Deep learning techniques in e-commerce recommender systems and their impact on business marketing strategies

Hongxing Tang[1,a], Jieying Zhong[2,b], Guanlin Liu[1,3,c*]

[a]thx1390@163.com, [b]jaye_zhong@163.com, [c*]liuguanlin.article@foxmail.com

[1]Guangxi Logistics Vocational and Technical College, Guigang China
[2]Guangxi University of Foreign Languages, Nanning China
[3]Graduate University Of Mongolia, Ulaanbaatar, Mongolia

Abstract. Recommender systems could mitigate the problem of "information overload", understand the additional value of data, provide the specific information to costumer, and make information fully used. The integration of the characterization capability of deep learning (DL) with the recommendation system assists to deeply explore custumer requirements and offer details and specific recommendation services. The formulation of enterprise marketing strategy is a systematic problem that integrates various factors. DL technology can help enterprises take in opinions from their customers, optimize their marketing strategies, and improve their marketing results. This paper first introduce the merits and dismerits of conventional recommendation systems, and then review the latest achievement of DL system. At the same time, DL technology is applied to the formulation of enterprise marketing strategy. Finally, it analyzes and mention the future trend direction of intelligent recommendation systems.

Keywords: Deep Learning; E-Commerce; Recommender Systems; Enterprise Marketing; Strategy

1 Introduction

E-commerce recommender system [1,2] is an information technology-driven strategy that is widely used in e-commerce websites. Recommender systems will mitigate the issue on "information overload", dig out more on the data, push personalized information to users in need, and improve information utilization. Recommender systems could be broadly divided into three types: content filtering-based models, collaborative filtering-based models, and hybrid models. Content filtering-based recommendation models infer the degree of user preference for an item by calculating the similarity of profiles between the user and the item. The content filtering approach is adaptable to the cold start problem as compared to collaborative filtering. Collaborative filtering algorithms can be reclassified into model-based algorithms and neighborhood-based algorithms. Hybrid models integrate collaborative filtering and content filtering in a unified system and tend to have superior results. In recent years, DL technology has been widely used in recommender systems, and the integration of DL's characterization capabilities with recommender systems helps to deeply custumer requirements and offer user-defined recommendation services.

ECIT 2024, March 15-17, Changsha, People's Republic of China

DOI 10.4108/eai.15-3-2024.2346153

As a popular technology, DL [3-5] has shown unlimited potential on associate areas eg. computer vision, natural language processing, and also provides a new method for recommendation systems. With the powerful characterization ability of DL technology, we learn the hidden vector representation of users and items, mine the historical behavior data of users, the diverse data of products, and the contextual scene information, capture the potential preferences of users, and generate a more accurate personalized recommendation list for users [6-7].

In this paper, we sort out the external literature, give the definition of AI marketing, discuss its foundation, characteristics, and purpose, summarize the practical application of AI in enterprise user insight, content management, interactive placement, monitoring and evaluation, and further sort out the psychological and behavioral responses of users to data collection, placement recommendation, and human-computer interaction in the process of AI marketing. Finally, this research summarizes framework of AI marketing and discusses future research directions, to provide a reference for future research and marketing practice.

2 Traditional recommendation algorithms

2.1 Collaborative filtering

Collaborative filtering is the most common recommendation algorithm, the core idea is to synthesize the explicit feedback information of users and items and filter out the items that target users may be interested in for recommendation. The dominant methods of collaborative filtering algorithms could be seperated into user-based collaborative filtering and item-based collaborative filtering, both types of algorithms need to be based on the construction of the binary co-occurrence matrix of users and items, with the entire matrix data to predict the user's rating of the item. User-based collaborative filtering requires calculating the similarity between users, finding users similar to the target user, weighting and summing the ratings of similar users as the predicted ratings of the target user for the items, and sorting the ratings to generate a list of recommended items. Based on the co-occurrence matrix, the algorithm finds the items with high ratings from the target user, calculates the similarity between the items using the item vectors, and finally takes the similar items with high ratings as the result of the recommendation list. The collaborative filtering algorithm is interpretable and can discover new user interests, but as the size of users and items increases, the co-occurrence matrix data will become sparse, and the accuracy of similarity computation will be reduced, which affects the practical effect of the algorithm. Moreover, the head effect of recommendation results is obvious, popular items with high ratings will be recommended many times, while new items with less rating information will be recommended less often, and the algorithm's generalization ability is poor.

2.2 Matrix decomposition

The main idea of the matrix decomposition algorithm is to generate a hidden vector for objectives with decomposing the co-occurrence matrix, and use the hidden vector to represent the information, which can be used to explore the deep potential connections between them, and thus improve the prediction accuracy. The matrix decomposition algorithm uses singular value decomposition and eigenroot structure decomposition to decompose the co-occurrence matrix to obtain the user and item hidden vectors, respectively. The dot product between the user vector and the item vector is the user's prediction score for the item, and the difference between the predicted value and the true value of the item is used as the loss function, as shown in Eq. (1): the user's prediction score and the true value of the item are the loss function.

$$e_{ui} = r_{ui} + q_i^T p_u \tag{1}$$

Where, is the label of user's true rating of item i, and are the user and item vectors respectively, and the dot product of and is the predicted rating of user u on item i. The model is trained using a gradient descent algorithm, and a regularization term is added to prevent overfitting. A gradient descent algorithm is adopted to train the model, and a regularization term is implemented to prevent overfitting.

2.3 Logistic regression model

Collaborative filtering and matrix decomposition algorithms only utilize the interaction information between users and items, while logistic regression models can integrate user profile features, item attributes, and contextual information, convert the features into numerical vectors, input them into the network for training, learn the weight of each feature, and then the output layer predicts the probability of the sample being positive.

2.4 Factorization machine model

The FM algorithm introduces a low-dimensional dense hidden vector feature for each feature, and uses the inner product of vector features as the weight of feature crossover, as in equation (2). The correlation between two features can be measured even if they co-exist in a small amount of data, thus alleviating the problem of difficult to compute feature interactions due to sparse data. Compared with the logistic regression model, the FM model is more expressive. However, it is limited by the combinatorial explosion problem, which prevents the combination of features from being extended to the third order and above.

$$\dot{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{w_0} + \sum_{i=1}^{n} \boldsymbol{w_i x_i} + \sum_{i=1}^{n} \sum_{j=n+1}^{n} < \boldsymbol{v_i}, \boldsymbol{v_j} > \boldsymbol{x_i x_j} \tag{2}$$

Where $\boldsymbol{w_0}$ is the global bias, $\boldsymbol{w_i}$ denotes the weight of the ith feature, $< \boldsymbol{v_i}, \boldsymbol{v_j} >$ is the inner product of the hidden vectors of the features, and the value of the inner product is used as the weight of the feature crossover, and the final prediction value $\dot{\boldsymbol{y}}(\boldsymbol{x})$ is the sum of the first-order features and the second-order crossover features. Based on the FM model, the FFM model groups the features with the same properties into the same domain and refines the representation of the feature combination.

2.5 Combined GBDT+LR model

Considering the gradient boosting decision tree and logistic regression, a combined model is used to accomplish the recommendation task, and the model structure is shown in Fig. 1. One main idea of the model is to use the gradient boosting decision tree for automated feature engineering, extracting important features and feature combinations, and the last leaf node of the tree generates new discrete features, which are used as inputs for the logistic regression model to output predictions after activation functions. The proposed combination model advances the feature engineering modeling process, reduces the workload of manual feature combination and feature screening, and realizes end-to-end model training. The comparison is provided in Table 1.

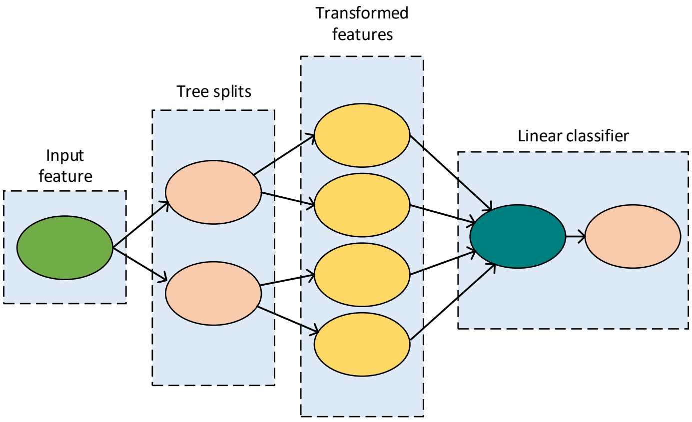

Fig. 1. GBDT+LR Combined Model Structure.

Table 1. Comparison of algorithms for DL in recommender systems

Recommendation Algorithms	Advantage	Disadvantage
Collaborative filtering	Simple; No domain knowledge required; uncover new points of interest	Sparse data; Cold start problem; Significant header effect
Matrix decomposition	Enhanced generalization capabilities; Mitigates data sparsity problem	Loss of other relevant user items and contextual information
Logistic regression	easy to implement; Learning individual feature weights, with interpretability	Poor expressive; No feature combination and feature filtering
Factorizer	Solve sparse data cross-feature combination problem; Enhance the expression ability of the model	Model has many parameters, difficult to train; Overfitting
Gradient boosting tree + logistic regression combination model	Automated feature combination; End-to-end training; Reduce manual feature combination	Poor generalization ability; overfitting

3 DL Technology

DL technology has achieved a lot in the field of artificial intelligence, DL and recommender system, can alleviate the problem of insufficient expression ability of the traditional recommendation model [8-9]. DL has a stronger characterization ability, requires a large amount of data to train the model, and can alleviate the problem of large data size and data sparsity [10]. The basic structures of DL are: multilayer perceptron, convolutional neural network (CNN), recurrent neural network (RNN), attention mechanism, etc.

3.1 Multilayer perceptron

Multilayer perceptron is a feed-forward structure of neural networks, the data through the input layer, through a number of hidden layers, converge into the output layer to calculate the final result, the network structure as shown in Fig 2, the use of BP backpropagation algorithm to supervise the training of neural networks, adjust the weight of each layer of neurons, fit the nonlinear function, reduce the error between the estimated value and the true value. Multi-layer perceptron is often used in recommender systems to mine higher-order feature intersections and learn latent data patterns.

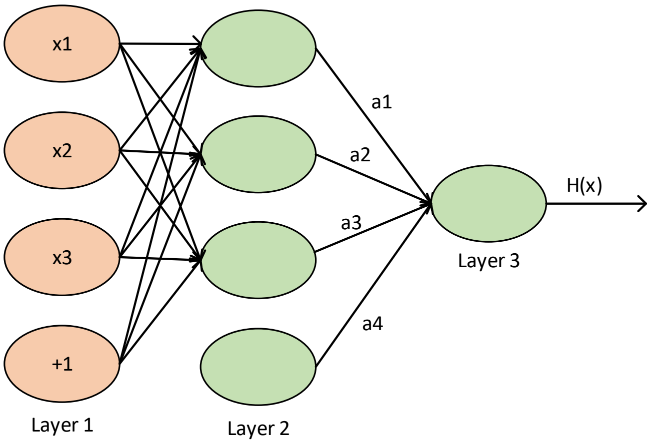

Fig. 2. Multilayer perceptron architecture.

3.2 CNN

CNN is one of the most important and commonly used algorithms in DL and it has impelemented achievements in the fields of image recognition, computer vision and natural language processing. CNN use the principles of linear algebra to recognize patterns within an image.CNNs extract features from the input data through operations such as convolution, pooling, and full connectivity, and then feed these features into a classifier for classification. The core building block of a convolutional neural network is the convolutional layer, which is responsible for performing most of the computation. The convolutional layer requires several components, including input data, filters, and feature maps. The convolutional layer is the first layer of a convolutional network, which requires a convolution operation to detect features in the input data. The pooling layer, also known as the downsampling layer, performs dimensionality reduction operations designed to reduce the number of parameters in the input. The fully connected layer, whose name aptly describes it, performs the classification task based on the features extracted through the previous layers and their different filters. Structural variations of convolutional neural networks can be achieved by adding or subtracting convolutional, pooling and fully connected layers.

3.3 RNN

RNN is a type of neural network with short-term memory and is suitable for temporal-related problems such as video, speech, and text. In RNN, neurons receive information from other neurons and themselves, forming a network structure with loops. The parameters of the recurrent neural network could be obtained by the backpropagation over time algorithm, i.e., the error is passed forward step by step in the reverse order of time. The principle of RNN is given in Fig 3.

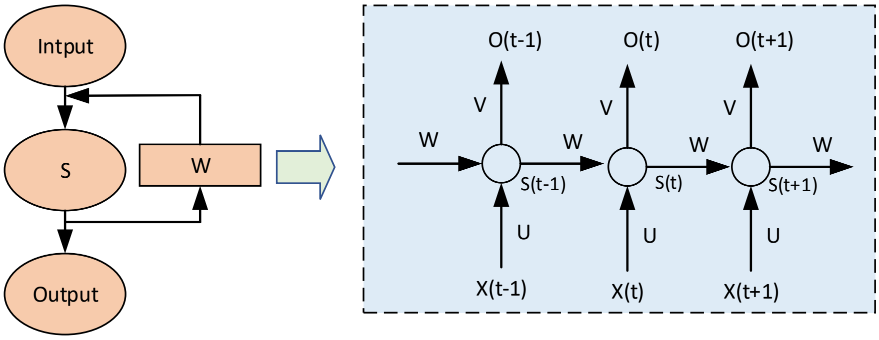

Fig. 3. RNN architecture.

3.4 Attention mechanisms

The attention mechanism is a method in DL that mimics the human visual and cognitive system with the NN to pay attention to the input data as it is processed. By introducing the attention mechanism, the NN is able to automatically learn and selectively focus on the key point in the input, improving the performance and generalization of the model

4 Algorithmic construction and experimental results

Reinforcement learning is mainly a dynamic trial-and-error mechanism to continuously interact with the environment and learn how to obtain the optimal behavioral strategy. Therefore, during the interaction with the environment, the agent not only needs to consider the action that maximizes the value function, i.e., exploitation, but also needs to choose as many different actions as possible to find the optimal strategy, i.e., exploration. Table 2 describes the algorithmic flow.

Table 2. Dueling bandit gradient descent

Step 1:	Import γ, δ, w_t
Step 2:	For query qt (t=1,2,...,T) do
Step 3:	The sample unit vector is u_t
Step 4:	$W_t' \leftarrow P_w(w_t + \delta u_t)$
Step 5:	Compare w and W_t'
Step 6:	If the character of W_t' is better than w
Step 7:	$w_{t+1} \leftarrow P_w(w_t + \delta u_t)$
Step 8:	Else
Step 9:	$w_{t+1} \leftarrow w_t$
Step 10	End

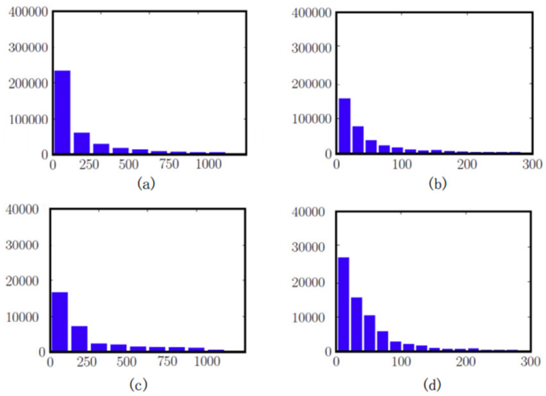

Fig. 4 Basic statistics of households and commodities (a) Number of times users requested access to recommended commodities in the dataset; (b) Number of times each commodity in the dataset was recommended; (c) Number of times users requested access to recommended commodities in the on-line recommendation dataset; (d) Number of times each commodity in the on-line recommendation dataset was recommended.

As shown in Fig. 4, the basic data statistics of users and commodities, Fig. (a) is the number of times each user requests access to recommended commodities, the horizontal axis indicates the number of times each commodity is accessed, and the vertical axis indicates the number of users, Fig. (b) is the number of times each commodity is recommended, and the horizontal axis represents the number of commodities recommended, and the vertical axis indicates the number of commodities, and it is found that the two sets of datasets are skewed by observing Fig. (a) and Fig. (b). By observing Figure (a) and Figure (b), it is found that both data sets are skewed, indicating that the number of times users visit the products has long-tailed distribution characteristics, i.e., most of the users visit less than 500 times, and the number of times each product is recommended also has long-tailed distribution characteristics, and most of the products are recommended less than 100 times.

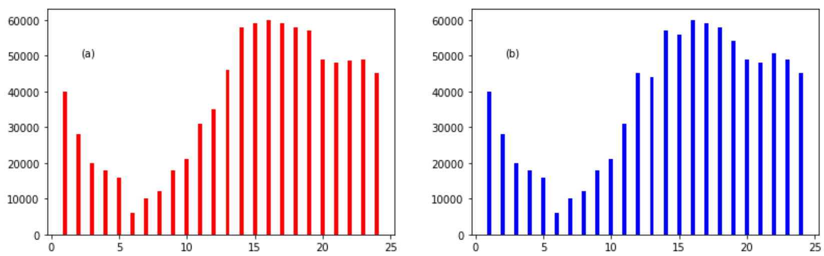

Fig. 5 User and Product Interaction Time (a) Offline Data Set (b) Online Data Set

Fig. 5 (a) and (b) represent the statistics of user and product interaction time in the recommender system dataset and online recommender dataset respectively, where the horizontal axis represents 24 hours a day and the vertical axis represents the number of times user behaviors (clicking/adding to the cart/purchasing) occur, and the number of times user behaviors occur from 0:00 to 6:00 is in a decreasing trend, and the number of times user behaviors occur from 7:00 to 16:00 is in an ascending trend, and the number of points from 17:00 to 24:00 firstly appears to be in a declining trend, and then tends to be stable after a small fluctuation, which is basically in accordance with the normal work and rest time of the human being.

5 Conclusion

Nowadays, the amount of data on the Internet soar up, the accompanying "information overload" problem significantly show up, the recommendation system serves as a critical part in alleviating the problem of information overload. DL technology is integrated with a recommender system to construct a model that fits the user's interest and produce a unique recommendation table. In comparison on conventional recommendation algorithms, DL enhances the scalability and characterization ability of the model, allowing the model to incorporate more diverse features, capture the user's interests, and improve the performance on accuracy. This paper can help researchers in the field of recommendation algorithms to clarify the situation and provide useful help. In the future, the direction of the development of intelligent recommender systems will be more focused on the application of DL, but also need to address the challenges of DL technology in recommender systems.

References

[1] Wei, K., Huang, J., & Fu, S. (2007, June). A survey of e-commerce recommender systems. In 2007 international conference on service systems and service management (pp. 1-5). IEEE.

[2] Hu Q, Zhu DJ, Wu HL, Wu LH. Survey on Intelligent Recommendation System. Computer Systems and Applications, 2022, 31(4): 47-58(in Chinese).http://www.c-s-a.org.cn/1003-3254/8403.html

[3] Shoja, B. M., & Tabrizi, N. (2019). Customer reviews analysis with deep neural networks for e-commerce recommender systems. IEEE Access, 7, 119121-119130.

[4] Almahmood, R. J. K., & Tekerek, A. (2022). Issues and Solutions in Deep Learning-Enabled Recommendation Systems within the E-Commerce Field. Applied Sciences, 12(21), 11256.

[5] Anil, D., Vembar, A., Hiriyannaiah, S., Siddesh, G. M., & Srinivasa, K. G. (2018, December). Performance analysis of deep learning architectures for recommendation systems. In 2018 IEEE 25th International Conference on High Performance Computing Workshops (HiPCW) (pp. 129-136). IEEE.

[6] Gu, Y., Ding, Z., Wang, S., Zou, L., Liu, Y., & Yin, D. (2020, October). Deep multifaceted transformers for multi-objective ranking in large-scale e-commerce recommender systems. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (pp. 2493-2500).

[7] Addagarla, S. K., & Amalanathan, A. (2021). e-SimNet: A visual similar product recommender system for E-commerce. Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 22(1), 563-570.

[8] Shankar, D., Narumanchi, S., Ananya, H. A., Kompalli, P., & Chaudhury, K. (2017). Deep learning based large scale visual recommendation and search for e-commerce. arXiv preprint arXiv:1703.02344.

[9] Gu, Y., Ding, Z., Wang, S., & Yin, D. (2020, January). Hierarchical user profiling for e-commerce recommender systems. In Proceedings of the 13th International Conference on Web Search and Data Mining (pp. 223-231).

[10] Da'u, A., & Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. Artificial Intelligence Review, 53(4), 2709-2748.