service and user interaction to simulate every user scenario The application of intelligent service team received customers during the "double 11" and other activities, effectively alleviating the pressure of artificial customer service [14, 15]. The advertising of e-commerce website is different from that of other media in the following aspects.

With the increase of products and users, and the diversification of user demand, in the practical application of e-commerce website, the commodity recommendation system is facing new challenges, mainly including the diversification of recommendation results, personalization and the timeliness of intelligent terminal recommendation, so as to maximize the satisfaction of user demand and stimulate economic growth [16, 17]. Compared with the traditional recommendation algorithm, the current recommendation algorithm requires more dynamic information absorption and processing. On the basis of static description information of goods and users, we need to further consider the history of users' shopping behavior, as well as the current time, location, and even social network information, so as to obtain the latest trends of the people we are concerned about, and at the same time, the behavior of these people will also affect our personal behavior.

(1) Diversified recommendation list

At present, diversified technology has been used in search engine, but it has not been widely used in commodity recommendation system. The existing commodity recommendation system usually directly recommend the pre ordered commodities which are ranked from high to low according to a certain similarity value to users, resulting in a single type of recommended list commodities, which greatly affects the satisfaction of users [18, 19].

(2) Personalized recommendation

Content-based recommendation algorithm automatically generate user files, covering all the key information of users, including personal information, shopping history, and even important information in social network.

(3) Recommendation of intelligent terminal

According to the user requirements and data characteristics of intelligent terminal, there are three difficulties in commodity recommendation on intelligent terminal. The data of intelligent terminal is more abundant, such as various information of user's current environment, including time, geographical location and weather, etc. it needs to consider how to use these information to reflect the user's environment, meet the user's specific needs, and improve the recommendation quality [20, 21].

2.2 Comparison between machine learning and deep learning

According to the probability of deep learning, deep learning is a special machine learning, which has high flexibility and performance. It can express the world through the probability of network hierarchical learning. Each network layer is connected with another network layer to form a computing network. At the same time, deep learning is not a single technology or theory, but a set of comprehensive methods combining multiple theories and achievements of neural network [22], the results of comparison of machine learning and deep learning as shown in Table 1.

The main purpose of recommender system is to construct user portrait and generate product recommendation list according to related algorithm by analyzing user's historical behavior data and product browsing data. For example, e-commerce platform's product recommendation algorithm, music recommendation algorithm, content recommendation algorithm of today's headlines, etc. The recommendation algorithm based on collaborative filtering relies on the user item matrix, which will lead to the problems of data sparsity and cold start. The application of neural network in recommendation system is a hot research direction. In e-commerce recommendation system, commodity retrieval is inevitable. The main technology of image retrieval system is applied, and the most important part is the calculation of similarity between images. The calculation process of image similarity is shown in Fig. 1. In the center of the image, some areas that may contain the target goods are intercepted, and then the features are extracted by neural network, and the extracted image features are transformed and the similarity is measured.

(1) Recommendation algorithm of multi-layer perceptron

Neural network is generated by simulating human brain neurons. Artificial neural network is usually composed of input layer, hidden layer and output layer. The main function of input layer is to receive external information and transmit it to the next layer.

(2) Self coding recommendation algorithm

Autoencoder (AE) also known as self-encoder, is a common infrastructure widely used in the field of deep neural network. It is an unsupervised learning framework. By minimizing the reconstruction error, the important features of

Table 1 Comparison of machine learning and deep learning

	Application scenario	Amount of data required	Time consuming	Usage method
Machine learning	Application in fingerprint recognition and object feature detection	It can be applied to large data volume and small data volume	Less training time than deep learning	Task breakdown processing
Deep learning	Application of text, image, sound and intelligence	Large amount of data can make the effect better	It takes a long time to train	Centralized processing of tasks without decomposition

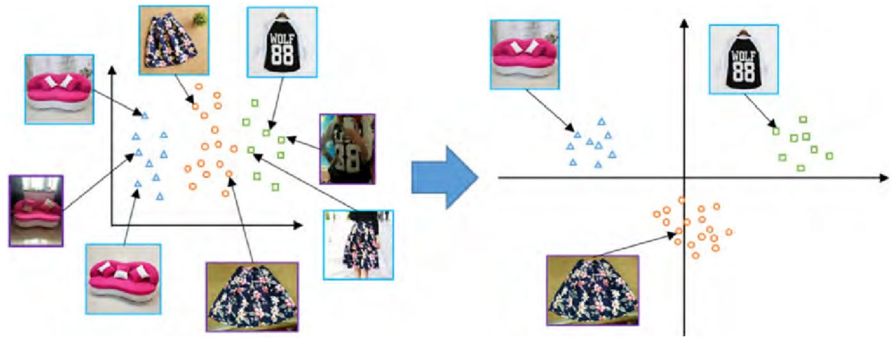

Fig. 1 Sketch of image feature extraction

the system are extracted. The encoder consists of a three-layer or more than three-layer neural network, including encoder, hidden layer and decoder.

(3) Recommendation algorithm of cyclic neural network

Recurrent neural networks (RNN) is a deep-seated neural network model, which mainly increases the number of hidden layers for the neural network, that is, the corresponding parameters such as neuron connection weight and threshold value will increase correspondingly, so it can have a strong learning ability and generalization ability.

(4) Restricted Boltzmann machine recommendation algorithm

Restricted Boltzmann machine (RBM) is essentially a codec. The coding process of restricted Boltzmann machine is to map the original input data from the visualization layer to the hidden layer so as to obtain the hidden factor vector representation of the original input data. In the decoding process, the hidden layer vector is used to map back to the visual layer, so that new visual layer data can be obtained.

2.3 Data mining and feature learning framework

The solution to the problem of commodity category recognition and retrieval under complex system is not only the design of network structure, but also the need of many types of annotation data to constrain the training of the whole network. These data include commodity location, commodity category, commodity attribute and the same item data, so that the retrieval results and query images can have global apparent similarity and local semantic consistency. In general, the exploration and innovation of the above-mentioned deep learning methods will lay a solid foundation for the commodity retrieval technology to become practical. As shown in Fig. 2, these methods can be summarized into a set of feature learning framework. The meanings of the three parts are as follows:

(1) Commodity image preprocessing. Products are divided into rigid bodies (such as shoes, bags, cosmetics, etc.) and non rigid bodies (such as men's wear, women's

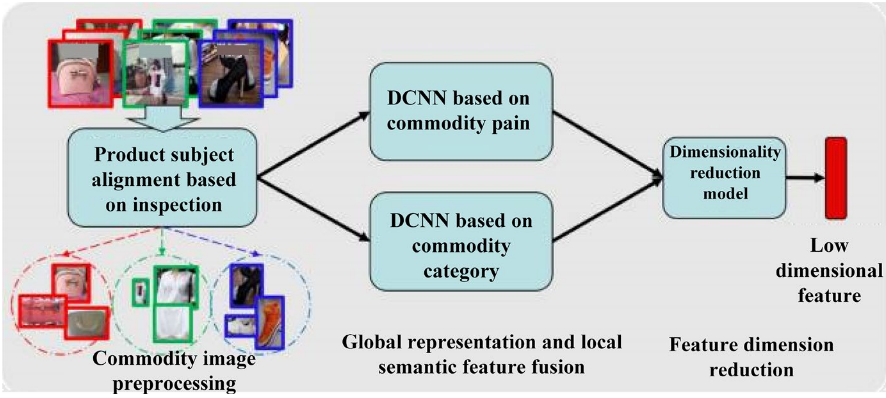

Fig. 2 Learning framework of vertical characteristics of commodities

wear, children's wear, etc.), with great differences in posture, deformation, size, etc.

(2) The fusion of global and local semantic features. There are many ways to map a commodity image to a feature. In order to make the features have good discrimination, a variety of semantic supervision information is used to guide the learning of the model.

(3) Feature dimension reduction Feature learning is a process of continuous improvement. Only features with low dimensions and good discrimination can ensure the performance and efficiency of retrieval.

3 Model framework design for product advertising recommendation

3.1 Deep learning model framework

The data used for dimension reduction learning is generally the same item data; the commonly used dimension reduction methods include linear discriminant analysis (LDA), image classification and metric learning. Based on this framework, feature learning can rely on a large number of annotation data. How to get annotation data Simple and rough full data annotation is very time-consuming and labor-consuming Here, the data mining methods are given for the same data and category data, as shown in Fig. 3. When training the deep neural network, we will encounter some super parameters. By observing the monitoring indicators in the training process, we can judge the current training state of the model. Adjusting the super parameters in time to train the model more scientifically can improve the resource utilization rate.

Figure 4 is a click rate prediction model framework based on deep learning, which is mainly divided into three parts. In the experiment of advertising click through rate prediction based on integrated learning method, the performance of three models of random forest, gradient ascending tree and limit gradient ascending tree are compared, and it is concluded that the limit gradient ascending tree

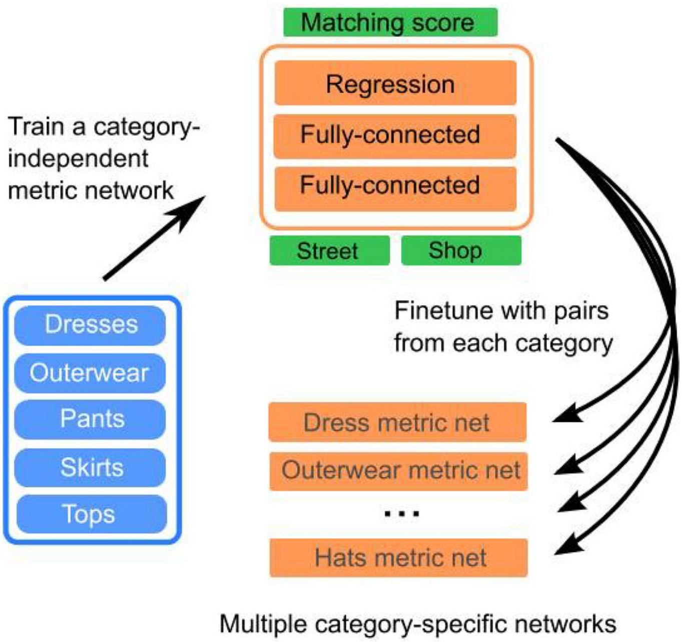

Fig. 3 Category data mining

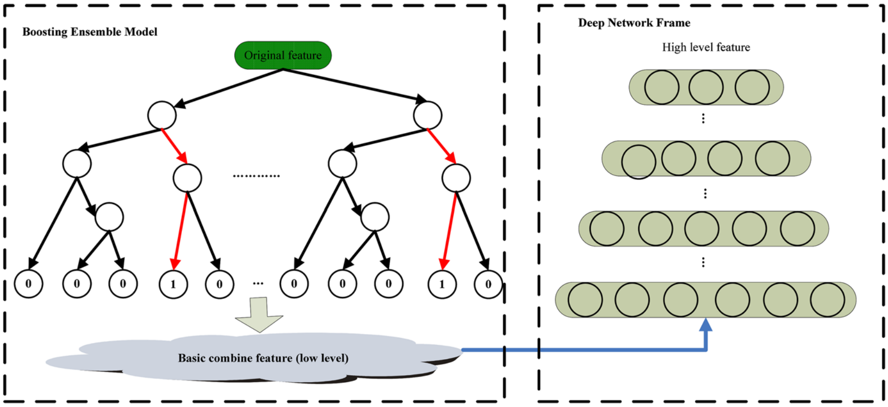

Fig. 4 Stack self-encoder and integration model to build the model framework of hit rate prediction

model is superior to gradient ascending tree and random forest in feature construction and training speed. So this paper use the integrated learning limit gradient rising tree model to construct basic multi-dimensional composite features, and then use self-encoder to extract high-level features from multi-dimensional composite features to learn those potentially abstract and effective features; finally, the click through rate prediction model is established.

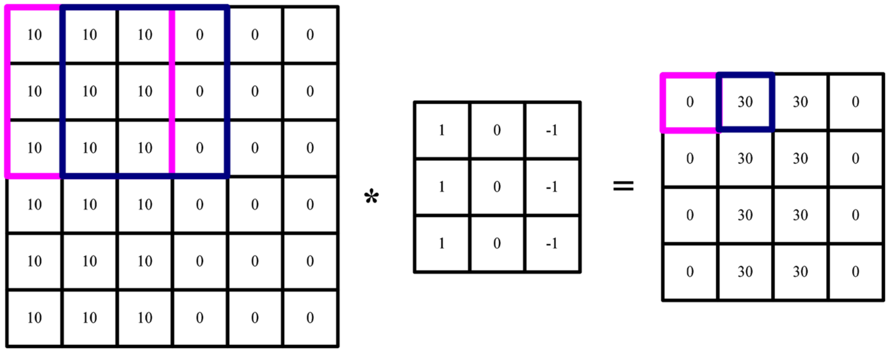

Fig. 5 Convolution operation

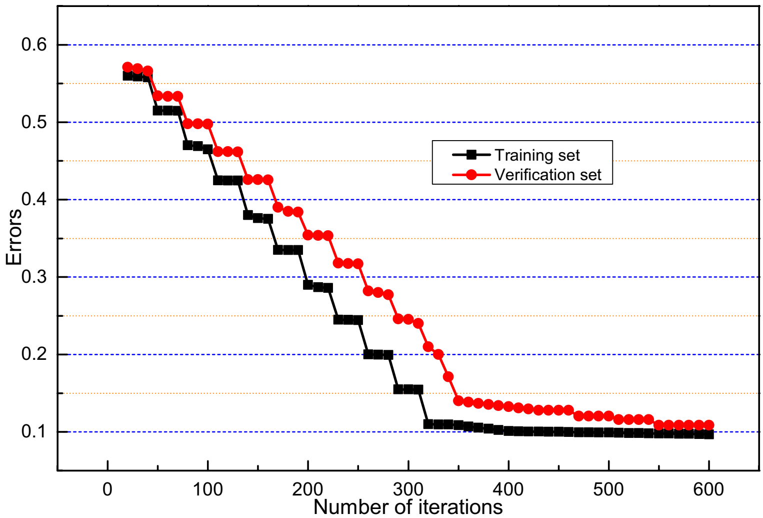

Fig. 6 Model training process

The second common network structure is convolution neural network. The parameter sharing and sparse connection mechanism of convolution operation in this network can alleviate the problem of too many parameters and easy to fall into over fitting in the full connection network. With the improvement of the network, the convolution neural network trained by back propagation algorithm can train 152 layer neural network. As shown in Fig. 5, it is a convolution module in the convolution neural network module. Through the parameter sharing mechanism, compared with the full connection network, the parameters are much less. And parameter sharing is that a feature map formed by convolution operation is convoluted by a filter and input as shown in Fig. 6. Each result in the output feature map is calculated by a set of weighted filters. The second method that makes the convolution neural network parameters less is sparse link. The median value in the red box on the right side is connected by a filter and the local pixel points in the input characteristic graph, which is only affected by the local pixel points but not all the pixel points. There is a pooling layer in the module of convolution neural network to reduce the size of the model. The method is to maintain the maximum value of a feature as long as it is extracted in a window of any filter. Large value means that some specific features may be detected, so as to transfer these features to the next layer. Various new convolution network structures based on parameter sharing and pooling are proposed and widely used. Convolution neural network is the most widely used network structure in the field of computer vision.

3.2 Performance improvement of deep neural network

In the traditional machine learning field, the trade-off between deviation and variance is needed when the parameters are adjusted, which means that the increase or decrease of one is always accompanied by the decrease or increase of the other. In depth learning, many tools are developed to adjust the trade-offs between them, such as batch normalization, random deactivation, end of training ahead of time, and traditional 1-norm and 2-norm regularization.

(1) Random inactivation

Random deactivation will traverse every layer of the network, and then according to the probability of deactivation in each layer, some neurons will be eliminated, so that they will not participate in the work of the network temporarily, and a network with fewer neuron nodes and smaller scale will be obtained.

(2) Batch normalization

For the deep learning model, the output of each layer is the input of the next layer. Batch normalization is to normalize the neural network of each layer before it is input to the activation function, or to normalize the output value of the activation function. The calculation formula is as follows:

$$u = \frac{1}{m}\sum_i z^i \tag{1}$$

$$\sigma^2 = \frac{1}{m}\sum_i \left(z^i - \mu\right)^2 \tag{2}$$

$$z^i_{norm} = \frac{z^i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{3}$$

The data is standardized by the above formula with the mean value of 0 and the unit variance, but sometimes it is not expected that the hidden unit always contains the mean value of 0 and the variance of 1. Maybe other different distributions will be more meaningful, so use the following formula to modify the distribution:

$$\overset{\sim i}{z} = \gamma z_{norm}^{i} + \beta \tag{4}$$

Parameters γ and β are the learning parameters of the model, which need to be learned in the process of model training. The distribution of data after batch normalization is controlled by these two parameters.

Batch normalization and normalization have similar functions. Input data are scanned in similar range. There are three layers of neural network, the first neuron node a_2^1 in the second hidden layer needs to receive the output from the neurons in the front layer. In the process of neural network training, forward propagation is used to calculate the error between the output value and the real value, back propagation is used to transmit the error signal and update the parameters. In this process, the parameters of each layer are changing, and the output distribution of each layer is naturally changing. From the perspective of a_2^1 neurons, a_2^1 input distribution is constantly changing, and what batch normalization does is to reduce the change of output value distribution of neurons in these hidden layers. Batch normalization ensures that these values become more stable, and the layer after the neural network will have a solid foundation.

(3) End early

In the process of training model, the loss value of the model in the training set is always in a downward trend, while the loss function will first decline and then rise after a certain node. The operation mechanism of the early end of training is to set a monitoring at the early end of training. If we monitor the performance of the verification set, if the loss value of the model on the verification set in a certain iteration has reached the optimal value. However, in the following training iterations, the loss function value of the model in the verification set does not decrease but increases, so the training of the model will end, so the loss value of the training model in the verification set is the best. Figure 6 shows the performance changes on the training set and verification set of the training model process. After adding the early training end mechanism, the training of the model will end the training process after the minimum value of the verification set increases.

There are some disadvantages when the training of the model is finished ahead of time. When the training of the model is finished ahead of time, the loss value of the model in the training data set is not small enough, and the fitting degree of the model to the training data set does not reach the optimal fitting. The mechanism of ending the training ahead of time also makes the model weigh the deviation and variance.

3.3 Recommendation system realization

Recurrent neural network (RNN) is a kind of deep learning model which can accurately model sequence data. In theory, the computing power of recurrent neural network is Turing complete. As a kind of typical sequence data, natural language has been widely used in many fields of natural language processing, such as language model, semantic role annotation, semantic representation, machine translation, dialogue and so on. In recent years, recurrent neural network and its variants (such as LSTM recurrent neural network, GRU recurrent neural network) have performed well in many fields of natural language processing.

The first layer of the network is the input layer, the input data is a one-dimensional vector, and the training data of this recommendation system includes two different types of data, the first kind of item label data. For this kind of data, this paper directly uses one hot method to encode the sparse vector, because the types of label data are not too many, so there is no need to reduce the dimension. The second is the past behavior data of users. In this paper, the objects that users click are processed into time series. The second layer of the network is the hidden layer. Through the previous analysis, this paper has known that the recursive neural network model is different from the ordinary BP neural network model, and the data in the recursive neural network is not only from the input layer to the output layer in a single direction. The data is transmitted circularly in the recurrent neural network, that is, the data in the input hidden layer includes not only the data in the current input layer, but also the output data in the previous hidden layer.

The operation process of the original RNN model is predicted at one time as shown in Fig. 7 below. X [t] represents data input at time t, H [t] represents hidden state at time t, and Y [t] represents output result at time t. In order to find the value of Y [t], we must know the value of x [t] and H [t]. X [t] is easy to obtain as input. To get h [t], we must know h [t − 1], H [t − 1] and H [t − 2] and x [t − 2]. In this way, it is easy to extend the state of the initial zero time to the t time through H.

The parameter W_xh represents the weight matrix from input layer to hidden full link. It is a two-dimensional matrix with input layer dimension * number of hidden layer nodes. Wh is the weight matrix from the hidden layer to the hidden layer. It is a two-dimensional matrix consisting of the number of hidden layer

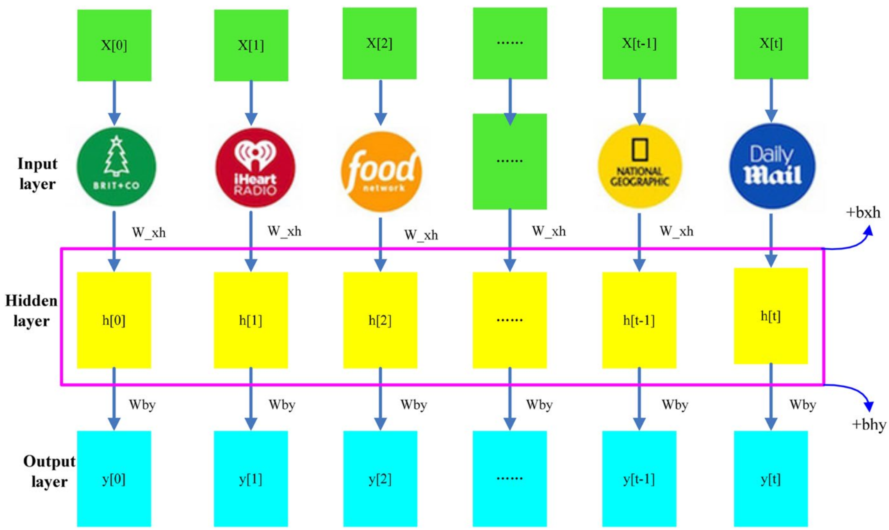

Fig. 7 Recommended diagram of recurrent neural network

nodes * the number of hidden layer nodes. Why is the weight matrix of the full link between the hidden layer and the output layer. The dimension of this matrix is the number of neurons in the output layer * the number of neurons in the hidden layer. BXH is the offset term from the input layer to the hidden layer. Its function is to generalize the linear relationship, and it is a matrix of the number of hidden layer nodes * 1. b_{hy} is the offset term from the hidden layer to the output layer. It is an offset vector of Y dimension * 1.

Using several important parameters mentioned above, we can simply create a common recurrent neural network model, and then use the random gradient descent algorithm to train the model to obtain the optimal parameters of the model. After that, we can use the recursive neural network model with the optimal parameters to recommend the next product based on the user's click sequence.

When applying recurrent neural network to the field of recommendation system, we should consider the special situation in the field of recommendation system. In fact, n should choose a reasonable value. Too large or too small T will have a great impact on the effect of recommendation. For example, what can reflect the current preferences of users may be the five to ten products that consumers have clicked before, but the possibility of the current interests of users is very small in the previous two months, or a few years ago due to the click behavior sequence of consumers. The difference between the modified recurrent neural network and the common recurrent neural network is that the improved recurrent neural network model introduces the concept of time window, in order to be able to change the size of the product sequence of consumer behavior actively. In this paper, the improved model is called W-Rnn model.

The core idea of the algorithm is to specify a time window size artificially. In the prediction of the user's current click on the item, the data satisfying the time window limit is retained, and the item sequence outside the time window is removed. It can also be said that the hidden layer within the time window is retained, while the hidden layer outside the time window is no longer transmitting the data. The detailed steps of the improved algorithm are as follows:

(1) Get the click sequence data of goods sorted by time: $\left(x_1, x_2, x_3 \cdots x_{t1}, x_t\right)$.
(2) Add time window: when the recommendation hasn't started, manually specify a time window size, reserve the data that meets the time window limit when predicting the user's current click on the item, and remove the item sequence outside the time window.
(3) Calculate the probability of the user's clicking objects at n time: first, use formula (7) to get the hidden layer data of the middle hidden layer in the time window. Then the final output of the middle hidden layer is obtained by formula (8), and then the output of the middle hidden layer is transformed into probability distribution by using the flexible maximum function.

$$hh_t = \prod_{start\_ix}^{win} w_{nn} * \left(w_{hx} x_{start\_ix} + b_{hx}\right) \tag{5}$$

$$h_t = \sigma\left(hh_t + W_{hx}x_t + b_{hx}\right) \tag{6}$$

$$\hat{y} = soft\max\left(W_{hy}x_t + b_{hy}\right) \tag{7}$$

(4) Classified prediction: according to formula (9), get the probability of each kind of goods, and select the goods with the highest probability as the output of this time.

After completing the training of neural network model, the distributed expression vector and label vector information can be input into the neural network model. Finally, the trained neural network model is used to get the probability value of users' click on each item. On this basis, the values are arranged in order of size, and the first n videos with higher scores are taken to generate a recommendation list, which is pushed to the user to complete the recommendation.

4 Performance analysis of recommended system

4.1 Implementation of e-commerce recommendation system

In a word, the large framework of e-commerce recommendation system is as follows. Implementation of e-commerce recommendation system is shown in Fig. 8.

Trigger: trigger stage, which is the source of recommendation. It can be the user's historical browsing of goods and user preferences. Generally, these data are obtained directly from the log for simple processing.

Match: in the recall stage, after we get some goods or trigger sources (categories, brands) through triggers, we associate some goods. An important part of this is

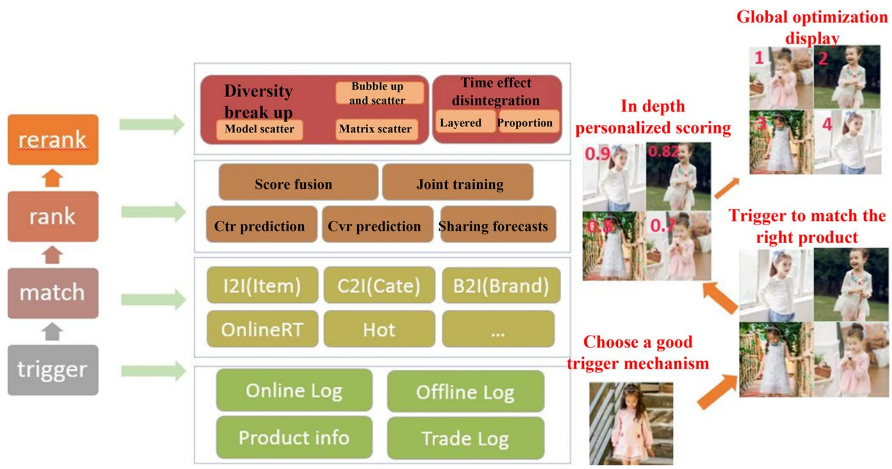

Fig. 8 Implementation of e-commerce recommendation system

through the commodity commodity relationship. This is where collaborative filtering works in recommendations.

Rank: in the sorting stage, there is a large amount of recalled goods. We can use some CTR (click through rate) and CVR (conversion rate) models to fine sort topN.

Rerank: reorder. It is mostly global. You need to reorder the results again. For example, category break-up and brand break-up are also called break-up stages.

According to the above figure 100, the following recommended process is briefly introduced:

Trigger: a user browsing a dress.
Match: the user who saw this dress also saw a lot of other dresses.
Rank: estimate the click through rate of the associated n items, and add personalization here.
Rerank: to disperse the time factor and brand factor, to avoid over concentration of brands, etc.

According to the discussion in this paper, deep learning technology in e-commerce product advertising recommendation about recommendation business and recommendation system thinking and concerns are as follows:

(1) Thinking about recommendation business
(a) Huge difference in recommended scenarios. There are great differences in recommended activity scenarios, shopping pre link and shopping post link, which need to be treated differently. The demand of shopping front link may be higher than that of shopping back link. We should cultivate the user's mind for each scenario, rather than a unified model.
(b) How to find the iteration point. Any place that seems unreasonable is an optimization point, which can be as few as Ali. Many companies' recommendations and searches can see the problem when they look carefully. This is the iteration point.
(c) Iterative process. Find an iteration point, find out where there may be gains, and evaluate the benefits in advance if possible. Develop, evaluate, verify, and think differently. If there is no promotion: when all the ideas that can be proved are proved, there is still no promotion, you can terminate this iteration and draw a conclusion. Generally, if you iterate over an unreasonable place, you can improve it. Don't give up easily.
(d) About deep learning. Good WDL may be enough for many scenarios. It may be more effective to do a good job in details than to build a model on a large scale. At the beginning of the system, it is not recommended to spend a lot of time chasing a model on a large scale.
(e) About offline assessment. It is recommended to do CTR, CVR and other indicators. So it's not easy to evaluate offline. But it's important to find a suitable indicator. For example, AUC, Gauc (average by user level) and xgboost. For example, the commonly used AUC may not be a good indicator, because the user level is different, and the click rate distribution is different. Any model can improve the AUC, but the average AUC is not always the same. So an appropri-

ate offline indicator is very important. Sometimes it needs to be compared with online multiple times to find a suitable indicator.

(f) High cost performance of recommendation system. Reuse components: componentization can facilitate development. AB test: effective online test is the basis of iteration. Monitoring: monitoring the core indicators enables us to find problems in time.

(2) Thinking about recommendation system

(a) Data is important, data is important, data is important! The same is the data. Compared with the data in NLP, the recommended data source is very miscellaneous, the controllability is not so high, and there is a lot of noise in it. It needs to be checked more carefully. When making any model, it needs to check all indicators of the data comprehensively. Whether the distribution is reasonable, numerical range, default value, data size, data source, and whether it is duplicate, etc. Be sure to have a good idea of the data before you go to the model.

(b) Exploring the case as a detective. After an iteration, if there is no effect or negative effect, carefully compare each step to find a reason to convince yourself. If you can't find it, analyze it in more detail, by user and by trigger. Just like the case investigation, the recommendation is very explanatory in many cases. It is closely linked. Don't be confused because of the complexity of the system. Where there is no explanation, there may be surprises.

(c) Rapid deployment. The recommendation system is more complex, and building a faster deployment scheme can greatly reduce the development time. After the system is completed, the algorithm students are better focused on solving problems, rather than spending a lot of time on deployment.

4.2 Comparison of recommendation effect

In this section, two groups of experiments are carried out. In the first group, we compare and analyze the recommendation effect of the recursive neural network (W-Rnn) and the recurrent neural network (Rnn) which are introduced into the time window. It is found that the W-Rnn model proposed in this paper is indeed better than Rnn (recurrent neural network model). The W-Rnn model proposed in this paper has higher accuracy than Rnn in short-term prediction and faster convergence speed. The difference between the modified recurrent neural network and the common recurrent neural network is that the improved recurrent neural network model introduces the concept of time window, in order to be able to change the size of the product sequence of consumer behavior actively. In this paper, the improved model is called w-rnn model. The core idea of the algorithm is to specify a time window size artificially, reserve the data satisfying the time window limit when predicting the user's current click on the item, and remove the item sequence outside the time window, which can also be said to retain the hidden layer within the time window, while the hidden layer outside the time window will no longer transmit the data.

Table 2 Super parameter setting of W-RNN and RNN models

Usage model	Learning rate	Number of hidden layers	Window size	Iteration times	Gradient descent algorithm
W-RNN	0.15	300	30	600	A dagrad
RNN	0.15	300	Null	600	A dagrad

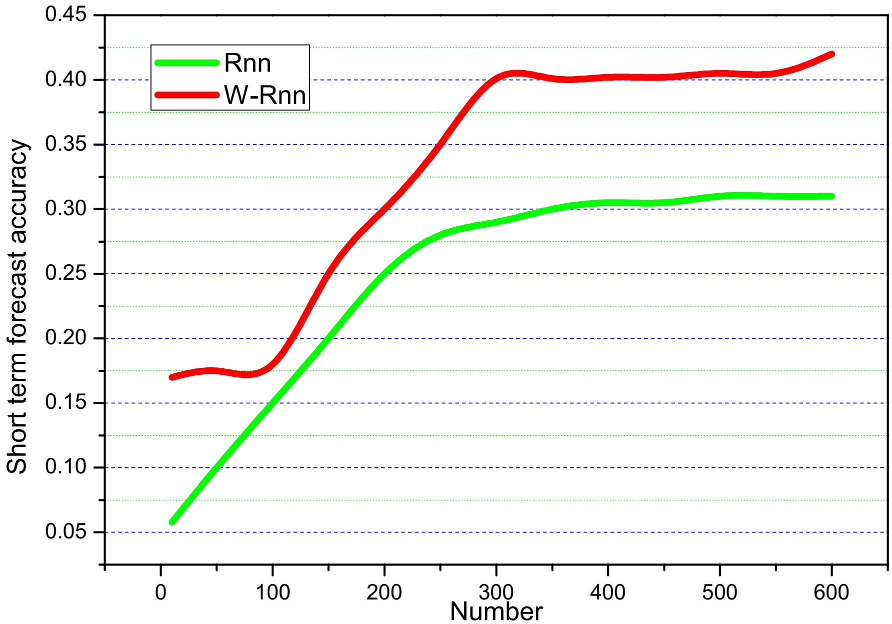

Fig. 9 Comparison of experimental results between W-RNN and RNN models

(1) Experimental comparison between W-Rnn model and Rnn model

Table 2 below is the comparison experiment of the two models under the same super parameter setting, and Fig. 8 is the comparison diagram of the two algorithms. The following experiment is to compare the effect of the two models under the condition of the same learning rate, the length of each calculation gradient, the number of iterations and the number of hidden layer nodes. The evaluation index is the short-term prediction accuracy on the test set of the model.

In the Fig. 9, the blue line is the training result chart of W-Rnn with window size of 30, and the red line is the training result chart of w-Rnn model. It can be seen from the figure that the short-term prediction accuracy of w-Rnn algorithm after convergence is higher than that of Rnn, and the convergence speed of W-Rnn is faster.

Table 3 Parameter setting of recurrent neural network model

Learning rate	Number of hidden layers	Window size	Iteration times	Gradient descent algorithm
0.15	300	30	600	A dagrad

Table 4 Alibaba mobile recommendation data set

Algorithm	Short term forecast accuracy (%)	Accuracy rate (%)	Recall (%)	User coverage (%)
MC	24.31	22.87	6.66	84.74
Item KNN	11.74	44.37	15.73	94.74
TOP	12.57	25.74	6.14	83.51
W-RNN	43.27	38.52	7.64	87.63
RNN	33.34	35.65	6.89	84.52

(2) Experimental comparison between W-Rnn model and traditional recommendation algorithm

In order to test the actual effect of the proposed algorithm, this paper compares its results with the existing classical algorithm. In this paper, we use the comparative experimental algorithms: Rnn, MC, item Knn and top. In addition to recurrent neural network, only item Knn algorithm has super parameters that need to be set manually, as shown in Table 2. Now, the optimal parameters of the recursive neural network recommendation model (W-Rnn), which introduces the concept of time window, obtained in the process of parameter adjustment of the experimental verification set, are given in Table 3 below.

This section mainly compares and analyzes the results of the five recommended algorithms in the test set of the data set. Table 4 shows the results of the recommended algorithm in the recommended data set of Alibaba mobile when the number of recommended items n is 10. And through Fig. 10, the recommendation results of various recommendation algorithms on this dataset are more intuitively expressed.

Through the data analysis, we can see that w-rnn is higher than RNN in the short-term prediction accuracy, accuracy, recall and user coverage.Through the comparative analysis of the above experimental results, comparing the performance of the recommended algorithm and the comparative experimental algorithm in multiple evaluation indexes, these conclusions can be summarized:

(1) Although the short-term prediction accuracy of the recommendation algorithm based on Markov chain is higher than that of the collaborative filtering algorithm and the popularity based algorithm, the accuracy and recall rate are relatively low. From this, we can see the limitations of the algorithm.
(2) The recurrent neural network model w-rnn with time window is superior to the common recurrent neural network in four evaluation criteria. From this we can see the necessity of introducing time window.

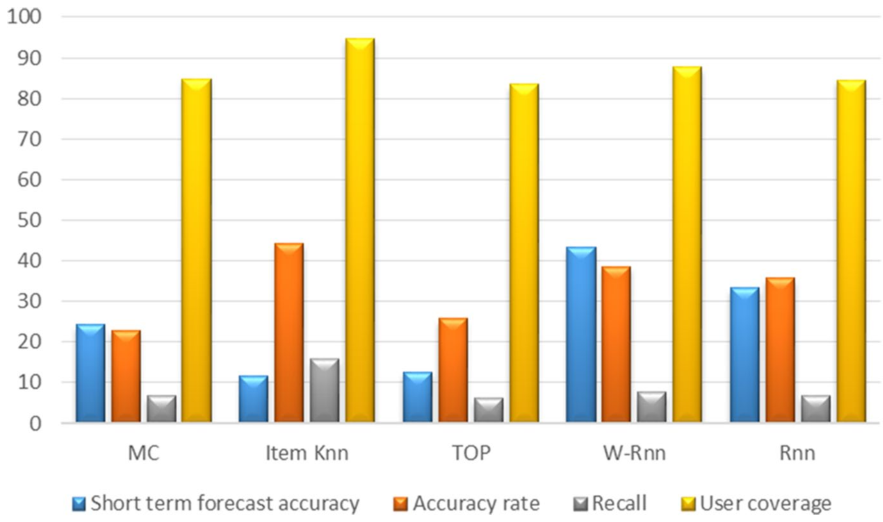

Fig. 10 Comparison of recommended algorithms

(3) How to observe these recommendation algorithms on the whole to get indicators in four evaluation criteria, it is obvious that the w-rnn model proposed in this paper has achieved good results in all evaluation indicators. And the most important thing is that the W-rnn model proposed in this paper is far higher than these comparison algorithms in the evaluation index of SPS, which shows that the recommendation algorithm proposed in this paper is more suitable for sequential recommendation than these traditional comparison algorithms.

To sum up, the performance of the same recommendation algorithm in different evaluation indexes of different recommendation algorithms is different, which reminds this paper that in order to get better results in different evaluation indexes, this paper needs to select the recommendation algorithm with better effect in this rating index. In the task of recommendation based on user's click time sequence, compared with the traditional method, this algorithm has better recommendation effect.

5 Conclusion

With the increase of the number of mobile users and the rapid development of Internet advertising, the precise advertising will become the development trend of the advertising industry, and the advertising recommendation system has become the research object of many scholars. Based on the analysis of users' interest and behavior track, this paper proposes an application method of e-commerce product advertising recommendation based on deep learning and distributed expression. Aiming at the particularity of the recommendation algorithm, this paper improves the traditional recurrent neural network. From the semantic level of advertisement, based on the theme distribution of advertisement, the similarity network of advertisement is established, and the time window is introduced to control the hidden layer data transmission of recurrent neural network. In this paper, the traditional algorithms are compared with the traditional recommendation algorithms in the real data set of e-commerce website. The experimental results show that the recommendation algorithm based on recurrent neural network and distributed expression proposed in this paper has strong time feature extraction ability, and it is obviously superior to the comparison algorithm mentioned above in the evaluation index of short-term prediction accuracy.

Compliance with ethical standards

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Lange, T., Hoefges, M., & Ribisl, K. M. (2015). regulating tobacco product advertising and promotions in the retail environment: A roadmap for states and localities. *The Journal of Law Medicine & Ethics, 43*(4), 878–896.
2. Adhikary, A. (2014). Advertising: A fusion process between consumer and product. *Procedia Economics and Finance, 11*(14), 230–238.
3. Zhu, B., Yang, C., Yu, C., et al. (2018). product image recognition based on deep learning. *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics, 30*(9), 1778–1784.
4. Nair, R., & Gupta, S. (2017). Wildfire: Approximate synchronization of parameters in distributed deep learning. *IBM Journal of Research and Development, 61*(4), 7:1–7:9.
5. Wei, J., He, J., Chen, K., et al. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications, 69*, 29–39.
6. Zhang, H., Ji, Y., Huang, W., et al. (2019). Sitcom-star-based clothing retrieval for video advertising: A deep learning framework. *Neural Computing and Applications, 31*(11), 7361–7380.
7. Park, J., Lee, S. J., Lee, S. J., et al. (2011). Online video recommendation through tag-cloud aggregation. *IEEE Multimedia, 18*(1), 78–87.
8. Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing, 107*(1), 241–265.
9. Li, M. X., Yu, S. Q., Zhang, W., et al. (2019). Segmentation of retinal fluid based on deep learning: Application of three-dimensional fully convolutional neural networks in optical coherence tomography images. *International Journal of Ophthalmology, 12*(6), 1012–1020.
10. Moeskops, P., De, J. B., Kuijf, H. J., et al. (2018). Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. *Neuroimage Clinical, 17*(C), 251–262.
11. Minarro-Giménez, J. A., Marín-Alonso, O., & Samwald, M. (2014). Exploring the application of deep learning techniques on medical text corpora. *Studies in Health Technology and Informatics, 205*, 584–588.
12. Sun, Y., Liu, D., Li, B., et al. (2019). Application of deep reinforcement learning in demand response. *Dianli Xitong Zidonghua/Automation of Electric Power Systems, 43*(5), 183–191.
13. Ardila, D., Kiraly, A. P., Bharadwaj, S., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine, 25*(6), 954–961.
14. Jeong, H. J., & Lee, M. (2014). Effects of recommendation systems on consumer inferences of website motives and attitudes towards a website. *International Journal of Advertising, 32*(4), 539.

15. Wenzhe, Y. U., Rong, Z., & Li, W. (2013). Recommendation in E-commerce. *Journal of East China Normal University, 53*(3), 46–53.
16. Xue-Qing, G., Li, W., & Xiao-Feng, H. E. (2013). Advertising targeting with user segmentation. *Journal of East China Normal University, 53*(3), 70–79.
17. Williams, G. W., Capps Jr, O., & Hanselka, D. U. S. (2017). National economic contribution of generic food and agricultural product advertising. *Journal of International Food & Agribusiness Marketing, 30*(1), 1–20.
18. Morgan, F. W., & Stoltman, J. J. (2013). Advertising and product liability litigation. *Journal of Advertising, 26*(2), 63–75.
19. Lee-Wingate, S. N., & Xie, Y. (2013). The influence of the number of presented symptoms in product-claim direct-to-consumer advertising on behavioral intentions. *International Journal of Pharmaceutical and Healthcare Marketing, 7*(3), 265–284.
20. Vandecasteele, F., Vandenbroucke, K., Schuurman, D., et al. (2017). Spott: On-the-spot e-commerce for television using deep learning-based video analysis techniques. *ACM Transactions on Multimedia Computing Communications and Applications, 13*(3s), 1–16.
21. Hao, T., Wang, Q., Wu, D., et al. (2018). Adaptive recommendation for photo pose via deep learning. *Multimedia Tools and Applications, 77*(17), 22173–22184.
22. Yi-Ou, L., Hang, L., Xiao-Yu, L. I., et al. (2017). Deep learning in NLP: Methods and applications. *Journal of University of Electronic Science and Technology of China, 46*(6), 913–919.