Accurate Recommendation Approach of Psychological Consultation Information Based on User Portrait and TAG_SVD_CF

De-gan Zhang1,2, Zi-xian Liu1,2, Tian-zheng Fu1,2, Xin Zhang1,2, Kai-yan Wen1,2, Bing Liu1,2,*
1Key Laboratory of Computer Vision and System (Tianjin University of Technology), Ministry of Education, 300384, China;
2Tianjin Key Lab of Intelligent Computing & Novel software Technology, Tianjin University of Technology, 300384, China;
E-mail: liubing@email.tjut.edu.cn, gandegande@126.com, *Correspondence author.

Abstract- A hybrid recommendation algorithm of psychological counseling information based on user profile and item tag attribute with singular value decomposition (SVD) technology is developed. To solve the problem of data sparsity of the recommendation algorithm, the SVD technology is applied to collaborative filtering algorithm for optimizing the user item rating matrix. The recommendation algorithm includes two parts: generating medical user profile in accurate recommendation of medical information, and realizing the storage, query and update of user profile, where the index system of medical portrait was established from demographic attribute, interest label dimension and business social dimension. The performances of the developed algorithm are investigated by compared with the traditional cosine similarity and Pearson similarity. The results show that the proposed similarity has a lower mean absolute error and significantly improves the accuracy of the system.

Index terms- User profile, Psychological consultation, Information recommendation, Singular value decomposition, Collaborative filtering

1 Introduction

With the blowout development of the Internet industry, more and more ways to obtain information are available. People gradually change from active access to information to passive access to information, and the amount of information is also exploding in a geometric multiple. In today's medical security has also been significantly improved with high-tech support and complete hospital facilities. However, in the eyes of many patients, even if the hospital services are perfect, the hospital is far from being a user-oriented place[1]. Although a large number of scientific and technological achievements have been rapidly put into medical and health applications, they have never reversed the bad user experience that hospitals give patients. Unlike traditional search engines, recommendation systems can actively help people filter information and provide personalized services by using data mining, artificial intelligence and other technologies. With the increasing expansion of medical health data, recommendation algorithms have drawn attentions from many researchers due to its importance and wide application in dealing with the massive medical data collected efficiently and efficiently.

SVD is one of the most classic and commonly used recommendation algorithms in the industry and academy areas, which was investigated by many researchers from theoretical and experimental perspectives SVD is a commonly used matrix decomposition algorithm to reduce data dimension. Li Chunchun and others pointed out that SVD algorithm used in collaborative filtering recommendation shows excellent prediction accuracy and stability, and quickly becomes one of the most popular recommendation algorithms[2][3]. Wan Guangfang and others put forward svd2 algorithm to add bias to users and projects respectively. In order to reduce the number of parameters and regard user features as a function of project features, nsvd and nsvd2 algorithm are further proposed[4]. Wang Quanmin proposes an SVD++ algorithm based on implicit feedback information, which regards user-scored items as implicit information only with explicit scoring information. TimeSVD++ further considers that user and project characteristics will change with time[5]. The domestic academic circles have also made remarkable achievements in the research and application of user portraits: Fei Peng, Lin Hongfei and others have proposed a method to construct user portraits from the perspective of multi-perspective fusion framework[6]. Wang Qiangbing et al. proposed a user
portrait model that integrates content and user gesture behavior to improve the efficiency of user portrait construction[7]. In the current medical field, Momeqi and Xia Zhiping made detailed analysis on the characteristics of medical data and realized a hierarchical CF recommendation based on subject words[8]. This system is an earlier system for recommending medical information, and named Demo Meb PRS. Master of Zhejiang University, Yu Baofu and others also proposed personalized medical information recommendation based on hobbies[9], but did not propose an effective doctor recommendation and case recommendation for patient data information in a specific hospital environment.

Generally speaking, the researches on recommendation algorithms in the field of medical health have made some breakthroughs and the recommendation technology for patients' portraits to meet their own conditions is also developed. Due to the complexity of the environment in Psychological Consultation, the problems of sparse data, cold startup and user interest migration should be solved and optimized for the recommendation algorithms of psychological consultation. The traditional collaborative filtering recommendation algorithm only considers the user's product rating information for generating recommendation, which is vulnerable to the impact of many missing data scores. To solve the problem of data sparsity of the recommendation algorithm, a hybrid recommendation algorithm of psychological counseling information based on user profile and item tag attribute with SVD technology is proposed in the paper. The improving performances of the proposed recommendation algorithm will also be carried out by compared studies with the traditional algorithms.

2 User Portrait Model and Principle

2.1 User Portrait Label Model

The user medical portrait model proposed in this paper has three main features: the basic characteristics of users (demographic attributes), the medical domain characteristics of users, and the business and social dimensions of users[10]. Because in this precise recommendation system in the field of health care, the social relationship and similarity between users are the same as the user's interest tendency, which will not only affect the choice of customer goals, the accuracy of user's medical portraits, the rationality of doctors'and hospitals' recommendation, but also affect the final result. The effect of personalized recommendation system and the user's social relationship (similarity) should also be fully taken into account[11]. At the same time, it is necessary to fully tap the information of users'Internet consultation and various registration information provided by users, and to describe a skeleton and detail user (including patients and doctors) with disease knowledge base.

This article defines a triple User =< Demographics, MedAttr, Relation > to represent a user portrait information. As shown in Figure 1.

![Figure 1 Multi-level user portrait model](image)

Demographics represents the user's static attribute model, MedAttr represents the user's medical domain model, and Relation represents the user's social relationship model. Among them, MedAttr is composed of sub-models Topic and Tag at different levels. Relation is mainly represented by Similarity similarity vector[12]. In the above multi-level user portrait model, the user's business dimension is merged into tag vector Tag in the form of tag in the field of household medicine. When processing, the user's social relationship is taken out separately, which is more convenient for our calculation. Each model is described in detail below.

1) Static Attribute Vector Model

The static attribute vectors of users are Demographics =< ID, Sex, Occupation, Address, Age, IsMarried > , ID is user registered ID, Sex = <male, female>, Occupation = <Marketing, Basic Education, Doctor/Health Care, Agriculture, Lawyer, Engineer, Artist, Scientist, Writer, Programmer, Other>, Address stands for city, Age = <60, 70, 80, 90, 00, 10> is divided into six grades, Is Married = <0, 1, 2>, 0 for unmarried, 1 for married, 2 for divorced.

2) Label Vector Model in User Medical Domain

MeAttr in a triple is used to represent the medical domain label vector model of user portraits. Similarly, MedAttr is represented by a binary <Topic, Tag>. Topic is a topic vector in the medical field and a label vector in the user's medical field. Its weight is expressed by the dimension on the Topic vector. The same user's label vector is represented by Tag, and its weight is expressed by the dimension on the Tag vector. In this paper, according to the relevant knowledge in the medical field, the topic is divided into four categories: disease information, drug information,
doctor's advice information and patient evaluation to express topic information $\text{Topic} = \langle \text{Topic}_1, \text{Topic}_2, \text{Topic}_3, \text{Topic}_4 \rangle$. Because the traditional methods mostly use keywords and their values to represent the space vector model, keywords represent features and values represent weights, but too many keywords in the medical field cause problems in processing data, which is time-consuming in the updating and maintenance of images, and under each big classification. A lot of keywords have brought inconvenience to our medical user portraits. Traditional keyword descriptions lead to too long vectors, resulting in sparse data, which is not conducive to our similarity calculation. Therefore, four major classifications are used to describe the labels and weights of user portraits in the medical field. At the same time, this method is more convenient for the extension of the subject model, and the dimension of each subject vector accurately describes the user's tendency to the classification. $\text{Tag} = \langle \text{tag}, \text{weight} \rangle$ is a binary group in which tag is used to denote the key words of the tag, and weight is also used to denote the weight of the key words in the tag. That is the weight of the user on the tag, the bigger the weight is, the higher the matching score is. Tag is used as the vector of the tag. The key words in the tag are through notes. The historical medical record information provided by the users of the book and the information of online inquiry are obtained. Tag vectors provide data basis for accurate calculation of tag similarity between images. In addition, the dimension of setting label vector is not too high. In order to facilitate storage and update, the number of labels describing users is usually set to 10. In the medical field attribute vectors of $\text{MedAttr} = \langle \text{Topic}, \text{Tag} \rangle$, the acquisition of Topic vectors is mainly based on the analysis and generation of text information generated by a large number of user's inquiry data, search records, collection browsing and attention, and the corresponding Tag vectors are obtained through the relevant basic knowledge base.

(3) User Social Relations Model

As for the user's social relationship model Relation, it is mainly used to express the degree of similarity and intimacy between users and other users in the medical field, namely $\text{Relation} = \langle \text{Similarity} \rangle$. Because users with similar label vectors in both domains have similar views and emotional perceptions about certain things in the medical field. In real life, two users are similar, so one user's view and evaluation will have a certain reference effect on the behavior and preferences of another user. In recommendation system, the user social relationship model designed in this section provides a basis for personalized recommendation or recommendation algorithm rating$^{[13]}$. The closer the user social relationship is, the more similar the users are. The recommendation results based on these neighbor users are better than those of ordinary users. Here use $\langle \text{Similarity}_{u_1}, \text{Similarity}_{u_2}, \ldots, \text{Similarity}_{u_n} \rangle$ is used to represent the similarity of social relationships, and the dimension Similarity is used to represent the similarity of medical features between user $u$ and user $i$. $\text{Similarity}_{u_i} = ((\text{Topic}), (\text{Tag}))$, where $\text{Topic}$ denotes the similarity of user $u$ and $i$'s subject domain vectors. $\text{Tag}$ represents the similarity of label vectors between user $u$ and user $i$. $f(S(\text{Topic}), (\text{Tag}))$ represents the fusion of user topic domain vector similarity and tag vector similarity. In the above medical portrait social relationship model, $\text{Relation} = \langle \text{Similarity} \rangle$ is obtained through subject domain vectors and label vectors. Theme domain vectors and label vectors are derived from two vector components of the user's medical domain label model. In the medical information recommendation system, the system carries out medical portraits for each user. Each user can use Users vector to represent himself, and every patient or doctor registered on this website. The similarity of medical image label vectors between them can reflect their similarity to some extent$^{[14]}$.

### 2.2 User Portrait Generation

The generation of medical user portraits is actually the generation of Users models: demographics, relationship and MedAttr. Because demographic attributes are provided to the system by registered information of patients, the dimension is a static attribute vector that can be extracted directly without specific training.

1. Generation of User Theme Vector in Medical Domain

   In this paper, the user medical domain topic label model is mainly used to describe the patient user by synthesizing the basic information of the patient user and the text information of the inquiry, and processing the text segmentation technology into the form of each key word. But each time, it must merge with the previous topic domain vector model, that is, every time the user topic domain vector is updated. In this paper, according to the relevant knowledge in the medical field, the topic is divided into four categories: disease information, drug information, doctor's advice information and patient evaluation to express topic information $\text{Topic} = \langle \text{Topic}_1, \text{Topic}_2, \text{Topic}_3, \text{Topic}_4 \rangle$. Users have corresponding values corresponding to the degree of conformity of the subject area$^{[15]}$. The following four steps
are used to build the user topic domain vector model:

Step1: Grab the basic information, browsing behavior, search records and interactive text data that users fill in when they register on the website.

Step2: Processing the eigenvalue extraction algorithm for all the text data information obtained in Step1: including step, and constructing the text eigenvector $\text{Doc} = \langle W_1, W_2, W_3, \ldots, W_n \rangle$.

Step3: The $\text{Doc}$ feature vectors obtained by Step2 are used to compute the probability of patients' classification in the four areas mentioned above by using Naive Bayesian algorithm (1), (2). The probability vectors $P = \langle P_1, P_2, P_3, P_4 \rangle$ are used to obtain the probability vectors in the field classification.

$$P(C_i|\text{Doc}) = P(C_i) \cdot \prod_{w \in \text{Doc}} P(w|C_i)$$ (1)

$$P(w|C_i) = \frac{N(w \in C_i) + 1}{N(\text{Doc} \in C_i) + 1}$$ (2)

In formula (1), $P(w|C_i)$ represents the probability of keyword $w_i$ under the condition of $C_i$, in a specific medical field, that is, the proportion of occurrence times of keyword $f$ in field document $C_i$. $N(w \in C_i)$ represents the number of $w_i$ contained in all documents in the field of category $i$, while $N(w \in C_i) \neq 0$. In formula (2), $N(\text{Doc} \in C_i)$ represents the number of texts in the $j$ th domain classification. $N(w \in C_i) + 1$ is to avoid the denominator can not be zero.

Step4: By simply adding all the vectors of Step3's formula (1) and formula (2) about these four fields, we can get the active subject domain vectors of the patient.

It can be concluded that the calculation of subject area vector is shown in the table below.

| Start | Obtain all relevant text data provided by user registration, search history and consultation interaction information. |
|-------|------------------------------------------------------------------------------------------------------------------|
| Do    | According to the relevant text data of each medical field, the feature keywords are calculated, and then the document vector $\text{Doc} = \langle W_1, W_2, \ldots, W_n \rangle$ is obtained. By calculating the probability of formula (1) and formula (2) in each domain classification, the domain classification probability vector of all the query data texts is further generated. |
| Done  | Fuse the classification probability vectors of all texts to generate the user's subject area vectors. |

(2) Generation of User Label Features

Tag = $\text{tag<Name, tag Weight>}$ is the format of user tag features obtained in this paper. Tag Name represents the name of the tag, and tag Weight is the corresponding weight of the tag. Tag Weight represents the user’s conformity with the tag. It is trained by the medical consultation information recommendation system according to the medical history information and the interactive information provided by the user. This tag feature further provides better similarity factors for calculating the user's similarity.

The dimension of label vectors should also be considered when building user label feature vectors. That is to say, some users have more dimensions of label vectors. Therefore, the length of tags should be truncated to facilitate updating, management and maintenance[16]. Some less active registered users may have few tag features for their tags, and these vectors need to be extended appropriately. In this paper, a dimension is defined to handle the user's image tag feature attributes more conveniently, and the length of the tag is specified to be 10. When the number of dimension of user's label eigenvector is more than 10, labels should be sorted according to their weights. Only the first 10 labels with larger weights are selected as the label attribute vector of the user. On the contrary, when the number of tags is less than 10, it is necessary to consider extending the tag vector. First, the tag feature vector (N < 10, N stands for dimension) is obtained through the API interface provided by the system, and then the recommended tags provided by the system are obtained. Finally, according to the user's original label and recommendation label, the first 10 label feature attributes are selected.

(3) Generation of Indicators of User Social Relations

In this paper, we use the similarity between users to represent the characteristics of users'social relations. That is, $\text{Relation} \bowtie \text{Similarity}$ in the portrait index system, $\text{Similarity}$ is used to represent the social relationship index vector. The user portrait $\text{Similarity}$ vector is generated as follows:

User portrait vector $\text{Similarity}$ is a vector used to represent similarity. In this paper, $\langle \text{Similarity}_u, \text{Similarity}_z, \ldots, \text{Similarity}_n \rangle$ is used to represent a vector of similarity. Among them, $\text{Similarity}_u$ represents the similarity between the target user $u$ and other user $i$. The generation of the vector dimension is usually composed of two aspects. The similarity between
Topic of user \( u \) and \( \text{Topic}_u \) of other user \( i \)'s thematic medical domain tag vectors.

The component \( \text{Topic}_u \) of \( \text{Topic} \) can be directly obtained from the MedAttr in the user portrait model. In this paper, a similarity \( \text{Similarity}(\text{Topic}_u, \text{Topic}_i) \) is defined to represent the similarity of subject domain vectors between two users. \( \text{Topic}_u = \langle TP_{u1}, TP_{u2},..., TP_{un} \rangle \) is used to represent the subject vector of the target user \( u \) and \( \text{Topic}_i = \langle TP_{i1}, TP_{i2},..., TP_{in} \rangle \) is used to represent the subject vector of other users, then there is the following formula to calculate \( \text{Similarity} \):

\[
\text{Similarity}(\text{Topic}_u, \text{Topic}_i) = \cos(\text{Topic}_u, \text{Topic}_i)
\]

\( \text{Similarity} \) represents the similarity degree of label vectors of \( u \) and \( i \).

In the medical information accurate recommendation system of this paper, a label is formed for each user through the label annotation engine. The label is obtained through related text processing and feature extraction, not the final user portrait model. After getting tag vector, cosine similarity is used to calculate tag similarity between users\(^{[17]} \). But sometimes there is a synonym in the tag, that is, there is a difference in the expression. For example, the tag vector of user \( u_i \) contains "hyperglycemia, educator", while the tag vector of user \( u_k \) contains "teacher, hyperglycemia". The similarity calculated for this problem is 0, but it is not the case in fact. Therefore, it is necessary to preprocess or synonym and synonym of the collected tag features, and then calculate the similarity. First, establish a thesaurus of synonyms, and then the specific steps are as follows:

Step 1: the system itself obtains the tag library through the dialogue data training between the user and the intelligent robot. When extracting the tag, the tag attributes of user \( u \) and user \( i \) can be obtained directly from the API interface provided by the system.

Step 2: to solve the influence of synonyms, look up whether there are synonyms in the labels of user \( u \) and user \( i \) in the established synonym dictionary. If it exists, it is represented by the unified label keywords in the dictionary;

Step 3: the tag vectors obtained through step 2 are all processed tag vectors without synonyms. Get the label vector \( \text{Tag}_u \) of user \( u \) and the label vector \( \text{Tag}_i \) of user \( i \), expand the length of the vector or reduce the weight. The rule of extension is to add a blank weight to the processed label vector if a synonym exists.

Step 4: next, use the following formula (4) to calculate cosine similarity:

\[
\text{Similarity}(\text{Tag}_u, \text{Tag}_i) = \cos(\text{Tag}_u, \text{Tag}_i) = \frac{\text{Tag}_u \cdot \text{Tag}_i}{|\text{Tag}_u| \times |\text{Tag}_i|}
\]

The two similarity values \( \text{Similarity}(\text{Topic}_u, \text{Topic}_i) \) and \( \text{Similarity}(\text{Tag}_u, \text{Tag}_i) \) described in the above formula (3) and (4) will be simply superposed, and then regularized.

3 Precise recommendation algorithm

3.1 Collaborative filtering algorithm based on user rating

Because the similarity between users can not only be calculated by the user's rating of the recommended items, but also be analyzed by the user's tagging of the recommended items. When two users label the target item similarly, it can be shown that the two users have similar interests and preferences to a certain extent, that is, the two users have high similarities, and the items recommended by one user should also meet the needs of another user well\(^{[18]} \). If user \( u \) is interested in label attributes of a project, then each user in the neighborhood user set \( u \) similar to user \( u \) should have some common interests and preferences with user \( u \). For example, the information of disease and doctor in user \( u \)'s subject interest domain classification is too much, so the neighbor users of user \( u \) should be more interested in the subject interest domain related to disease and doctor. Therefore, more and more recommendation algorithms begin to use user information and project information to improve the core algorithm, and the more perfect the modeling of users and projects, the more accurate the information of recommendation results.

Collaborative filtering algorithm process integrating Tag attributes:

1. Calculated to the preference matrix \( S \) of tag attribute and the scoring data matrix \( R \) of the user. Among them, the dimension of attribute preference matrix \( S \) is \( m \times k \). \( m \) represents the number of users, the total tag feature of the whole is \( k \). \( \text{Weight}_v \) represents the total
weight of the \( j \)th tag feature of all items evaluated by active users. The \( R \) matrix is used to represent the scoring data of users. Its dimension is \( m \times n \), dimension \( m \) represents the number of registered users, and dimension \( n \) represents the number of items to be recommended. \( r_{ij} \neq 0 \) indicates that the user \( i \) has scored item \( j \) in the past period of time, and the score value is \( r_{ij} \). When \( r_{ij} = 0 \), it means that user \( i \) has not scored item \( j \) in the past period of time. The schematic diagram of the matrix is as follows:

\[
S = \begin{bmatrix}
\text{Weight}_{u1} & \text{Weight}_{u2} & \ldots & \text{Weight}_{un} \\
\text{Weight}_{u2} & \text{Weight}_{u3} & \ldots & \text{Weight}_{un} \\
\vdots & \vdots & \ddots & \vdots \\
\text{Weight}_{um} & \text{Weight}_{u2} & \ldots & \text{Weight}_{un}
\end{bmatrix}
\]

(5)

Where, \( \text{Weight}_{ui} = \sum_{j \in C} T_{ij} \), \( C \) is the set of all items scored by the user \( i \).

(2) Calculate the similarity between user tag preference vectors and tag preference vectors. According to tag attribute preference matrix \( S \) and formula (7), calculate each user's tag attribute preference vector: \( P_i = (P_{i1}, P_{i2}, \ldots, P_{in}) \). Then use formula (8) to calculate the similarity \( \text{Similarity}_{uv} \) of each user's tag preference vector, \( u \) and \( v \) represent two users.

\[
P_u = \frac{\text{Weight}_{ui}}{\text{Weight}_{u}}
\]

(7)

Where, \( \text{Weight}_{ui} \) represents the total weight of the user \( u \) to the \( i \) tag attribute of the item, that is, \( \text{Weight}_{ui} = \sum_{j \in C} T_{ij} \), \( C \) is the set of all items scored by the target user \( u \). \( \text{Weight}_{u} \) is also used to represent the total weight of all tag attributes contained in all related items commented by user \( u \), that is, \( \text{Weight}_{u} = \sum_{i \in C} \text{Weight}_{ui} \).

\[
\text{Similarity}_{uv} = \frac{\sum_{k=1}^{n} P_{uk} P_{vk}}{\sqrt{\sum_{k=1}^{n} P_{uk}^2} \sqrt{\sum_{k=1}^{n} P_{vk}^2}}
\]

(8)

(3) Calculate the final similarity. The similarity measurement method can be combined with the similarity based on the Pearson correlation coefficient and the familiarity preference of tags as the final similarity. Firstly, the Pearson correlation coefficient similarity calculation for the scoring matrix \( R \) can get \( \text{Similarity}_{k}(u,v) \); secondly, the dynamic weight weight is used to weigh the ratio of the two measurement methods \( \text{Similarity}_{k}(u,v) \) and \( \text{Similarity}_{w}(u,v) \), which is a reasonable fusion of the two similarity of user scoring and tag attribute preference. Finally, the following formula is used to measure the comprehensive similarity between user \( u \) and user \( v \):

\[
\text{Similarity}(u,v) = w \times \text{Similarity}_{k}(u,v) + (1 - w) \times \text{Similarity}_{w}(u,v)
\]

where, \( w \) and \( 1 - w \) are used to express the proportion of Pearson similarity and tag attribute preference in this recommendation algorithm, or the proportion of two similarity, the higher the proportion, the higher the importance. The value of \( w \) can be set dynamically. After a large number of experiments, we can finally find a value of \( w \), so that the performance of our recommendation system can reach the best in Mae index. The corresponding weight is the value we ultimately need as the similarity measurement parameter. Using the above formula (9), we can get the similarity matrix between the target user and other users as follows.

\[
\text{Similarity} = \begin{bmatrix}
\text{Similarity}_{11} & \text{Similarity}_{12} & \ldots & \text{Similarity}_{1n} \\
\text{Similarity}_{21} & \text{Similarity}_{22} & \ldots & \text{Similarity}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\text{Similarity}_{n1} & \text{Similarity}_{n2} & \ldots & \text{Similarity}_{nn}
\end{bmatrix}
\]

(9)

Among them, \( u \) and \( v \) represent two target users. Their intersection of scoring items is represented by \( C \), and \( \bar{r}_{ui} \) and \( \bar{r}_{vi} \) represent the average score of \( u \) and \( v \) in all their items respectively, that is, the ratio of the total score of the user to the number of items evaluated. \( w \) and \( 1 - w \) are used to express the proportion of Pearson similarity and tag attribute preference in this recommendation algorithm, or the proportion of two similarity, the higher the proportion, the higher the importance. The value of \( w \) can be set dynamically. After a large number of experiments, we can finally find a value of \( w \), so that the performance of our recommendation system can reach the best in Mae index. The corresponding weight is the value we ultimately need as the similarity measurement parameter. Using the above formula (9), we can get the similarity matrix between the target user and other users as follows.

\[
\text{Similarity} = \begin{bmatrix}
\text{Similarity}_{11} & \text{Similarity}_{12} & \ldots & \text{Similarity}_{1n} \\
\text{Similarity}_{21} & \text{Similarity}_{22} & \ldots & \text{Similarity}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\text{Similarity}_{n1} & \text{Similarity}_{n2} & \ldots & \text{Similarity}_{nn}
\end{bmatrix}
\]

(10)

(4) When recommending to the target users, it is mainly to select \( k \) neighboring users with high matching degree as their neighbor users according to the similarity matrix \( \text{Similarity} \), which is customarily called \( k \)-neighbors, that is, the user's neighbor set. Then formula (10) can be used to predict the unknown score of the target user. The score value is \( \text{pre}_{ij} \) :
3.2 Convolutional Neural Network

3.2.1 Related Definitions

Definition 1: Singular value decomposition theorem: let \( A \in \mathbb{R}^{m \times n} (r > 0) \), then there are \( m \)-order orthogonal matrix \( U \) and \( n \)-order orthogonal matrix \( V \), so that \( U^T AV = \begin{bmatrix} \sum_{i=1}^{r} \sigma_i U_i V_i^T & 0 \\ 0 & 0 \end{bmatrix} \), where \( \sum_{i=1}^{r} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r) \), \( \sigma_i (i=1,2,\ldots,r) \) are all non-zero singular values of matrix \( A \), satisfying \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0 \), the first several values are relatively large, they contain most of the information of matrix \( A \). The column vector of \( U \) (left singular vector) is the eigenvector of \( A^T A \), and the column vector of \( V \) (right singular vector) is the eigenvector of \( A A^T \).

3.2.2 Singular value recommendation algorithm

For a matrix \( M \) whose dimension is \( m \times n \) (\( m > n \)), it can be decomposed into three matrices after SVD Optimization: matrix \( U \), matrix \( \sum \) and matrix \( V \).

\[ M = U \sum V^T \]  
(11)

The \( U \) in formula (11) is an orthogonal matrix, its dimension is \( m \times m \), and the definition of the orthogonal matrix meets the following requirements: \( U \times U^T = 1 \); the matrix \( V \) is also an orthogonal matrix, and its dimension is \( n \times n \); the matrix \( \sum \) is the final singular value diagonal matrix after decomposition, that is to say, only the elements in the matrix on the diagonal can not be 0, and the elements in other positions are 0. The diagonal elements must also meet certain conditions: \( \sum_{i=0}^{n} \delta_i \) is greater than 0, and the eigenvalues \( \delta_i \) on the diagonal are arranged in descending order from large to small, that is, \( \delta_1 \geq \delta_2 \geq \ldots \geq \delta_n > 0 \).

Algorithm: SVD matrix decomposition algorithm

Input: rating matrix \( R \)
Steps:
Step 1: replace the unknown value in matrix \( R \) with the average value \( r_i \) of all the values in the column;
Step 2: replace the element \( r_{ij} \) of matrix \( M \) with \( r_{ij} = r_i - r_{ij} \) to get matrix \( R \); decompose the singular value of \( R \) to get three matrices \( U \), \( \sum \) and \( V \);
Step 3: simplify \( \sum \), set threshold filter for the elements on its diagonal, replace all the values lower than the threshold with 0, for example, keep the first \( K \) eigenvalues, and then the values after the \( K \) eigenvalues are all 0, then delete the row and column corresponding to 0,

and finally get a \( K \)-dimensional matrix \( \sum_k \), which is still a diagonal matrix;

Step 4: use matrix \( \sum_k \) in step 3 to simplify \( U \), \( V \), get \( U \) and \( V \), then matrix \( R \) can be simplified to \( R = U \sum_k V^T \).

Output: rating matrix \( R' \).

The recommendation based on SVD optimization can well preprocess the scoring data through dimension reduction technology, alleviate the sparsity of the data, and make the system can find more hidden main feature information while processing[19]. Not only for users, users' subject interests and hobbies can be represented by vectors of keywords of different quantity dimensions; similarly, the items to be recommended in the system can also be represented by vectors of different quantity of tag dimensions (tag attribute)[20]. Finally, the recommendation of the target user can match the user's image feature vector with the item tag vector to obtain the similarity calculation.

3.3 Detailed steps of accurate recommendation algorithm based on TAG_SVD_CF

This paper designs a hybrid recommendation algorithm based on SVD optimization, which combines collaborative filtering recommendation with tag attribute information.
Input: user rating matrix $R$ and item tag attribute matrix $T$

Steps:

Step 1: in this paper, SVD technology is used to optimize the sparse problem of the original scoring matrix $R$, and then fill in the missing value: according to the existing literature [21], for the data on the training set, The prediction $P_u$ of the target user $u$ to the unknown score represents the prediction value of $u$ to the item $i$. $P_u$ can be calculated by the following formula:

$$P_u = \hat{R}_u = U_s \times \sqrt{S}^T \times V^T (k)$$  \hspace{1cm} (12)

In the above formula (12), $\hat{R}_u$ is the average score of the target user $u$ on all items that have been scored can be obtained by simple calculation. Matrix $U_s$, matrix $S$ and matrix $V$ are obtained by SVD decomposition technology. After processing in formula (11), three matrices $U_s$, $S$, $V$ with dimension $K$ can be obtained, that is, the dimension parameters reserved after processing. The selection of $K$ is mainly through the following experimental part, and then a complete scoring matrix $R$ without missing value can be obtained.

Step 2: for the matrix obtained in 1) above, calculate $\text{Similarity}_{sim} (u,v)$ based on the traditional Pearson correlation similarity, and finally get a matrix of similarity between users.

Step 3: when calculating the predicted value, it is necessary to calculate the preference of the tag attribute of the item. According to the formula (8), (9) and (10), the user rating matrix and item tag attribute matrix $T$ are calculated, and the values of Pearson similarity $\text{Similarity}_{sim} (u,v)$, attribute preference similarity $\text{Similarity}_{w} (u,v)$ and final comprehensive similarity $\text{Similarity} (u,v)$ are obtained. Finally, the predicted value is obtained according to formula (10).

Step 4: recommend according to the predicted value of step 3.

Output: recommended result sets

The accurate recommendation algorithm based on Tag _SVD_CF is applied to the recommendation of medical information. The recommendation system studied in this paper is mainly through the research of user portrait, and the design of personalized recommendation algorithm for accurate recommendation of medical information. The core recommendation algorithm is to use SVD technology to optimize the matrix of the user disease score data obtained and processed, and to solve the data sparse problem caused by the absence of some score values in the matrix. Then, we introduce the similarity of user profile's label attribute while calculating the user similarity. We use $w$ as the coordination coefficient to get the hybrid similarity based on $\text{Similarity}_{user, profile}$ and $\text{Similarity}_{user, disease}$, and then implement the recommendation algorithm based on the hybrid collaborative filtering algorithm. The recommended algorithm is as follows: The recommended algorithm is as follows:

Algorithm input:
User image label matrix $PM$, user disease score information matrix $IM$.

Algorithm output: recommended disease medical record list $Top-N$.

Algorithm steps:
Step 1: get all user profile data according to the acquired user data, disease knowledge base and data acquired after processing. Extract the top 10 labels of each user's portrait label weight according to the needs, and calculate the user's portrait label matrix $PM$.

Step 2: the user's disease score matrix $IM$ is mainly provided by the user's medical history. The corresponding disease situation corresponds to the user's score $P_u$ for this disease, which is a five point system$^{[22]}$:

$$P_u = \left[ \frac{\text{Illness}_i}{\text{Illness}_{max}} \right]$$  \hspace{1cm} (13)

Where $\text{Illness}_i$ represents the number or active score of the user's medical history for the $i$th disease. $\text{Illness}_{max}$ represents the maximum value in $\text{Illness}_i$. Square brackets represent rounding down.

Step 3: SVD singular value decomposition technology is used to optimize the user rating matrix to solve the problem of sparsity of the user rating matrix$^{[23]}$. Then, by calculating the tag attribute preference matrix of $PM$, the tag information of user profile is applied here, and the attribute tag of disease is not used, so our tag attribute preference can be directly replaced by tag weight$^{[24]}$. Then we calculate the similarity between users and the traditional user similarity of $IM$. By combining a weight $w$, we get the $\text{Similarity}_{w} (u,v)$, $\text{Similarity}_{f} (u,v)$ and the total similarity $\text{Similarity} (u,v)$, which are respectively replaced by formula (8), and formula (9), where $\text{Similarity}_{w} (u,v)$ is used to replace $\text{Similarity}_{f} (u,v)$.
Step 4: use the Similarity(u,v) in step 3 as the final similarity to predict the unknown scoring data r of the target user.

Step 5: fill in the initial scoring matrix IM of the target user according to the unknown scoring data Pre. Since the patient's past disease may still suffer from this disease in the future, the patient's condition recommendation should integrate the initial scoring data[25]. Finally, it sorts and outputs the relevant information back to the user.

3.4 Algorithm complexity analysis

3.4.1 Time complexity analysis

Input: user disease rating matrix R_{ux}, disease tag weight similarity matrix L_{sim_{ux}}, disease rating similarity matrix S_{sim_{ux}}, in which n is the number of diseases, m is the number of users, and α is the sparsity of rating matrix.

Output: k forecast scoring results of users.

- First, the disease neighbor selection of the target disease, ranking the score similarity matrix and disease label matrix, with a time complexity of O(2log(α*n)). SVD algorithm is used to select r disease neighbors, so the time complexity of calculating the score difference between disease and target disease is O(r + 2log(α*n)). To calculate the user score difference between any two diseases, each user's score data needs to be scanned. The time complexity is O(m), so the time complexity of each prediction score is O(m(t + 2log(α*n))), and the time complexity of k records is O(km(t + 2log(α*n))).

3.4.2 Spatial complexity analysis

The spatial complexity determines the number of parameters of the model. Due to the limitation of dimension disaster, the more parameters the model has, the more data it needs to train the model. However, the data set in real life is usually not too large, which makes the training of the model easier to over fit[26].

When using TAG_SVD_CF for information recommendation, SVD decomposition is carried out according to the score matrix after collaborative filtering based on user score. It can greatly reduce the size of feature vector, discard redundant features, and finally get a matrix of similarity between users. In this way, the number of parameters obtained is less than that of traditional CF (n,m is the dimension of input/output), which effectively solves the problem of data sparsity. It can be seen from this that the algorithm of TAG_SVD_CF has a lower spatial complexity than that of single CF.

4 Experimental Tests

4.1 Experimental data

The experimental data are from 31 days' IIS log records of the intelligent system of psychological counseling from May 13, 2019 to June 12, 2019. The total number of log records is 16171, removing anonymous users from IIS logs. At the same time, by removing the invalid link of the status code, 1856 users in this data set were randomly selected by the system, and 4128 kinds of target consultation were provided by these users, and more than 200000 scoring records were counted. The attributes of psychological counseling items in the data set are divided into many kinds, each disease has one or more attributes, and also provides the tag information tags of users for some diseases. The following is a simple calculation of the sparsity of the data set:

$$\frac{1 - \frac{200000}{1856 \times 4128}}{1} = 0.9738957$$

From the above data, we can see that the data set is relatively sparse. For this data, 75% of the training data used in the algorithm model and 25% of the test data are selected by random generation. The most common MAE (mean absolute error) is still used to evaluate the experimental results.

$$MAE = \frac{\sum_{i \in T} |p_i - q_i|}{N}$$

Among them, p_i is the prediction score, that is, the score of unknown score prediction generated by the system according to the designed core algorithm. q_i is the actual score of the user to the movie in the test set. The smaller the MAE is, the closer the predicted score is to the actual score.

4.2 Analysis of experimental results

(1) Comparison of similarity measurement methods

The similarity between Pearson similarity and user's preference for tag is measured by value. The proposed algorithm based on the similarity of label attribute preference and cosine similarity measure and the traditional Pearson correlation coefficient are compared, and then the experimental results are observed and analyzed. In this paper, the number of k-neighbors of users is gradually increased from 4 to 60, the step size is 4, W value is set to the reasonable 0.6 of the last experiment, and then the experiment is carried out. The experimental results are shown in Figure 3.
Figure 3 The comparison of MAE between this algorithm and traditional recommendation method

It can be seen from the figure that no matter how to select k-neighbors (number of nearest neighbor users), the hybrid recommendation algorithm proposed in this paper reduces the value of MAE to a large extent and improves the accuracy of recommendation. The result of the recommendation algorithm using Pearson correlation coefficient similarity and cosine similarity is poorer than that of the hybrid recommendation algorithm. At the same time, it can be also seen that the best recommendation effect of the hybrid recommendation algorithm can be achieved when the number of nearest neighbors is 40. It can be clearly seen from the figure that when k value is less than 30, the MAE value of the algorithm drops sharply as the K increases, and the change slows down when it reaches the vicinity of 30. When the number of user neighbors K reaches 40, the calculated MAE value is the minimum, so the optimal number of user neighbors is 40 in the paper due to the computed cost.

(2) Selection of K value in SVD optimization

Dimension K is the key of the next experiment when Figure 4 that the MAE value firstly decreases and then increases as the number of user neighbors K increases. Because the selection of K value is too small (such as K=1), which means that users have less preferences, and the main attribute information of the project is also relatively small, which means that most attribute information is lost after decomposition. However, if the K selection is too large, it will lead to the significance of SVD technology optimization(such as K=30). Therefore, selecting the appropriate K value will improve the performance of the recommendation system. It is necessary to select reasonable values through experiments.

As shown in Figure 4 and Figure 5, different retention dimensions K have a great impact on the Mae and precision of the recommendation system. However, as a whole, it can be determined that when K=13, MAE in Figure 4 reaches the lowest point.

Figure 4 Mae value optimized by SVD for K value

(3) comparison of different recommendation algorithms

The algorithm designed in this paper is evaluated as a whole, and the algorithm proposed in this paper is compared with the algorithm proposed in COS-CF, ACOS-CF, Pearson CF, US-CF (Ding Shaoheng et al 2015) based on modified cosine similarity.

Figure 5 Comparison of recommendation quality from different algorithms

The experimental results are shown in Figure 5, which shows the quantitative relationship between the MAE values of various algorithms under different K values. From Figure 5, the MAE value of the improved algorithm in this paper is significantly lower than the results of the traditional algorithms and the results of the improved algorithm are relatively stable. When the number of nearest neighbors K is less than 30, the MAE values of various algorithms are higher, but the results of the improved algorithm are lower than the others. With the increase of the number of nearest neighbors, the MAE values of various algorithms are sharply reduced. When the number of nearest neighbors is 5, the MAE value of the improved algorithm is smaller and the change range is smaller than other algorithms. When the K
value is more than 30, all kinds of algorithms tend to be stable gradually. The MEA value obtained by this algorithm is significantly lower than other algorithms, which shows that the algorithm proposed in this paper is superior to other algorithms in terms of recommendation accuracy and can significantly improve the recommendation quality of the recommendation system.

5 Conclusion

A hybrid recommendation algorithm of collaborative filtering and item tag attribute based on SVD technology is proposed for Psychological Consultation Information in the paper. The user item rating matrix optimized by SVD technology is used to solve the missing data cores of traditional collaborative filtering algorithm. The index system of developing recommendation algorithm was established from demographic attribute, interest label dimension and business social dimension. In the aspect of user portrait modeling, the tag attributes of user portrait and the calculation of each tag weight are generated for the developing recommendation algorithm, and the updating of user portrait is also analyzed. Compared with the traditional cosine similarity and Pearson similarity, the proposed similarity has a lower cosine similarity and Pearson similarity, the proposed algorithm is poor in time and space complexity because of the combine item tag attribute algorithm with collaborative filtering algorithm, due to the large number of improved fusion models, and there is still a large room for improvement for the future work.

Compliance with Ethical Standards:

Funding: This research work is supported by Intelligent Psychological Consultation System Based on NLP and User Portraits (201910060022), College Students' Innovative Entrepreneurial Training Plan Program (National Program, 2019.6-2020.6), National Natural Science Foundation of China (Grant No. 61571328), Tianjin Key Natural Science Foundation (No.13JCZDJC34600), CSC Foundation (No. 201308120010), Major projects of science and technology in Tianjin (No.15ZXDSGX 00050), Training plan of Tianjin University Innovation Team (No.TD12-5016, No.TD13- 5025), Major projects of science and technology for their services in Tianjin (No.16ZXFWGX00010, No.17YFZC GX00360), the Key Subject Foundation of Tianjin (15JCYB JC46500), Training plan of Tianjin 131 Innovation Talent Team (No.TD2015-23).

Informed consent: Informed consent was obtained from all individual participants included in the study.

Contribution: De-gan Zhang, Zhi-xian Liu, Tian-zheng Fu jointly designed the algorithm and experiment, and wrote the core content of the algorithm; Xin Zhang, Kai-yan Wen, and Bing Liu jointly discussed the application direction of user portrait in psychological counseling information recommendation, tested the algorithm, revised and corrected the grammar and sentences of this thesis.

Reference

[1] Wu Tao, Gao Changchun, Lin Jingsheng. Application of information intelligence technology in hospital operation decision analysis [J]. Chinese medical journal, 2017, 97 (12): 894-896.
[2] Li Chunhun, Li Jun. Collaborative filtering recommendation algorithm based on ASVD [J]. Minicomputer system, 2018, 39 (6): 168-172.
[3] Zhang Chao, Qin Yongbin, Huang ruizhang. Collaborative filtering algorithm combining confidence and SVD [J]. Computer and digital engineering, 2015, 43 (5): 758-761.
[4] Wang Jianfang, Zhang Pengfei, Liu Yongli. Research based on improved decomposition algorithm of probability matrix with bias [J]. Computer application research, 2017, 34 (5): 1397-1400.
[5] Wang Guomin, Liu Xin, Zhu Rong, et al. A new hybrid personalized recommendation algorithm [J]. Computer and modernizaton, 2013 (8): 64-67.
[6] Fei Peng, Lin Hongfei, Yang Liang, et al. A multi perspective fusion framework for building user portraits [J]. Computer science, 2018, 45 (1): 179-182.
[7] Wang qiangbing, Zhang Chengzhi. Design and implementation of user portrait construction system integrating content and user gesture behavior [J]. Data analysis and knowledge discovery, 2017, 1 (2): 84-90.
[8] Zhang Wei, Mo Meiqi, Xia Zhiping, et al. Design and implementation of personalized service recommendation system for medical information resources [J]. Library Journal, 2016, 25 (6): 43-45.
[9] Wen Yuxin, Wang Huaqiong, Tian Lili, et al. Research on semantic based regional medical information integration method [J]. China Digital Medicine, 2013 (9): 28-30.
[10] Chen Gong, Huang ruizhang, et al. Multi dimensional text representation based on social characteristics [J]. Computer engineering and science, 2016, 38 (11):
Hu Jiming, Hu Changping, Hu Jiming, et al. User modeling based on topic hierarchy tree and semantic vector space model [J]. Journal of information science, 2013, 32 (8): 838-843.

Yang Jing, Cheng Weiqing, Guo Changzhong, et al. Research on user interest model based on standard label [J]. Computer technology and development, 2013 (10): 208-211.

Sallam A A I. An intelligent system to boost the recommendation for individual and group users in mobile marketplace[J]. International Journal of Computational Systems Engineering, 2016, 1(1):33-41.

Poompuang P. Compromise between quality and effort of recommender systems using the neighbor filtering approach based on reliable neighbors[C]/ International Conference on Science & Technology. 2016.

Gong Xiaocui, an Xinxin, Gong Xiaogui, et al. Detection of medical subject splitting and fusion based on LDA model [J]. Library and information work, 2017, 61 (18): 76-83.

Eltaher M A. Social user mining: User profiling of social media network based on multimedia data mining[J]. Dissertations & Theses - Gradworks, 2015.

Bokde D K, Girase S, Mukhopadhyay D. An Approach to a University Recommendation by Multi-criteria Collaborative Filtering and Dimensionality Reduction Techniques[C]/ IEEE International Symposium on Nanoelectronic & Information Systems. 2015.

Guo X, Yin S C, Zhang Y W, et al. Cold Start Recommendation Based on Attribute-Fused Singular Value Decomposition[J]. IEEE Access, 2019, 7:1-1.

Mansouri A, Aznaveh A M, Torkamani-Azar F, et al. Image quality assessment using the singular value decomposition theorem[J]. Optical Review, 2019, 16(2):49-53.

Bunse-Gerstner A, Byers R, Mehrmann V, et al. Numerical computation of an analytic singular value decomposition of a matrix valued function[J]. Numerische Mathematik, 2017, 60(1):1-39.

Li Yongchao, Luo Jun, Li Yong Chao, et al. Collaborative filtering algorithm based on partial characteristics of users [J]. Computer system application, 2017, 26 (3): 206-210.

Han Yanan, Cao Han, Liu Liangliang, et al. Collaborative filtering recommendation algorithm based on scoring matrix filling and user interest [J]. Computer Engineering, 2016, 42 (1): 36-40.

Wang Na, Ge Yubin. Research on user fuzzy clustering method based on tag weight [J]. Information theory and practice, 2016, 39 (3): 140-144.

Yi HW, Zhang Fu Zhi. Robust recommendation algorithm combining k-distance and project category information [J]. Minicomputer system, 2017, 38 (11): 2476-2481.

Wang Yongkang, Yuan Weihua, Zhang Zhijun. A recommendation algorithm for the fusion of temporal argot filling and subgroup Division [J]. Computer engineering and application, 2019 (16): 130-137.

Zheng Dan, Wang Mingyang, Chen Guangsheng, et al. Research on user clustering recommendation algorithm based on weighted slope one [J]. Computer technology and development, 2016, 26 (4): 51-55.