Research of Recommendation System Based on Deep Interest Network

Shan-shan Wang1,*, Yuan-yuan Pan1 and Xiao Yang1
1 Anhui Institute of International Business, Hefei, 230011, China

*Corresponding author e-mail: 258693224@qq.com

Abstract. With the development of e-commerce, it provides users with many choices. But how to quickly and accurately recommend goods to users is an important topic in this field. Matrix factorization recommendation model based on scoring data is widely used, but data sparsity affects the recommendation quality of the model. In the paper, a product recommendation algorithm based on deep interest network is proposed. First, the user's purchased goods and the user's search are embedded coded, and the sparse features are transformed into low-dimensional dense features. Then, the feature vectors of the purchased goods and the search text vectors are joined together to input the deep interest network as the feature to predict the user's interest. Finally, the effectiveness of the model is validated by using record data. The validity of the model is verified by comparing with other models.

1. Introduction

With the development of the Internet and the increase of data, it is a huge challenge, how to provide information to customers. Recommendation system has become a research hotspot. At present, collaborative filtering recommendation method is based on explicit rating to model user generated information, so there is a sparse problem of rating data. User purchasing records reflects users' interests.

2. Relevant Work

2.1. Recommendation Method on Scoring Matrix

User rating data is one of the UGC information generated by users. Researchers mainly propose two kinds of recommendation methods.

User-based collaborative filtering method calculates the similarity between users and users, and finds the most similar users to the target users. Item-based collaborative filtering method is based on similarity between items to predict, the earlier application is Amazon merchandise recommendation system. Neighborhood-based recommendation methods usually focus on how to improve the traditional similarity calculation methods such as Euclidean distance and cosine distance to improve the recommendation quality.

The model-based recommendation method is to establish a mathematical model, estimate the parameters of the model by using the user rating data, and then use the trained model to predict the user's score on the item. One of the most popular methods is Latent Factor Model. A probability-based matrix decomposition method is proposed by Salakhutdinov. PMF, which added the characteristic
Gauss probability distribution to the SVD matrix decomposition and improved method of adding social information.

2.2. Recommendation Method
Wide-deep networks are trained models that have the ability to acquire memory and generalization. Memorization is based on historical data to recommend related items. Generalization realizes the diversity of recommendation. In the system, memorization generalization reflects novelty and reflects accuracy[1].

3. Product Recommendation Based on Deep Interest Network
Among various interests, the user's attention is only focused on some items. Although the wide and deep network captures all kinds of interests, only some historical data will affect whether the recommended items are purchased by users. It reflects the locality of interest[2]. So, we use product recommendation method based on deep interest network, which can not only capture diverse interests, but also focus on them.

3.1. Feature Representation
The features in UGC data of e-commerce platform are user features, user behavior, purchase items and context.
1. User features: including gender, age and other static attributes;
2. User behavior: Select text keywords that search for item records and change over time;
3. Purchase items: the characteristics of the items actually purchased by the user, one user may purchase more than one item;
4. Context: The time, date, online time of the user's purchase, etc.

We transform these features into high-dimensional sparse features by encoding. Encoding vector of i-th feature group is formulated as \( \mathbf{t}_i \in \mathbb{R}^{K_i} \), \( t_{ij} \) is the j-th element of \( \mathbf{t}_i \) and \( t_{ij} \in \{0, 1\} \). \( \sum_{j=1}^{K_i} t_{ij} = k \cdot K_i \) represents the dimensionality of feature group \( i \), It means feature group \( i \) contains \( K_i \) unique ids. \( t_{ij} \) with \( k = 1 \) refers to one-hot encoding and \( k > 1 \) refers to multi-hot encoding.

Because the search items are recorded as keyword data, the bag-of-word feature of this text is sparse. We pre-train embedded coding vectors as user input features according to a large number of user search keywords. In order to capture the short-term interest of users, we select a time period \( T \) to train the user to search for text, and the search text itself has a temporal relationship, a time period \( T \) is divided into \( m \) short time series \( \{t_1, t_2, ..., t_m\} \), each \( t_i \) is less than \( T \), and its length is determined by parameters \( n[3] \).

We can think that in each time \( t_i \), the search behavior of users is concentrated, and its temporal relationship can represent the relevance of search items. The text uses Word2vec to pre-train the text search words and get the word vector.

| CATEGORY          | FEATURE GROUP       | DIMENSIONALITY | TYPE       |
|-------------------|---------------------|----------------|------------|
| USER FEATURES     | Gender              | 2              | one-hot    |
|                   | Age                 | \( \sim 10 \)  | one-hot    |
|                   | ...                 | ...            | ...        |
| USER BEHAVIOR     | Item Key Words      | 300            | embedding  |
| PURCHASE ITEMS    | Items               | \( \sim 10^9 \) | multi-hot  |
|                   | shop                | \( \sim 10^7 \) | multi-hot  |
| CONTEXT           | Shopping date       | \( \sim 10 \)  | one-hot    |
|                   | Online time         | \( \sim 10 \)  | one-hot    |
|                   | ...                 | ...            | ...        |
In this way, the input feature used is shown in Table 1. User features and context are all one-hot coding. Due to the diversity of purchase items, it is multi-hot coding. User behavior data are encoded by pre-trained embedding features and updated in the interest network.

3.2. Baseline Model

The baseline model is composed of embedded layer and MLP. Each type of feature shares a set of parameters. The baseline model is shown in Figure 1.

First, the original input are represented high dimensional Binary Vector. Embedded layer are represented high-dimensional vectors with a dense vector of fixed dimensions. We define the i-th feature as $t_i$, define $W^i = [w^i_1, ..., w^i_j, ..., w^i_k] \in \mathbb{R}^{D \times K_i}$, represents the i-th dictionary, $w^i_j \in \mathbb{R}^D$ represents an embedded vector whose dimension is D. According to the different input vectors, the following operations are performed:

1) If $t_i$ is the j-th element $t_i[j] = 1$ of one-hot vector, then the embedded vector of $t_i$ is represented by a single vector, that is, $e_i = w^i_j$.

2) If $t_i$ is multi-hot vector and $t_i[j] = 1$, of which $j \in \{i_1, i_2, ..., i_k\}$, then the embedded vector of $t_i$ is represented by a list of vectors, that is, $\{e_{i_1}, e_{i_2}, ..., e_{i_k}\} = \{w^i_{i_1}, w^i_{i_2}, ..., w^i_{i_k}\}$.

![Figure 1. The structure of the baseline model](image)

Different users have different choices and interests in purchasing goods. Therefore, the characteristics of purchasing goods are multi-hots, and the quantity of purchased goods is different for different users, which makes the length of embedded vector output different. In order to ensure that the MLP layer inputs a fixed length vector, the pooling layer is added. We select sum pooling to perform pooling operations on embedded vector lists:

$$e_i = pooling(e_{i_1}, e_{i_2}, ..., e_{i_k})$$

Finally, all dense vectors are connected together and fed into MLP. Full connection layer is used to learn the fusion features, and softmax function is used to output the predicted items (we only select five types of goods for prediction).
3.3. **Deep Interest Network Model**

In the baseline model, only the purchase items, user features and context are used. In order to ensure that the length of the input MLP layer is fixed, a pooling layer is added. However, the operation only adds up the information of the purchased items, and the user's historical behavior does not do any processing, which can not reflect the diversity of user's interests[4]. In addition, the user search text expresses the user's historical interest, which can provide useful information for the user's purchase forecast.

Therefore, using the deep interest network, the features of purchased items are weighted based on Attention Mechanism, and the text word vectors of user history search are also weighted. At the same time, text word vectors are connected directly with other embedded feature vectors and sent into MLP, pre-trained word vectors also participate in updating parameters in deep interest networks.

After adding Attention Unit, the user’s interest expression is calculated as follows:

\[ v_u(A) = f(v_1, v_2, \ldots, v_K, e_1, e_2, \ldots, e_H) = \sum_{j=1}^{H} a(e_j, v_{A_j})e_j = \sum_{j=1}^{H} w_j e_j \]

\( \{e_1, e_2, \ldots, e_H\} \) is an embedded vector feature list of items purchased by U users. Its length is H, and \( \{v_1, v_2, \ldots, v_K\} \) is a keyword vector list of U users' historical search, with the length of K. The introduction of Attention Unit is to model Local activation. Similarly, in order to ensure that the feature dimension of the input MLP layer is fixed, we need to add the pooling layer, that is, to model Diversity. At this time, due to the introduction of Attention mechanism, the feature is weighted sum operation, which distinguishes different items and different search keywords on the weight of user interest.

4. **Experiment**

We use e-commerce data as training-testing set for one month, totaling 7 million. With time dimension as the criterion for training, the 28 days of user data as the training set, the next 2 days of data as the test set. Choose 10 minutes as the time window of search record, and train the 28-day search text with word vector pre-training, about 3 million search keywords. Tensorflow, an open source code, is used for model training.
We choose recall rate, accuracy and F1 as the indicators. But only select 5 categories of products (Mom and baby, sports, electronic, books and cosmetics) as the target. The effect of the two model on the same data set is also compared. we can see the recommendation effect based on deep interest network is better than onther model. Among them, maternal and infant women and cosmetics women prefer more, so compared with the other two categories, the classification effect is better.

| Table 2. Comparison of recall rates |
|-------------------------------------|
| DIN | Base model | Wide&Deep |
|------------------------------|------------|------------|
| Mother and infant | 0.82 | 0.69 | 0.71 |
| sports goods | 0.80 | 0.70 | 0.77 |
| electronic products | 0.79 | 0.65 | 0.62 |
| books | 0.75 | 0.67 | 0.67 |
| cosmetics | 0.81 | 0.73 | 0.76 |

| Table 3. Accuracy effect comparison |
|-------------------------------------|
| DeepFM | Base model | Wide&Deep |
|------------------------------|------------|------------|
| Mother and infant | 0.80 | 0.67 | 0.70 |
| sports goods | 0.81 | 0.74 | 0.78 |
| electronic products | 0.82 | 0.70 | 0.74 |
| books | 0.80 | 0.69 | 0.65 |
| cosmetics | 0.83 | 0.74 | 0.72 |

| Table 4. Effect comparison of F1 |
|-------------------------------------|
| DeepFM | Base model | Wide&Deep |
|------------------------------|------------|------------|
| Mother and infant | 0.81 | 0.68 | 0.70 |
| sports goods | 0.80 | 0.71 | 0.77 |
| electronic products | 0.80 | 0.67 | 0.67 |
| books | 0.77 | 0.68 | 0.66 |
| cosmetics | 0.82 | 0.73 | 0.74 |

5. Summary

In this paper, a recommendation system on deep interest network is proposed to solve how to choose the use interest on the massive e-commerce data. Firstly, the sparse features are transformed into dense vectors by emmbedding, Then, we introduce the structure of baseline model and its corresponding drawbacks, introduce the characteristics of deep interest network, carry out attention weighting processing on the features of shopping goods and the vector of historical search terms, and model Local activation, which also lays particular emphasis on the points of interest on the premise of guaranteeing the diversity of users interests. Finally, the effects of deep interest network and baseline model as well as wide and deep network are compared through experiments. From the experimental data, deep interest network has obvious advantages in the prediction of commodity purchases in various indicators.

References
[1] Cheng H. et al Wide & deep learning for recommender systems. In Pro- ceedings of the 1st Workshop on Deep Learning for Recommender Systems. ACM,2016.
[2] Guorui Zhou, Chengru Song. Deep Interest Network for Click-Through Rate Prediction. [Alibaba] 2017.
[3] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. H. Cernocky. Strategies for training large scale neural network language models. In IEEE Automatic Speech Recognition & Understanding Workshop, 2011.
[4] Kun Gai, Xiaoqiang Zhu, et al. Learning Piece-wise Linear Models from Large Scale Data for Ad Click Prediction. 2017.
[5] Tom Fawcett. 2006. An introduction to ROC analysis. Pattern recognition letters 27, 8 2006.