Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

A Hybrid Recommendation Method Combining User Profile and Product Profile

Yong Xu, Qian Wang, Shuqin Huang*, Xiaoyu Li, Xizhi Lv
School of Management Science & Engineering, Anhui University of Finance & Economics, Bengbu 233030, China
*Corresponding Author

Abstract:

In view of the cold start problem, the sparse rating and the deviation of emotions in the answers of the same rating of personalized product recommendation service in the e-commerce field, a personalized product recommendation method HRUP, which integrates user portrait and product portrait, was proposed. Firstly, user portrait was constructed from the basic attributes, interactive attributes, feedback attributes, interest attribute and situation attribute. Secondly, the product portrait was constructed based on five aspects: the user characteristics of the product, the popularity of the product, the feedback effect of the product, and the purchase situation of the product. Then, the candidate recommended product set was obtained by user similarity calculation. Next, the product similarity was used to sort and recommend the products. The experiment results show that the MAE and RMSE values of the proposed method are lower than those of the collaborative filtering method and the recommendation model based on single user profile or product profile.

***Keywords**: Personalized recommendation, User portrait, Product portrait, Sentiment analysis, Clustering analysis.*

I. MULTI-FEATURE USER PORTRAIT CONSTRUCTION

With the popularity of the Internet, problems such as "information overload" and "information trek" caused by massive data on e-commerce platforms become more and more serious, and personalized services emerge at the historic moment. In order to further improve the accuracy of personalized recommendation services, user portraits have been applied to this field. User painting was first proposed by Alan Cooper, the father of interaction design, in Why High-tech Products Drive Us Crazy and How to Restore the Sanity. A user model based on a set of real data [1]. The early application fields of user portrait are mainly concentrated in the field of interaction design or product research and development. Generally, user portrait is described as outlining the characteristics of target users' requirements for products or services, which is a kind of outlining target customers An effective tool for connecting customer demands with design direction provides a basis for product designers to concretize their subjective imagination into contour features of target users and then construct prototype systems [2,3].

In recent years, relevant scholars have carried out in-depth research on the concept of user portrait,

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

construction technology of user portrait and application of user portrait [4] At the same time, some scholars began to gradually develop the single portrait constructed in the past to the present group portrait, from the original static portrait constructed gradually to the present dynamic portrait, and the depth and breadth of research continue to expand [5-7].

As an important product of the era of big data, personalized recommendation service deeply affects every aspect of people's life. However, the current personalized recommendation algorithm still faces some problems, such as: cold start problem [8] , sparsity of rating and comment data, and deviation of emotions expressed in the same score. In view of the above problems, this paper considers the five attributes of user feedback attributes and constructs a multi-attribute user portrait model. Construct product portrait model from five aspects such as product feedback effect; In order to improve the accuracy of product recommendation and overcome the cold start problem in the process of recommendation, the product similarity is used to calculate the product recommendation candidate set, and then the product similarity is used to rank the candidate products.

II. MULTI-FEATURE USER PORTRAIT CONSTRUCTION

By mining and analyzing the attributes and behavior data of users, the user portrait can label the users accordingly to achieve accurate depiction of user needs and preferences [2, 9] . At present, user portrait is often used in personalized recommendation, precision marketing, search engine, user statistics and other aspects [10].

2.1 Construction of User Label System

User tag system is one of the key problems to realize user portrait. Only a complete and effective user tag system can show the whole picture of the user. This paper mainly designs the label system for online shopping users in the field of e-commerce, which is divided into five attribute dimensions, as shown in Fig 1.

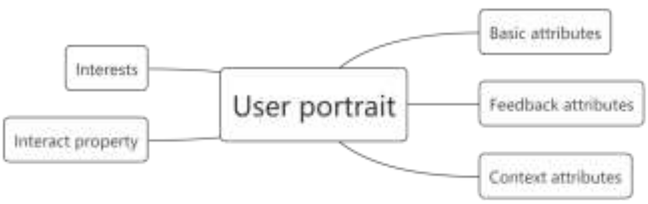

Fig 1: User label system design

The base attribute refers to the user's demographic information. Interactive attribute refers to the interactive behavior data between users, mainly including users' clicks, purchases, purchases, favorites, etc. Judge the activity of online shopping users through this attribute. The feedback attribute pointer conducts

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

sentiment analysis on the comments to get the sentiment value which can better show the user's attitude towards the product than the score. Then, the sentiment value of the score and the comment are fused to get the user's attitude towards the product purchased in the past. Interest attribute refers to the features extracted by deep mining based on the browsing and comment data of the user history. Furthermore, tF-IDF method is used to calculate the weight of interest features. Context attribute refers to the external environmental factors that may influence users' purchasing behavior. This paper lists several common external factors, including purchase time, product location and weather, etc.

2.2 Calculation of User Portrait

In order to describe users' interests and preferences more accurately, each feature in the user portrait label system must be quantified. In this paper, the user portrait model is defined as a 5-tuple in the following form.

$$\mathrm{MUP} = \{B, A, F, P, S\}$$

Where, MUP represents the user portrait model; B indicates basic user attributes. A represents the user interaction attribute; F represents user feedback attribute; P represents user interest attribute; S represents the user context attribute.

2.2.1 Basic attributes of the user

The indicators under this attribute can be directly extracted from the user's registration information and are relatively stable. Gender 0 and 1 represent male and female respectively. Age can be divided into four sections according to the classification standard of the United Nations World Health Organization, children are 18 years old and below, youth is 19-35 years old, middle age is 36-59 years old, and old age is 60 years and above, and they are represented by 1, 2, 3 and 4 respectively. Marriage can be divided into two categories, 0 means married, 1 means unmarried.

2.2.2 The interactive attributes of the user

In this paper, four indicators are selected to show the interaction: the number of clicks, the number of purchases, the number of favorites and the number of purchases. Then, the entropy weight method is used to calculate the overall interaction value A according to the collected index values [11] . The specific calculation formula is as follows:

$$A = \sum \omega_j \cdot A_{ij} \tag{1}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

$$A_{ij} = \frac{a_{ij} - \min a_{\mathrm{j}}}{\max a_j - \min a_j} \tag{2}$$

Where a_{ij} is each feature under the individual user interaction attribute; A_{ij} is obtained by data standardization processing of a_{ij}, $Mina_j$ is the minimum value of the J^{th} index, $Maxa_j$ is the maximum value of the J^{th} index.

ω_j is the weight value corresponding to each feature, which is calculated according to the uncertainty of each feature index represented by the entropy weight method, and the formula is as follows:

$$\omega_{\mathrm{j}} = \frac{1 - E_j}{\sum_{j=1}^{m} 1 - E_j} \tag{3}$$

Information entropy Ej is calculated as follows:

$$E_j = -\frac{1}{\ln(n)} \sum_{i=1}^{n} p_{ij} \ln p_{ij} \tag{4}$$

p_{ij} represents the probability that the J^{th} feature of the i^{th} user accounts for the j feature value of all users, namely:

$$p_{ij} = \frac{A_{ij}}{\sum_{i=1}^{n} A_{ij}} \quad (i=1,2,...,n) \tag{5}$$

2.2.3 User feedback attributes

In this paper, user rating SC and user comment sentiment value SE are considered as indicators of user feedback attributes [12] . The formula for calculating the feedback value is:

$$F = \alpha \times \mathrm{sc} + \beta \times \mathrm{se} \tag{6}$$

Where, when users only have ratings but no comments, α=1, β=0; Otherwise, α=β=0.5.

User score SC refers to the mean of all scores of a single user, namely:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

$$\mathrm{sc} = \frac{\sum_{\mathrm{i}=1}^{n_{\mathrm{sc}}} sc_i}{n_{\mathrm{sc}}} \tag{7}$$

The user emotion value SE refers to the average value of all comments made by a single user. Since the number of comments posted by users is inconsistent, the user's emotion value should be obtained by dividing the overall emotion of comments with the number of comments. That is:

$$\mathrm{se} = \frac{\sum_{j=1}^{n_{se}} se_j}{n_{se}} \tag{8}$$

2.2.4 The user's interest attributes

This attribute analyzes the user's interest points by collecting and processing all comments published by the user. This paper uses TF-IDF method to mine user key features and their weight values in text data.

2.2.5 Situational attributes of users

Indicators under this attribute are similar to the representation method of basic attributes, which are relatively fixed and can be obtained through directly collected data. They are mainly used as auxiliary attributes. For example, if the access source is mobile terminal or PC terminal, the mobile terminal is represented as 0 and the PC terminal as 1. The time period or season in which the action occurs.

III. CONSIDER ASPECTS OF EMOTIONAL PRODUCT PORTRAIT CONSTRUCTION

Aspect sentiment analysis is to understand the user's preference for different aspects of the product. In each comment, there is user's emotional tendency for one or more attributes of the product. Building product portraits according to the emotional value of different aspects of the product is one of the key factors to refine product features [13,14].

3.1 Aspect of Emotional Calculation

Aspect words are usually nouns or noun phrases, such as "clothes are nice in style, affordable, fast in delivery, comfortable to wear, and sizing standards will continue to be a concern." In, "style", "price", "logistics", "size" are the terms of the product. Therefore, this paper firstly selects nouns from corpus and then uses TF-IDF method to calculate the TOP 10 aspect words as the main aspect words of the product.

After obtaining the aspect words that users are mainly concerned about, it is necessary to determine the emotional polarity corresponding to the aspect words, and then determine the adjectives, adverbs and negative words that describe the word. Only by extracting the corresponding description words can the emotional intensity value be calculated. In this paper, we use dependency syntax to determine perspective pairs and then carry out emotion calculation.

Dependency parsing is to analyze the syntactic relationships between words and get syntactic structures so as to understand the relationships between words. This is very important for finding emotion words corresponding to aspect words. In this paper, the language cloud platform tool of Harbin Institute of Technology is used to analyze the syntactic dependency relationship, so as to select the required perspective pairs.

Because not all of the views on are the main aspects, so as to determine the main aspect of the word and its view of the later, need to filter the results of views on further, through the existing word to find the corresponding views on, and finally the description of the main aspects and the corresponding word.

In many product reviews, there are often many repeated aspect words, so one aspect word may contain N description words. At this point, in order to make the results of emotion calculation comparable, the frequency of occurrence of aspect words will be considered, so as to average the obtained emotion values and obtain the final emotion value.

3.2 Construction of Product Portraits

Product portrait MPP is defined as a 5-tuple.

$$MPP = \{C, H, F, P, S\}$$

Where, MPP represents product portrait model; C represents the user characteristics of the product; H represents the popularity of the product; F represents the feedback effect of the product; P represents product features; S stands for product purchase situation.

3.2.1 User characteristics of the product

The user characteristics of the product refer to the regular situation shown by the users who buy the product. For example, most of the users who buy dresses are female, so the main gender characteristics of the users of the product are female. For men's casual wear, its main user gender characteristics are male. At the same time, different types of products are suitable for different age characteristics of users, so they can also be distinguished.

3.2.2 The popularity of the product

The popularity degree of the product is one of the important factors that can affect consumers' decision

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

making, and the indexes that affect the popularity of the products, the main products are the number of comments, the number of purchases, the number of purchases and the number of collections.

In order to reduce the construction dimension of the image, entropy weight method was adopted in this paper to calculate the overall popularity value of the product, instead of comparing each influence indicator one by one. This paper divides the popularity of products into three categories: high, medium and low.

3.2.3 Product feedback effect

Product reviews and ratings are self-published product purchase opinions of users who have purchased the product. Therefore, it plays an important role to analyze reviews and ratings to get the feedback effect of the product and provide users with purchasing reference.

In this paper, the final F value is calculated by considering the emotional value of the score and the comment [15]. The specific calculation method here can refer to the calculation method of user feedback attribute in Section 1.

3.2.4 Product features

Product aspect feature P The attributes of a product described in a particular aspect.

3.2.5 The purchase situation of the product

In addition to the characteristics of users and products, the influencing factors of product purchase are also influenced by the external environment. This paper mainly considers the time of purchase, location of products, climate and other situational factors.

IV. HYBRID RECOMMENDATION MODEL

4.1 Identification of Similar User Groups

Through the multi-attribute feature user portrait model MUP, we can obtain the portrait of a single user. However, due to the large number of network users, in order to reduce the computational complexity of the recommendation process, k-means clustering algorithm is adopted to classify similar users into groups, further reducing the candidate set of recommended products and improving the recommendation efficiency. Fig 2 is a similar user group identification flow chart based on k-means algorithm.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

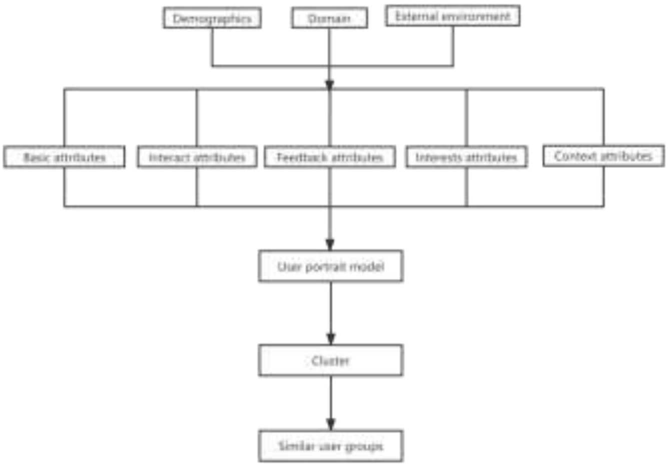

Fig 2: K-means-based similar user group identification flow chart

The user portrait model MUP is composed of user's basic attributes, interaction attributes, feedback attributes, interest attributes and context attributes, as shown in Table I.

TABLE I. User portrait model information

id	Basic		Interact	Feedback	Interests										Context
	Gender	Age	Interact	Feedback	Style	Fabric	Workmanship	Size	Color	Quility	Price	Deliver	Service	Thickness	Source
jd_1siu（1）	1	2	0.025	0.953	0.259	0.275	0.184	0.347	0.178	0.214	0.162	0.234	0.284	0.272	0
jd_1bjl（2）	0	2	0.129	0.919	0.251	0.272	0.172	0.214	0.184	0.147	0.354	0.112	0.083	0.195	1
jd_8kjs（3）	1	1	0.241	0.569	0.217	0.178	0.284	0.163	0.284	0.255	0.286	0.323	0.265	0.268	0

In this paper, TOP 10 features are calculated by product portrait as the interest feature set of all users. Because of the different dimensions, need to be normalized processing, the results are shown in Table II.

TABLE II. Normalized user portrait model information

id	Basic		Interact	Feedback	Interests										Context
	Gender	Age	Interact	Feedback	Style	Fabric	Workmanship	Size	Color	Quility	Price	Deliver	Service	Thickness	Source
jd_1siu（1）	1	0.33	0.025	0.953	0.25	0.27	0.184	0.34	0.17	0.214	0.16	0.234	0.284	0.272	0

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

		3			9	5		7	8		2				
jd_1bjl（2）	0	0.333	0.129	0.919	0.251	0.272	0.172	0.214	0.184	0.147	0.354	0.112	0.083	0.195	1
jd_8kjs（3）	1	0	0.241	0.569	0.217	0.178	0.284	0.163	0.284	0.255	0.286	0.323	0.265	0.268	0

For the normalized user portrait model information, k-means clustering was performed to obtain similar user sets.

Algorithm SU:

Input: normalized user portrait model information, number of clustering K

Output: K clusters

(1) K class clusters m_1, m_2..., m_k is initialized to empty, denoted by the set M $=\{m_{1,} m_{2,} \ldots, m_k\}$;

(2) N user vectors constructed according to the multi-attribute feature user portrait model in Chapter 3 are recorded as sets MUP$=\{MUP_{1,} MUP_{2,} \ldots, MUP_n\}$;

(3) Randomly select K users as the initial center point {P1, P2... Pk};

(4) The distance dist ($MUP^{(i)}$, P_j) between MUP_i of other users and users in the cluster center was calculated, and the similarity between MUP of each user and P_k of K cluster centers was calculated successively, and each user was assigned to the nearest cluster m_k;

(5) Recalculate the cluster center P_k according to the existing user MUP in the cluster;

(6) Repeat Step 2 and Step 3 until there is no change in clustering [8].

4.2 A Hybrid Recommendation Algorithm Integrating User Portrait and Product Portrait

The purpose of personalized recommendation is to sort the products that users may be interested in and recommend them to users according to their different interests and preferences. At present, most recommendation algorithms only study users or product characteristics to find similar users or products, without considering the important role of users and product characteristics in recommendation. This paper proposes a hybrid recommendation algorithm HRUP based on user portrait and product portrait model which is shown in Fig 3.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

The algorithm idea is as follows:

(1) Using the multi-attribute feature user portrait model to obtain the portrait MUP_i of each user;

(2) According to the algorithm SU, k-means clustering method is used to obtain similar user set SIMU for each MUP_i, and the category is denoted as C;

(3) Calculate the similarity between the target user U and other users in the category C, denoted as sim(U,U'); Cosine similarity is used in the similarity calculation method, as shown in the formula:

$$sim(U,U') = \frac{\sum_{k=1}^{n} x_{1k} \cdot x_{2k}}{\sqrt{\sum_{k=1}^{n} x_{1k}^{2}}\sqrt{\sum_{k=1}^{n} x_{2k}^{2}}} \qquad (9)$$

Where, X_{1k} represents the value of the k^{th} attribute feature of the first user, similarly, X_{2k} represents the value of the k^{th} attribute feature of the second user.

(4) According to the size of SIM (U,U'), the TOP n users are taken as the most similar user group C' of user U, and the goods purchased by users in C' are taken as the recommendation candidate set, denoted as P_1'.

(5) Build product portrait MPP_i for commodities in P_1';

(6) Calculate the similarity between MPP of product portraits purchased by target user U and all product portraits in recommended candidate set P_1'using the same calculation method as Formula (9);

(7) Calculate the predicted score value of target user U for all products in P_1'. The calculation formula is:

$$R_{p,i} = \frac{\sum_{j \in p'} r_{u,j} \times sim(p, p')}{\sum_{j \in p'} sim(p, p')} \qquad (10)$$

$R_{u,j}$ represents the scoring value of user u on product j; sim (p,p') represents the similarity result of product p and product p'.

(8) The predicted score values are sorted from large to small, and the TOP N products are taken as recommendation candidate set P, so as to realize personalized recommendation.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

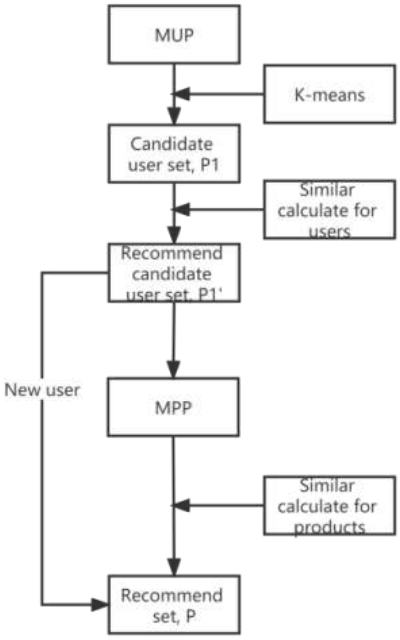

Fig 3: Recommendation system flow chart

For new users, HRUP method directly applies recommended steps 1-4. After obtaining the most similar user group C' of user U, the predicted score value of products in recommended candidate set P_1' is calculated according to the similarity of users, as shown in Formula (11):

$$R_{u,i} = \overline{r_i} + \frac{\sum_{u'\in C'}(r_{u',i} - \overline{r_{u'}}) \times sim(u,u')}{\sum_{u'\in C'}|sim(u,u')|} \tag{11}$$

Where $R_{u,i}$ is the predicted score value of user U for product I, $\overline{r_i}$ is the mean score of product i; sim(u,u') is the similarity result of user U and user U'; U' is in the most similar user group C'. $\overline{r_{u'}}$ is the mean score of user U '.

V. EXPERIMENT

5.1 Data Set

Data sources for quantitative analysis in this paper include user comment texts obtained from JINGdong mall based on Python crawler algorithm and data sets made public by Movielens provided by Grouplens group of Minnesota School of Computer Science [16], by mixing user demographic information contained in the U.user file of Movielens 100K public data set, rating data from U.data, and crawling comment text, to obtain relevant data for each user. Due to the excessive number of comments contained in

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

each product, the review data of the first 100 pages of each product is mainly collected, and the repeated comments in each product review have been deleted in the crawling process, and 26,550 comments are finally collected and kept in the file. Random numbers are used to supplement the missing data to obtain the final experimental data set. Data structure is shown in Table III.

TABLE III. Main data types

Data type	Attribute name
User's data	user id
	ugender
	uage
	uclicks
	upourchase
	ucollect
	ubuy
	uscore
	ucomments
	Access to the source
Product's data	product id
	pgender
	page
	pclicks
	pbuy
	pcollect
	ppourchase
	pscore
	pcomments
	pweather

5.2 Analysis of Experimental Results

Through multiple attribute features of the proposed user portrait model calculation method, to each user to build into A vector model of 15 d, the gender of the user by the 0 and 1, the age range is expressed as 1, 2, 3, 4, user's interaction values by the acquisition of A user clicks, the collection number, the number of purchase, and purchase number calculated by entropy weight method; The feedback value is obtained by the weighted sum of the average score and the sentiment value of the user's comments. The eigenvalues of interest are obtained by TF-IDF algorithm. It is normalized to eliminate the influence caused by dimension difference. The result is shown in Fig 4.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

```
[[1.    0.333 0.025 0.513 0.56  0.    0.    0.    0.154 0.    0.    0.393
  0.    0.265 0.   ]
 [1.    0.667 0.129 0.564 0.    0.778 0.    0.472 0.    0.414 0.    0.
  0.    0.    0.   ]
 [1.    0.333 0.241 0.62  0.562 0.375 0.152 0.304 0.314 0.399 0.    0.354
  0.213 0.    0.   ]
 [1.    0.333 0.007 0.503 0.74  0.423 0.286 0.057 0.355 0.2   0.07  0.133
  0.    0.    0.   ]
 [1.    0.333 0.23  0.615 0.536 0.    0.    0.    0.    0.508 0.    0.675
  0.    0.    1.   ]
 [1.    0.667 0.109 0.555 0.613 0.273 0.166 0.166 0.514 0.436 0.203 0.
  0.    0.    1.   ]
 [1.    0.667 0.047 0.523 0.419 0.186 0.339 0.226 0.468 0.397 0.139 0.
  0.477 0.    1.   ]
 [1.    0.667 0.116 0.558 0.534 0.317 0.192 0.    0.    0.674 0.    0.224
  0.27  0.    1.   ]
 [1.    0.667 0.145 0.573 0.    0.5   0.    0.603 0.624 0.    0.    0.
  0.    0.    1.   ]
 [1.    0.667 0.485 0.742 0.747 0.665 0.    0.    0.    0.    0.    0.
```

Fig 4: MUP vector representation of user portrait model

After all users are represented as 15-dimensional vectors, k-means algorithm is used for clustering. Here, the number of clustering clusters is set to 5, 10 and 15 respectively. Table IV-VI shows the clustering results:

TABLE IV. User clustering results of category cluster 5

User’s id	Label	Distance of cluster
1	1	0.827
2	1	0.422
3	1	0.452
4	1	0.872
5	5	0.452

TABLE V. User clustering results of category cluster 10

User’s id	Label	Distance of cluster
1	10	0.847
2	10	0.457
3	10	0.423
4	3	0.749
5	2	0.417

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

TABLE VI. User clustering results of category cluster 15

User's id	Label	Distance of cluster
1	6	0.839
2	4	0.294
3	6	0.429
4	4	0.657
5	12	0.401

Analysis Table IV-VI as you can see, the result of the user from class to class 10 to 1 1 6 class, while the user 2 by 1 class to class to class 4, 10 for these users in class cluster number 5, 10, 15, show the different category labels, candidate set according to the group of information at this time the product will also change, will affect the final recommendation. Table VII-IX compares product recommendations of different clusters.

TABLE VII. Recommended results for class cluster 5

User's id	Product's id	Score
	71	4.108
	24	4.088
1	62	4.076
	32	4.076
	66	4.066

TABLE VIII. Recommended results for class cluster 10

User's id	Product's id	Score
	71	4.108
	22	4.103
1	24	4.088
	62	4.076
	32	4.076

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

TABLE IX. Recommended results for class cluster 15

User's id	Product's id	Score
1	71	4.108
	92	4.102
	39	4.089
	24	4.088
	10	4.083

Can be seen from Table VII-IX, because of the different type of cluster, the user 1 recommended changes as a result, it is because of the need to use in the recommendation algorithm clustering to identify user 1 K - means algorithm is similar to the user, thus further according to the similarity between similar users get the preliminary recommended candidate set, therefore, class cluster for recommendation results will produce certain effect.

In this paper, evaluation indexes MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) are further selected to quantify the recommendation accuracy. MAE and RMSE are used to evaluate the deviation between the predicted value and the real value. RMSE, however, is more sensitive to outliers because of its ability to impose a greater variance penalty on those scores that deviate from larger ones. The calculation method is shown in the formula:

$$MAE = \frac{\sum_{i=1}^{N} \left| r_{ui} - \hat{r}_{ui} \right|}{N} \tag{12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (r_{ui} - \hat{r}_{ui})^2}{N}} \tag{13}$$

Where, r_{ui} is user u's score for the i^{th} product, $\widehat{r_{ri}}$ is user u's predicted score for the i^{th} product, and N is the number of recommended products.

As can be seen from the above experimental results, the hybrid recommendation algorithm HRUP proposed in this paper, which integrates user portrait and product portrait, is affected by the number of clustering clusters. Therefore, in order to find the optimal recommendation result under this data set, it is necessary to find the optimal number of clusters for different users. For example, for user 1, MAE and RMSE values are calculated when class cluster K is 5, 10, and 15 respectively. The results are shown in Fig 5 and 6.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

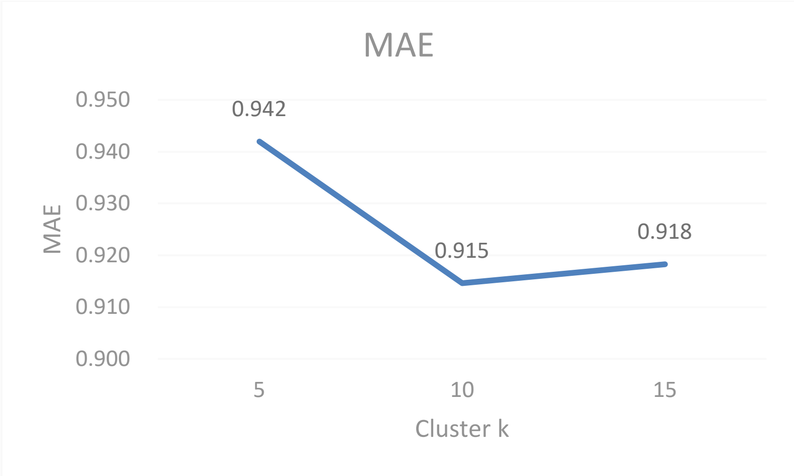

Fig 5: Comparison of MAE values under different class clusters

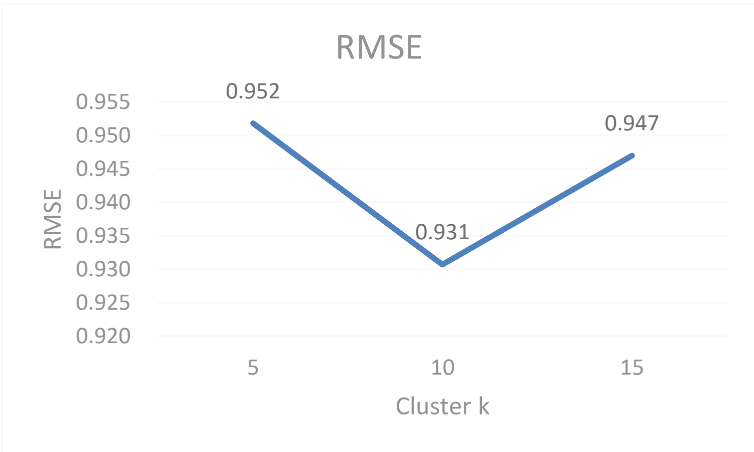

Fig 6: Comparison of RMSE values under different class clusters

It can be seen from Fig 5 and 6 that class cluster has an obvious influence on indicators. When the value of class cluster K is too small, users, as a large whole, are not distinguished, which makes the personal attribute characteristics of users not shown, unable to narrow the recommended product candidate set, can only be calculated and recommended in a large number of products. However, when the class cluster K is particularly large, each user may exist as an independent cluster, and the similarity between user groups cannot be mined to predict users' preferences for products they have not purchased. As can be seen from the figure, when the number of class clusters is 10, the values of MAE and RMSE are relatively minimum.

In addition to the influence of class cluster on the recommendation results, this experiment further compares MAE values of the following algorithms to judge the accuracy of the predicted values of the algorithm proposed in this paper.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

(1) UB-CF: This algorithm is a common user-based collaborative filtering algorithm, which only calculates user similarity based on user score matrix and then realizes recommendation;

(2) MUP-HR: this algorithm is a recommendation algorithm that only uses user portraits. The recommendation result of class cluster 10 is selected here.

(3) MMP-HR: this algorithm is a recommendation algorithm that only uses product portraits;

(4) HRUP: This algorithm is a hybrid recommendation algorithm proposed in this paper that integrates user portrait and product portrait. The recommendation result of class cluster 10 is selected here, as shown in Fig 7.

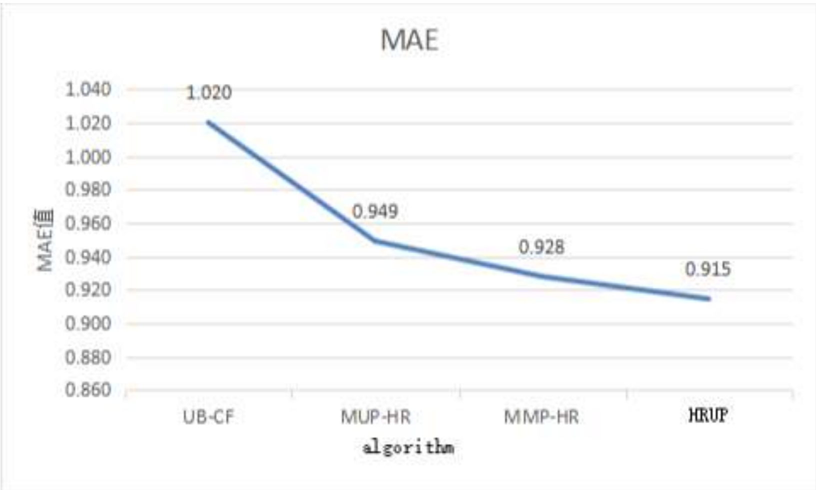

Fig 7: Comparison of MAE values under different algorithms

As can be seen from Fig 7, UB-CF only makes recommendations based on user scoring results, and MAE values are relatively high with large deviations. Secondly, compared with muP-HR, which only considers user portraits with multi-attribute features, and MMP-HR, which only considers sentiment product portraits with comments, the hybrid recommendation algorithm HRUP proposed in this paper has the lowest MAE, indicating that its recommendation results are relatively good. The reason is that when users' similarity is considered only, the recommendation is made after looking for similar users, so the deviation of predicted score value will be large. And only use a picture of the product of MMP - HR recommendation algorithm is similar to the recommendation algorithm based on content, by product, the similarity calculation between the to find related products with the past, therefore, the algorithm is requires users to have previous related data, to carry out the corresponding recommendations, so for the cold start problem is more apparent disadvantages.

VI. CONCLUSION

This paper proposes a hybrid recommendation method, HRUP, which integrates user image and product image. Firstly, based on user portrait model MUP, k-means clustering technique is used to obtain SIMU of similar user groups. Secondly, for SIMU, the similarity value of target users and users in SIMU is calculated. After sorting, the product set purchased by top N users is obtained as the recommendation candidate set P_1'. This step can reduce the product recommendation set. Then, aspect emotion is applied to quantify the differences between products, build product portrait model MPP, and calculate the degree of similarity between products purchased by target users and products in recommended candidate set P_1' according to portrait MPP. Finally, the product score was predicted according to the similarity result, and the top N products were obtained as the final product candidate set P. For new users, users' similarity can be used to calculate predicted score for products in P_1'to make recommendations. Relevant experiments verify that different values of the number of class clusters K have a significant impact on the recommendation results. Among them, for this data set, MAE and RMSE values of class cluster 10 are the smallest. At the same time, uB-CF, MUP-HR, MMP-HR and muP-MMP-HR algorithm in this paper are compared. The results show that MUP-MMP-HR algorithm has the smallest MAE value and the recommended result is relatively good.

ACKNOWLEDGMENTS

The work was supported by Anhui Provincial Natural Science Foundation Under Grant No. 1808085MF194; Philosophy and Social Science planning project of Anhui Province Under Grant No. AHSKF2021D31; The Graduate Research and Innovation Fund Under Grant No. ACYC2020365, ACYC2020368.

REFERENCES

[1] Cooper A. The inmates are running the asylum: Why high tech products drive us crazy and how to restore the sanity. The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity (2nd Edition), 2004,

[2] Gao L, Gao Q Q, Wang Hai, et al. Preference prediction method based on Datum similarity spatial Distribution Optimization. Journal of Computer Research and Development, 2018, 55(05): 977-985.

[3] Shang Y M, Cao Y N, Han Y, et al. Prediction of user dynamic behavior tendency based on theme and crowd influence. Journal of Computers, 2018, 41(07): 1431-1447.

[4] Yuan R, Wang Q. Construction and Demonstration of academic blog user portrait Model: A Case study of Sciencenet blog. Library and Information Service, 2019, 63(22): 13-20.

[5] Qiu Y F, Zhang W Z. Research on group portrait Construction method based on Network Structure and text Content. Library and Information Service, 2019, 63(22): 21-30.

[6] Shan X H, Zhang X Y, Liu X Y. User profile Based on online reviews: A case study of Ctrip Hotel. Information Theory and Practice, 2018, 041(004): 99-104,149.

[7] Li Y Y, Li X H. Research on user dynamic interest modeling based on ontology and social tag. Journal of Information Science, 2020, 39(04): 436-449.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 828-846
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

[8] Han L F. User cold start recommendation model combining user attributes and project popularity. User Cold Start Recommendation Model Combining User Attributes and Item Popularity. Computer Science, 2021, 48(02): 114-120.
[9] Xu G, Yu Z, Yao H, et al. Chinese text sentiment analysis based on extended sentiment dictionary [J]. IEEE Access, 2019, 743749-43762.
[10] Feng Y, Zhang B, Qiang B H, et al. Mn-hdrm: A Hybrid dynamic recommendation Model with long and short interest multi-neural Networks. Chinese Journal of Computers, 2019, 42(01): 18-30.
[11] Yuan R W. Construction and Demonstration of academic blog user portrait Model -- Taking Sciencenet blog as an example. Library and Information Service, 2019, 63(22): 13-20.
[12] Xu D, Tian Z, Lai R, et al. Deep learning based emotion analysis of microblog texts. Information Fusion, 2020, 64
[13] Lu Q, Zhu Z, Zhang G, et al. Aspect-gated graph convolutional networks for aspect-based sentiment analysis. Applied Intelligence, 2021, 51(7): 4408-4419.
[14] Hou Y, Yang N, Wu Y, et al. Explainable recommendation with fusion of aspect information [J]. World Wide Web, 2019, (22): 221–240.
[15] Hermanto, Kuntoro A Y, Asra T, et al. Gojek and grab user sentiment analysis on google play using naive bayes algorithm and support vector machine based smote technique. Journal of Physics Conference Series, 2020, 1641012102.
[16] Https://grouplens.Org/datasets/movielens/