Atlantis Highlights in Engineering, volume 5

International Conference on Precision Machining, Non-Traditional Machining and Intelligent Manufacturing (PNTIM 2019)

Shopping Recommendation System Design Based on Deep Learning

Li Haihan*
Department of Computer Science and Technology
Century College
Beijing University of Posts and Telecommunications
Beijing, China
E-mail: 1039984502@qq.com

He Nana
Department of Computer Science and Technology
Century College
Beijing University of Posts and Telecommunications
Beijing, China

Qi Guanglei
Mobile Media and Cultural Computing Key Laboratory of Beijing
Century College
Beijing University of Posts and Telecommunications
Beijing, 102101, China

Dong Xinri
Department of Computer Science and Technology
Century College
Beijing University of Posts and Telecommunications
Beijing, China

***Abstract*—With the development of internet shopping, the amount of user data generated is increasing day by day. In this paper, a shopping recommendation system based on deep learning is constructed. The user data crawling module and shopping recommendation module are mainly designed. Firstly, obtain important user review information and product information from Jingdong Mall by python crawler and build a user data crawling module. Then a shopping recommendation system was constructed based on deep learning, combined with recommendation algorithm, The system extracts the characteristics of users and commodities through neural network algorithms, proposing a coupled recommendation algorithm (referred to as U-S recommendation algorithm) based on user characteristics and product similarity. The algorithm calculates the best match rate between users and commodities. The results show that the proposed algorithm can improve the effectiveness of the recommendation system, compared with the algorithm based on similarity of products.**

Keywords-Machine Learning; Recommended Algorithm; Pathon; Jingdong Mall

I. INTRODUCTION

With the development of Internet shopping, more and more people are already inseparable from online shopping. Facing all kinds of products, choosing the most suitable and most suitable product has become a difficult task for many users. This is especially true for female users who have selected obstacles. Therefore, the company faces the increasing amount of user data every day, recommending the most appropriate product for users and building an accurate and personalized shopping recommendation system can not wait. At present, many recommendation systems have been produced at home and abroad. Today's recommendation systems can be roughly classified into the following four categories: Collaborative filtering recommendation systems, Rule-based recommendation systems, Content-based content and Recommendation system and Hybrid recommendation system.

In the 1990s, Resnick P, Hill W, Shardanand U and others proposed the term "Collaborative Filtering". Then, Marko Balabanovic and his partners introduced a personalized recommendation system, which marked the opening of the recommendation system[1]. In 1994, Prof. Resnick of the University of Minnesota and others created the first recommendation system, Group Lens[2], which was only a recommendation for some news. In 1997, the laboratory created Movie-lens through the combination of collaborative filtering and association rules, and built a movie recommendation system, mainly using the user's rating data for the movie. Domestic researchers also conducted some research on the recommendation system. The literature[3] compares the popular Internet recommendation system. The literature[4] proposes an algorithm that combines recommendation relationships between the recommended objects. Literature[5] studies the context-aware recommendation algorithm that is suitable for scene changes.

In this paper, the current machine learning algorithm and recommendation algorithm are deeply studied. A shopping recommendation system based on deep learning is constructed. The real-user data set and commodity data set of Jingdong Mall are crawled by pathon[6], and the U-S recommendation algorithm is proposed. The convolutional neural network[7] algorithm performs feature extraction on user reviews and product features, finally calculates the best match rate between the user and the selected product, and recommends the product to the user.

II. CONSTRUCTION OF SHOPPING RECOMMENDATION SYSTEM BASED ON CNN

The shopping recommendation system based on deep learning mainly includes two modules, one is a user and product data crawling module, and the other is a shopping recommendation module. Among them, the user shopping

recommendation module uses the U-S algorithm calculation and recall set, and then according to the CNN method, the user and the product are selected according to the corresponding salient features, and finally the best match probability of the user and the product is obtained, and the product is recommended to the user.

A. *The overall framework of the shopping recommendation system*

The overall framework is shown in Fig. 1. First, the recall set is screened according to the given information, and the cosine similarity and recall set of the goods to be predicted of the commodity are calculated through the user's history review information and the text information of the product title searched by user within a recent period of time. Then the features are obtained through the convolutional neural network, and the probability of collocation between the recall set and the test set is calculated by the logistic regression model.

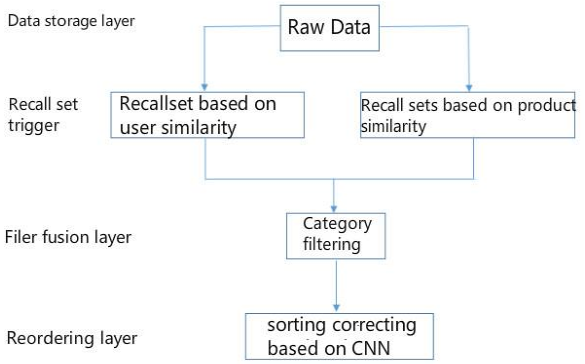


Figure 1. The overall framework of system

B. *Data crawling module*

According to statistics, Jingdong Mall will generate nearly 20TB of data each day, completing tens of thousands of transactions[8]. In this paper, data mining is implemented by scrapy framework which is widely used currently. Python is used to crawl the product information and user review information of Jingdong Mall, and a large amount of target data are obtained. The overall framework for data crawling is shown in Fig. 2.

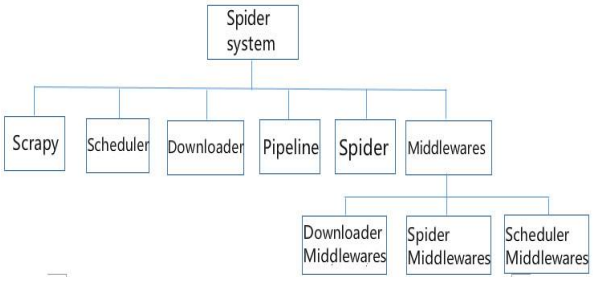


Figure 2. The overall framework of data crawling

First, the scrapy engine takes a URL link from the scheduler to grab the data, then encapsulates the URL link into a request and passes it to the downloader. The downloader downloads the resources and packages it into the response package. The spider crawler parses the response packet. If the item entity is parsed, it is passed to the item pipeline for further processing. If the URL link is parsed, the URL is passed to scheduler for fetching. Finally, the grab data is stored in the Redis and the MongoDB. The data crawl process analysis diagram is shown in Fig. 3.

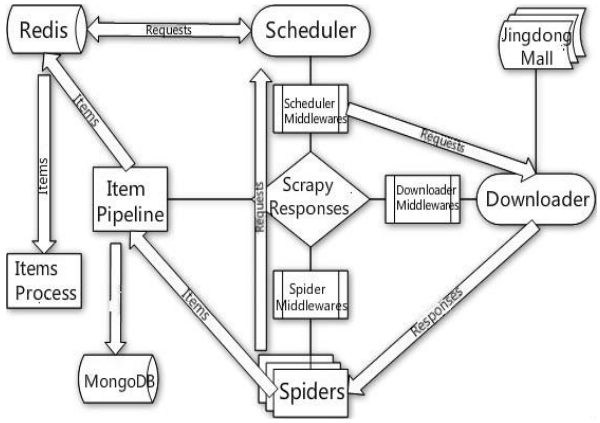
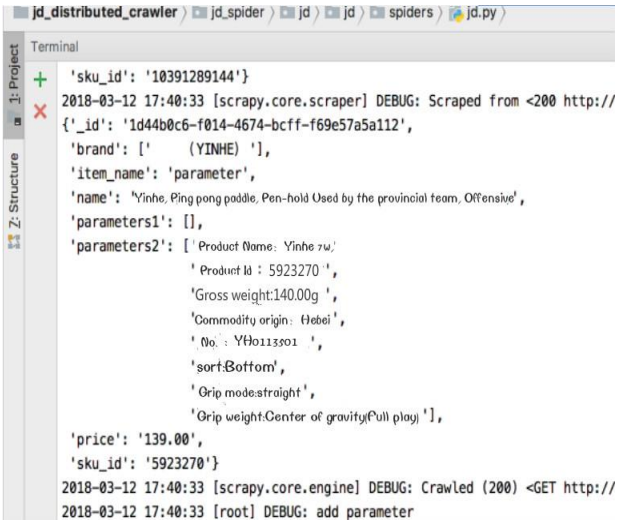


Figure 3. Data crawl process analysis diagram

This article crawls Jingdong Mall's real data set, as shown in Figure 4, where Fig. (a) is the crawled product information, and Fig. (b) is the user review information captured.



(a) Product information crawling diagram

(b) User review information crawling diagram

Figure 4. Data information crawling diagram

III. USER SHOPPING RECOMMENDATION MODULE

A. *Convolutional Neural Network(CNN)*

The Convolutional Neural Network (CNN) is a supervised neural network that is one of deep learning. The schematic diagram of the neural network is shown in Fig. 5.

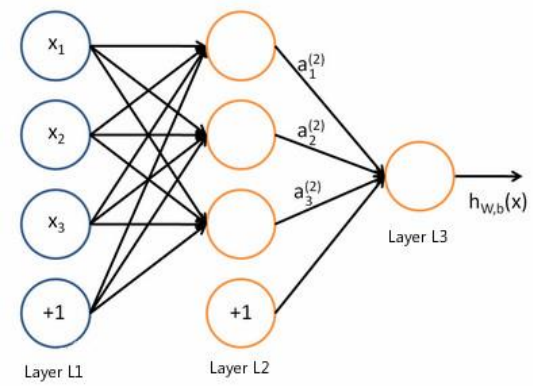


Figure 5. CNN schematic

This article uses the powerful feature extraction capabilities of convolutional neural networks to extract image features, and then passes the logistic regression model to the next step. This process is shown in Figure 6. Firstly, the image information of each commodity is extracted using the existing classical neural network model, and then the collocation data, recall set data and product image characteristic data are respectively connected, and the difference of product image characteristics is calculated. Finally, a certain non-collocated sample data is added to the collocation data as logistic regression training data to obtain the probability of predicting the recall set as a commodity collocation.

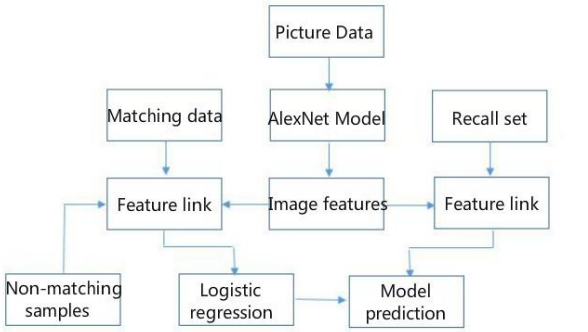


Figure 6. CNN-based process diagram

B. *U-S-based shopping recommendation algorithm*

The US-based recommendation algorithm mainly includes two parts, one is to obtain non-failure rating information on the purchase of the product by the user in the near future and the number of people who purchased the product, and the other is to obtain the title of the product for the above non-defective product or shopping cart product. The text gets the similar value of the merchandise so as to recommend the precise and appropriate merchandise to the user.

The weight is given based t_k on the frequency of the word segmentation of the product information in the text of the product, and is called a word frequency. tf_{kj} is the word frequency of the word in the document d_j. The formula is expressed as follows:

$$w_{kj} = tf_{kj} \cdots \quad (1)$$

The inverse document frequency can be expressed as follows:

$$idf_k = \log \frac{m}{n_k} \quad (2)$$

Where m is the total number of documents and n_k is the number of documents that the word t_k appears in document set D. According to formula (1) and formula (2) to obtain the weight, can be expressed as:

$$w_{kj} = tf_{kj} \times idf_k \cdots \quad (3)$$

According to formula (3), the number of people buying some kind of products can be calculated, where $(I \bullet)$ is an indicator function. Parameter It is any commodity, B_u is a collection of products purchased by user u, T_{ui} is a time of purchase of commodity i by user u, and τ is time interval. It is expressed as follows:

$$sim(It_i, It_j) \approx freq(It_i, It_j) = I_{u \in U}(It_i \in B_u \cap It_j \in B_u \mid T_{ui} - T_{uj} < \tau) \quad (4)$$

The user's rating for a product is expressed as:

$$P_{ui} = \bar{R}_u + \frac{\sum_{v \in N(u)} sim(u,v) \cdot (R_{vi} - \bar{R}_u)}{\sum_{v \in N(u)} | sim(u,v) |} \quad (5)$$

Among them, N(u) is the nearest neighbor set of user u, and $sim(u,v)$ is the similarity measure of user u and user v.

IV. Experimental Results and Analysis

This article chooses Alex Net[9] model as the special city extraction model, and Mean Average Precision (MAP)[10] index as the system evaluation index. The logistic regression model was used to output the matching probability. The maximum number of iterations of the model was 1000, the convergence error was 0.000001, L1 was regular, and the coefficient was 1. The sample for the experiment is Jingdong Mall data. The sample parameters are shown in Table 1.

TABLE I. Sample Parameters

Name	Amount
Matching information	371015
Product and Image Information	2658271
User information	68202113

The experiment first uses the features extracted from the convolutional neural network and the logistic regression model to classify and use the collocation data to evaluate its classification. A part of the sample is reserved as a verification data set during the evaluation process, and the data set does not participate in the training of the model. The results are shown in Table 2.

TABLE II. Training Results

Result	Correct rate	Recall rate
Verification result	0.56328341	0.19377105
Training results	0.57019582	0.19592814

In this experiment, the best match rate between the user and the selected product is calculated to obtain the final online MAP value of 5.17%. The experiment is compared with the literature[11], from the user recall set, the product similarity recall set and the recall set integration respectively. In these three aspects, the comparison results are shown in Figure 7. The final experimental results of this experiment show that under the condition of the sample set's balance, the features extracted using the convolutional neural network will produce better classification results in the logistic regression model, especially in the training results are close to the verification results, indicating that the model Without over-fitting or under-fitting, overall these features have a strong ability to describe goods. However, although it has improved but not much in the entire system, the main reason for this phenomenon is that the hit rate of the recall set is too low. Therefore, the U-S recommendation algorithm in this paper can improve the effectiveness of the recommendation system to a certain extent in the feature extraction of images based on convolutional neural networks.

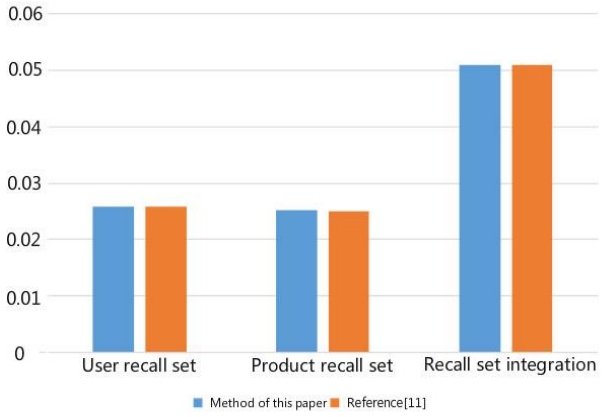


Figure 7. MAP comparison chart

V. Conclusion

With the development of the Internet, domestic and foreign companies have become increasingly keen to use data to provide personalized recommendations for users. In this paper, the current machine learning algorithm and recommendation algorithm are deeply studied. The shopping recommendation system based on deep learning is constructed. The real-user data set and commodity data set of Jingdong Mall are crawled by pathon, and the U-S recommendation algorithm is proposed. The convolutional neural network algorithm performs feature extraction on user reviews and product features, finally calculates the best match rate between the user and the selected product, and Compared with the algorithm based on similarity of products, the results show that the proposed algorithm can improve the effectiveness of the recommendation system.

References

[1] Deng Ying.Research on personalized information recommendation based on anonymous user usage mining[D]. Xi'an: Xi ' an Electronic Technology University, 2009.

[2] Resnick, P., Iacovou,N., Suchak,M., Bergstrom, P., Riedl, J.. Group Lens: an open architecture for collabo rative filtering of netnews[C]. Proceedings of the 1994 ACM conference on Computer supported cooperative work, 1994, 175-186.

[3] Xu Hailing, Wu Hao, Li Xiaodong, Yan Baoping. Comparative Study of Internet Recommendation System [J]. Journal of Software 2009,20:350-362.

[4] Guo Lei, Ma Jun, Chen Zhu-Min, Jiang Hao-Ran A Social Recommendation Based on the Correlative Relationship between

Recommendation Objects Chinese Journal of Computers, 2014(37):219-228.

[5] Wang Licai, Meng Xiangwu, Zhang Yujie.Context-aware recommendation system[J].Journal of Software.2012,01:1-20.

[6] Kim T, Haghani A, Kim H. Algorithms for One-to-One Time Dependent Shortest Pathon Real Networks[C]// Transportation Research Board 92nd Annual Meeting. 2013.

[7] Hu B, Lu Z, Li H, et al. Convolutional Neural Network Architectures for Matching Natural Language Sentences[J]. 2015, 3:2042-2050.

[8] Cavalieri R R. La rivoluzione dei Big Data:l'e-commerce in Cina[J]. Università Ca Foscari Venezia, 2014.

[9] Ballester P, Araujo R M. On the performance of GoogLeNet and AlexNet applied to sketches[C]// Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, 2016:1124-1128.

[10] Liu L, Özsu M T. Mean Average Precision[M]. Springer US, 2009.

[11] Meng Xiaolong Research on recommendation technology based on machine learning [D]. Jilin University, 2016.