Microprocessors and Microsystems 81 (2021) 103728

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Intelligent information recommendation algorithm under background of big data land cultivation

Haoxiang Tang [a], Wei Yang [b], Susheng Zheng [c,*]

[a] Carey Business School, Johns Hopkins University, Washington, D.C., 21218, America
[b] The Center of Collaboration and Innovation, Jiangxi University of Technology, Nanchang, Jiangxi, 330098, China
[c] Office of general services, Jiangxi University of Technology, Nanchang, Jiangxi, 330098, China

ARTICLE INFO

Keywords:
Big data
Intelligent informatization
Upper ecosystem
Explicit preference
Implicit preference

ABSTRACT

In order to solve the problem of serious information overload in the era of big data and improve the informatization of intelligent recommendation result, an intelligent information recommendation algorithm based on user preference mining was put forward. According to the background of big data, user behavior data is unified. The advantage of spark relative to compared Hadoop Map Reduce was analyzed through the operating architecture and upper ecosystem of spark, so that the data processing ability was improved. The user preference mining technology was integrated with the intelligent information recommendation algorithm. Moreover, the explicit user preference knowledge and implicit user preference knowledge were analyzed to obtain the user preference knowledge and the nearest neighbor community, and thus to complete the intelligent information recommendation. Experimental results show that the proposed algorithm can exactly reflect the user preferences in different groups, with good recommendation accuracy. In addition, the desired effect is achieved.

1. Introduction

In the rapid development of the Internet, massive data information has been produced. It is significant to mine valuable information from these data. In this context, the research and development of big data platform began to rise. For producers and consumers, the ability and speed of producing and transmitting information and data are growing exponentially. On the one hand, it provides extensive information resources. On the other hand, "information overload" makes it difficult to accurately find the required information. It is more difficult for users to find useful data from complex and diverse information. With the increase of data, how to find the required information has become an urgent problem. Therefore, the recommendation system has become the main mechanism to solve information overload. The search engine is different from the search engine technology. It requires user to search keywords and filter out information. However, users are often unable to accurately describe and locate the keywords that need to be queried, and the mechanism of search engines can't provide the automatic recommendation service according to user behavior. The recommendation system needs to build an interest preference model by analyzing the log data and the historical user behavior, and predicts the user's preference for information by the recommendation algorithm. And then, the recommendation system sorts the products or information that users may be interested in. Based on the advantages of personalization and intelligence, it provides the development power and commercial interests for the traditional Internet industry. At the same time, it also promotes more in-depth and extensive research in the field of recommendation.

In order to get more accurate recommendation, the weighted interest recommendation algorithm based on association rule analyzes the user data and establishes the association rule mining algorithm. By analyzing the correlation between user interests, the algorithm predicts the user interest score, and recommends the personalized services for users [1]. Meanwhile, a real-time community search model is built by incremental training algorithm. For dynamic weighted attributes of each node, the model can expand the attribute set when a node and a group of attributes are given, so that the matching between the two nodes is more significant [2]. Based on the improved tag recommendation quality method, we can find that the total number of tags used by users in a social tagging system changes with time. The concept of user marked state is introduced, including growth state, mature state and dormancy state. The algorithm to determine the status of user tags is adopted. Based on the statistical language model, three strategies are used to calculate the probability distribution of tags and thus to recommend the tags that are

* Corresponding author.
E-mail address: dshssdf6484@163.com (S. Zheng).

https://doi.org/10.1016/j.micpro.2020.103728
Received 7 November 2020; Received in revised form 14 December 2020; Accepted 18 December 2020
Available online 28 December 2020

most likely to be used by users [3].

The above methods query the corresponding content according to the input content mechanically, which can't meet the dynamic needs and mine the potential needs of users. Therefore, an intelligent information recommendation algorithm under the background of big data is presented. Integrating the user preference mining technology into intelligent information recommendation algorithm can help users find the required information as soon as possible. In addition, it can help technology suppliers implement active marketing, improve the success rate of technology and promote technological innovation.

2. Analysis of big data background

With the rapid development of Internet of things, cloud computing, the Internet and other new technologies, massive information and data is far beyond the accepting ability of users. In the face of massive information based on data, people can't accurately obtain the required information. From the perspective of information consumers, massive information has flooded information consumers. From the perspective of information producers, the exponential growth of data makes information producers unable to mine users' real interests and make an accurate recommendation according to the current interests of users, resulting in the loss of customer resource.

Big data has brought great influence and great value to social, economic and scientific research. In this context, the research on big data and its related fields has become a key issue [4]. As an effective method to solve the problem of information overload, the intelligent information system has been widely applied in e-commerce, music, film and television, social software, read, advertisement and other fields, with good economic benefits. Therefore, it has great theoretical significance and economic value to study the personalized recommendation in big data environment.

Theoretical significance: based on user behavior data, the theories of information entropy, complex network and machine learning are applied to the recommendation algorithm, and then the influence of explosive growth and severe sparsity of user historical behavior data on the recommendation algorithm in the big data environment are discussed. How to combine with above theories to process the recommendation algorithm is also discussed.

Immediate significance: the research on recommendation system not only helps users find the interested or valuable information, but also presents the information accurately to the users who are interested in this information, and thus to realize a win-win situation between information consumers and information producers. For the rapid development of e-commerce, the excellent personalized recommendation system can build a service system to meet the personalized needs. This not only helps e-commerce enterprises to make full use of the existing network resource, but also stabilizes the existing customer groups, develops new customer sources, improves customer loyalty and service level, and thus to strengthen their competitiveness. The core content of recommendation system is the recommendation algorithm, so it is significant to improve the efficiency of intelligent information recommendation algorithm and strengthen the function of recommendation system.

2.1. User behavior data

User behavior data mostly exists in the website in the form of logs. User behaviors (such as browsing, clicking, purchasing, scoring, etc.) are recorded and saved in the log, and many original logs are stored in the website background [5]. In the recommendation system, there are two kinds of feedback: explicit feedback and implicit feedback.

The explicit feedback is the behavior that users express their preference for the project actively and clearly. The common way to collect clear feedback from users includes rating and like/dislike. According to the needs of different websites, the scoring system adopts ten-point scale (Douban film review, IMDB, etc.) and five-point scale (Taobao, Amazon, movielens, etc.). The two scoring systems have their own advantages and disadvantages. According to actual needs, different websites design their scoring system.

Table 1
Unified Representation of User Behavior.

User ID	Unique identification of the user who generated the behavior
Item ID	Unique identification of the object that generated the behavior
Behavior Type	Types of behavior (e.g., buying, browsing, etc.)
Context	Context (including time, place, etc.) in which the behavior occurs
Behavior Weight	The weight of behavior (for example, the weight of video browsing behavior can be determined according to the viewing time, and the weight of scoring behavior is determined according to the score value)
Behavior Content	Behavior content (e.g. comment text of comment behavior, label of tagging behavior, etc.)

The behavior of browsing web and clicking page link may be caused by user's unintentional behavior or wrong operation, which can't reflect the user's preference. Compared with explicit feedback, the implicit feedback is not clear, and only has positive feedback. Due to the massive data, it is easy to collect.

There are many kinds of user behaviors on the Internet (such as browsing the web, shopping, commenting on movies and videos, evaluating movies and books, etc.), so it is impossible to express all user behaviors in a unified way. In general, the user behavior is divided into a generic user behavior in Table 1.

In practical application, the website should give different representations to different user behaviors according to the actual needs, instead of using a unified structure to represent all user behaviors. The unified representation of user behavior is to make it easier to understand the user behavior data.

2.2. Spark distributed computing framework

2.2.1. Research on spark operation architecture

Spark is a parallel computing framework that is similar to Hadoop Map Reduce. It is a cluster computing system based on memory. It is also the most active project of Apache. The Spark can provide a more general platform for big data processing [6]. Compared with Hadoop, Spark can increase the speed of program running on disk by ten times and the speed of program running in memory by one hundred times. Spark has become the fastest open-source engine for PB data sorting, and provides the following key features:

(1) It can run on a separate cluster or a cluster managed by Hadoop YARN or Apache Mesos.
(2) Currently, API for Java, Python, and Scala are given.
(3) It can be well integrated with Hadoop ecosystem and data source.

Spark introduces Resilient Distributed Dataset (RDD), which is created when loading external datasets or distributing the sets from driver applications. It can contain any type of objects. It is a set of immutable, fault-tolerant and distributed object.

The operation types of RDD include transformation and actions [7]. The transformation is to build a new RDD according to the original RDD. Actions are to return the results to the driver after the RDD operation. All RDD transformations adopt the delay loading mode. That is to say, Spark does not immediately calculate the results, but remembers the transformation between all datasets. Only when these transformations encounter the action, the calculation begins. This design makes Spark more efficient.

The operational architecture of Spark Application is composed of driver program (SparkContext) and executor [8]. In general, Spark Application operates in Spark Standalone, YARN and mesos clusters, which provide computing resource for Spark Application. Spark

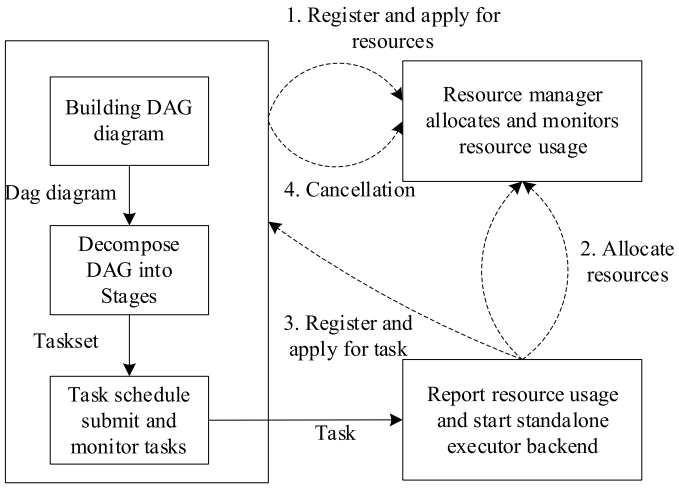

Fig. 1. Spark Operation Architecture.

Application consists of a driver program and multiple jobs. One job is composed of multiple stages, and one stage is composed of multiple tasks without shuffle relationship. Spark application calculates various transformations, and triggers the job through action. After submission, DAG diagram is constructed according to RDD dependency through Spark Context, and DAG Scheduler parses DAG diagram. When parsing, the shuffle is taken as the boundary, which belongs to the reverse parsing. In order to build dependencies in stage, the TaskSet needs to be submitted to the underlying scheduler, and then submitted to Task Scheduler in spark, thus generating Task Setmanager. After that, it is submitted to the executor for calculation. After the calculation is completed, it is fed back to Task Setmanager at first and then to the Task Scheduler, and finally to DAG scheduler. After all the operations are completed, the data is written.

The basic process of *Spark* operation is shown in Fig. 1.

Step 1: start spark context and register with the resource manager (Standalone, Mesos, or YARN) and request Executor resource.

Step 2: after receiving the application, the resource manager allocates the Executor resource, and starts the Standalone Executor Backend.

Step 3: Spark Context builds DAG diagram, DAG Scheduler decomposes DAG diagram into Stage, then sends Taskset to Task Scheduler. And then, Executor registers with Spark Context and applies for Task, and Task Scheduler sends Task to Executor.

Step 4: run the Task on the Executor and release all resources after it has completed running.

The characteristics of Spark operational architecture: each application has a dedicated executor process, which resides in the application and executes tasks in a multithreaded mode. From the perspective of the task scheduling of each driver, this Application isolation mechanism has its advantages. Spark is unrelated with resource manager, so it only needs to get the executor and maintain communication with each other. Due to the large amount of information exchange between Spark Context and Executor during the running of Spark Application, the submission for Client of Spark Context should be close to the Executor node. In the best case, they're in the same rack. Task not only uses the speculative execution optimization mechanism, but also the data locality optimization mechanism [9].

2.2.2. *Spark upper ecosystem*

The core of Spark is composed of Spark SQL, MLlib, Spark streaming and Graph X libraries, which can be applied to the same application program seamlessly [10]. Next, these libraries are described.

(1) SparkSQL supports SQL or Hive query data. It is a component of Spark. It was originally derived from Apache Hive project, and integrated into Spark to provide support for multiple data sources. It is destined to a very powerful query tool.
(2) Spark Streaming can process streaming data in real time. Spark Streaming can receive input data in message queues such as Twitter and Kafka, and divide it into several batches, and then process it with spark. Finally, the results in batch are generated into stream.
(3) MLlib is an algorithm component based on Spark, which provides classification, regression, clustering, collaborative filtering and other algorithms, some algorithms can also be used on stream data, such as K-means clustering and ordinary least squares.
(4) Graph X is a very important component in Spark, which is used to calculate graphs and parallelize graphs. It introduces a new abstract graph, namely directed multigraph with vertex and edge attributes, which extends Spark RDD. In order to support the graph computing, Graph X provides many basic operations such as subgraph and join Vrtices.

2.2.3. *Machine learning framework–mahout*

Mahout is an open-source project that can help developer create intelligent application more easily [11]. Mahout has three core topics such as clustering, classification and recommendation engine and other algorithms [12]. Mahout has completed many classic data mining algorithms, and these algorithm modules are extensible. It can parallelize some algorithms, so that massive data can be quickly processed on Hadoop distributed platform. The collaborative filtering library form Taste project provides the main interface for common collaborative filtering algorithms. The Mahout theme is shown in Fig. 2.

Mahout parallelizes some machine learning algorithms into Hadoop Map Reduce mode, and thus to improve the data processing ability of algorithms.

3. Intelligent information recommendation algorithm based on user preference mining

As an important way to solve the problem of information overload, the recommendation system has become a hot spot in the field of social network. The intelligent information recommendation algorithm based on user preference mining proposed in this paper can predict the demand of target users through the purchase behavior and preference knowledge of the nearest neighbor community, and thus to realize the research on intelligent information recommendation [13].

For the same user, if the prediction items are different, the adjacent users for prediction are also different. The adjacent users are related to the items which need to be predicted. In this way, we can ensure that the adjacent users for prediction and the current users have similar interests in the items to be predicted. This is the key to ensure the accuracy of algorithm.

3.1. *User preference mining*

The recommendation system should make personalized recommendation to technology supply and demand. In addition to processing the Web information such as technology supply and demand information published on the technology innovation platform, the personalized information of supply and demand sides should also be effectively processed.

Before the personalized recommendation for users, the information recommendation system should comprehensively analyze according to the personalized information of technology suppliers and demanders, so as to ensure the accuracy of recommendation [14]. The user preference mining is to use data mining, behavior analysis, trend prediction and other technologies to deeply mine and analyze user preference information and thus to obtain user preference knowledge. Traditionally, the

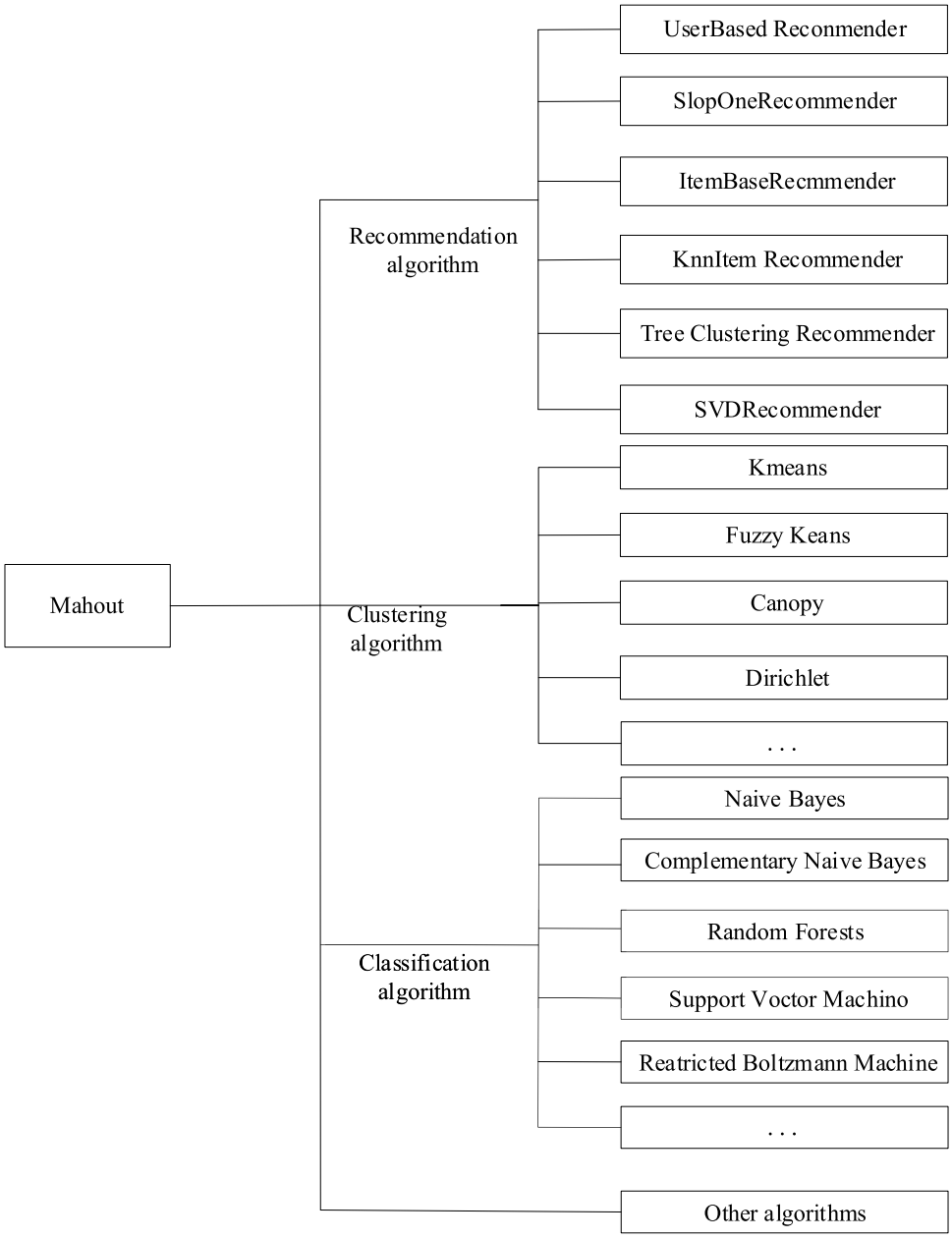

Fig. 2. Three Topics of Mahout and Other Algorithms.

user preference mining mainly obtains explicit preference knowledge by analyzing and processing explicit information such as user registration information, user rating, user comments [15]. However, it is not enough to analyze and process the implicit information such as the residence time on the web page and the number of clicks.

The proposed intelligent information recommendation algorithm comprehensively uses the data mining technology to obtain the user's explicit preference information and the user's explicit preference knowledge. In addition, the algorithm uses the behavior analysis and Web log mining technology to mine the user's implicit preference information, and thus to obtains the user's implicit preference knowledge. The trend prediction and association mining technology are used to analyze the real-time changes of users' browsing behavior and mine the trend information such as evaluation items, and thus to more reasonably predict the user preference. The processing flow is shown in Fig. 3.

(1) Explicit preference knowledge mining: the main object of explicit preference knowledge mining is the explicit preference information such as text information and comment information of web page [16]. In this article, K-means clustering algorithm is used to mine and analyze the user's explicit preference information, and thus to obtain the user's preference knowledge clustering. $D = \{d_1, d_2, \ldots d_n\}$ is a set of text documents formed by user explicit preference information collected after preprocessing the data. Firstly, the initial clustering $C = \{c_1, c_2, \ldots, c_i, \ldots c_n\} c_i = \{d_i\}$ is constructed, and all the documents in $D$are regarded as a single initial user preference category. Secondly, the similarity between any category$sim(c_i, c_j)$ is calculated, and then the initial clustering is merged and optimized according to the set threshold ε. The most similar category with maximum similarity is selected.

$$\max = Max\{sim(c_i, c_j)\} \tag{1}$$

In Formula (1), if $\max > \varepsilon$, the sub clusterc_i is regarded as the seed set

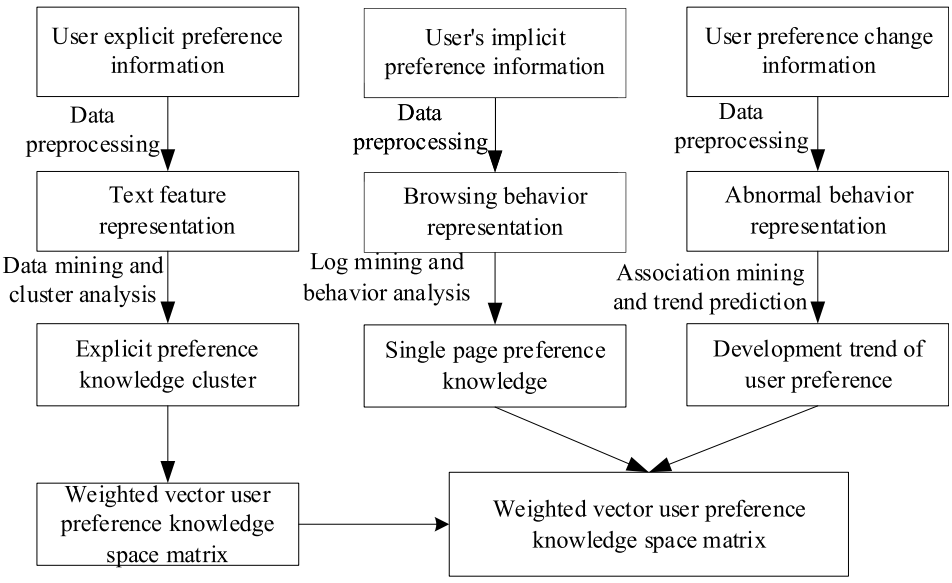

Fig. 3. Flow Chart of User Preference Mining.

of the initial clustering center of K-means algorithm, $S = \{s_1, s_2, \ldots s_k\}$. $sim(d_i, s_i)$ is calculated. The seeds with the greatest similarity are divided into the class with maximum similarity, $\max = sim(d_i, s_i)$. Finally, the clustering result$C^* = \{c_1, c_2, \ldots, c_k\}$ are obtained.

(1) Implicit preference knowledge mining: many behaviors can reflect user's preference when browsing e-commerce websites, such as residence time on relevant web pages, the web browsing times and the number of clicks [17]. In this article, we use the similarity within cluster and the similarity between clusters to calculate the comprehensive similarity of user's implicit preference. The sequential aggregation algorithm based on K-means central aggregation is used to mine the user preference knowledge of user groups with the same implicit preference and single page. According to the data preprocessing method, the collected user privacy preference information is preprocessed. The initial number of clusters of user behavior sequence set $D = \{d_1, d_2, \ldots d_n\}$ is k, and then the similarity within the behavior cluster is calculated.

$$S_w(k_i) = \frac{\sum_{i=1}^{n-1} Sim(d_i, d_j)}{c_n^2} \tag{2}$$

Where, k_i is the ith cluster in the clustering mode when the number of clusters is k. $S_w(k_i)$ represents the average similarity within the ith cluster.

The calculation for similarity between clusters is similar to intra-cluster similarity.

$$S_b(k_i, k_j) = \frac{\sum_{s=1}^{k_i} Sim(d_i, d_j)}{d_i, d_j} \tag{3}$$

Where, $S_b(k_i, k_j)$ represents the inter-group similarity between behavior sequence cluster i and cluster j.

Through the mining and analysis for user behavior data, we can not only obtain the user's potential preference and demand, but also predict the change of user's preference according to the abnormal information such as browsing time and clicks. According to the user's explicit preference and behavior analysis, we can also predict the commodity items which are not evaluated by the user, and thus to obtain the user's implicit preference knowledge [18].

After the above processing, the weighted keyword vector model is used to construct the matrix of user preference space.

$$U_{PS} = \{(i_1, w_1), (i_2, w_2), (i_k, w_k)\} \tag{4}$$

In Formula (4), $i_k$is the kth preference type of user. $w_k$is the weight of the kth preference type in the user's preferences.

3.2. Formation of nearest neighbor community

The key of intelligent information recommendation algorithm is to accurately locate the nearest neighbor of target user. The basis of determining the nearest neighbor is to calculate the similarity between users. There are three common calculation methods.

The first is Pearson correlation similarity [19], and its formula is:

$$sim(u, v) = \frac{\sum_{\alpha \in P_{uv}} \left(R_{u,\alpha} - \overline{R}_u\right)\left(R_{v,\alpha} - \overline{R}_v\right)}{\sqrt{\sum_{\alpha \in P_{uv}} \left(R_{u,\alpha} - \overline{R}_u\right)^2} \sqrt{\sum_{\alpha \in P_{uv}} \left(R_{v,\alpha} - \overline{R}_v\right)^2}} \tag{5}$$

In the formula, P_{uv}represents the set of scoring items. $R_{u,\alpha}$and $R_{v,\alpha}$represent the scores for item α, respectively. $\overline{R}_v$ represents the average score.

The second is cosine similarity [20]. Let's set the score of user$u$ and user$v$ in the n-dimension item space as the similarity between m and n. Its formula is:

$$sim(u, v) = \frac{m \times n}{|m| \times |n|} \tag{6}$$

The third is the modified cosine similarity [21], which fully considers the scoring methods of different users. Its formula is:

$$sim(u, v) = \frac{\sum_{\alpha \in P_{uv}} \left(R_{u,\alpha} - \overline{R}_u\right)\left(R_{v,\alpha} - \overline{R}_v\right)}{\sqrt{\sum_{\alpha \in P_v} \left(R_{u,\alpha} - \overline{R}_u\right)^2} \sqrt{\sum_{\alpha \in P_v} \left(R_{v,\alpha} - \overline{R}_v\right)^2}} \tag{7}$$

Where, $P_u$and $P_v$are item sets rated by user$u$and user$v$ respectively.

Based on the modified cosine similarity, the user preference knowledge is integrated into the calculation of user similarity. The formula is:

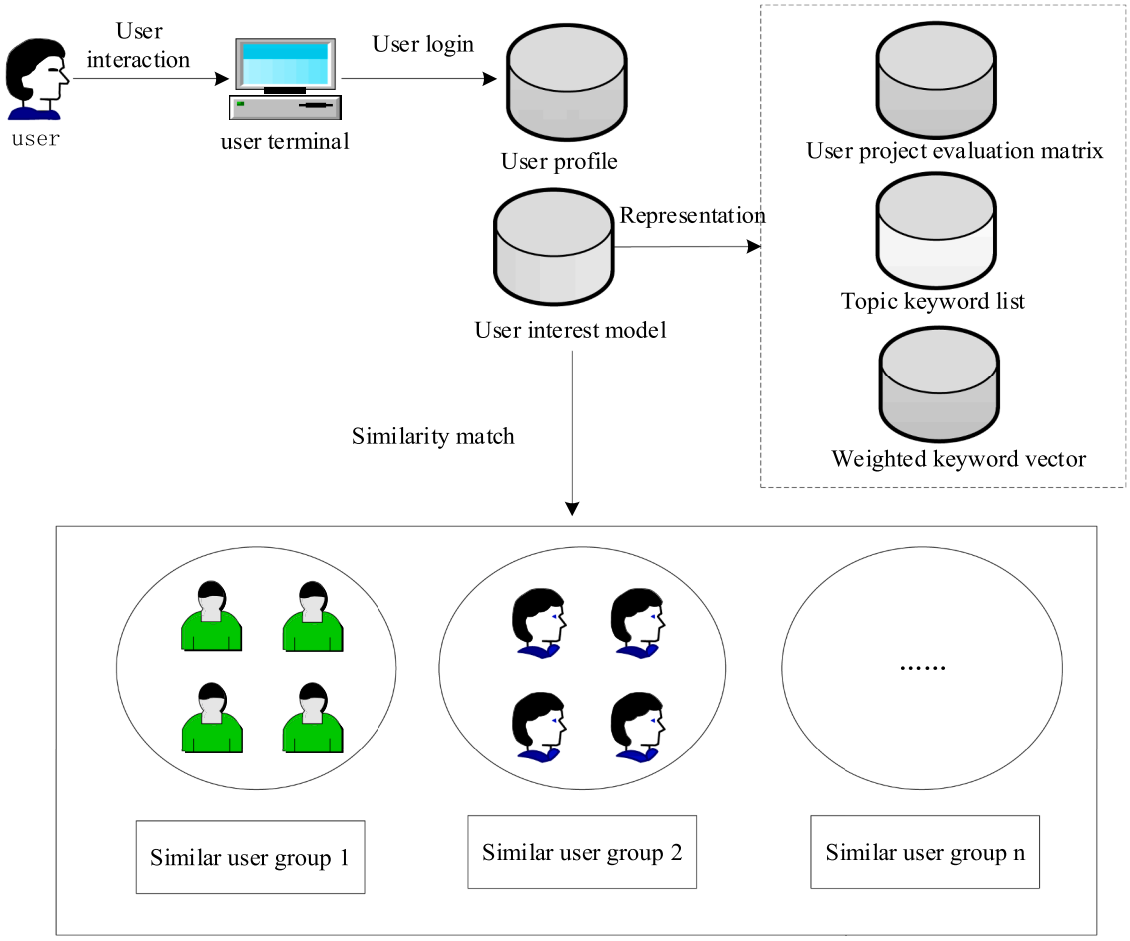

Fig. 4. Initial Intelligent Recommendation Model.

$$sim(u,v) = \frac{\sum_{i \in U_{PSuv}} \left(W_{u,i} - \overline{W}_u\right)\left(W_{v,i} - \overline{W}_v\right)}{\sqrt{\sum_{i \in U_{PSu}} \left(W_{u,i} - \overline{W}_u\right)^2} \sqrt{\sum_{i \in U_{PSv}} \left(W_{v,i} - \overline{W}_v\right)^2}} \tag{8}$$

Where, U_{PSuv}represents a set of shared preference types of user$u$ and user$v$. U_{PSu}, U_{PSv} represents the set of preference types of user$u$ and user$v$, respectively. $W_{u,i}$and $W_{v,i}$represents the weights of user u and user$v$ on preference types, respectively.

By calculating the similarity of any two users in user space, the users who meet the preset threshold are clustered to obtain the nearest neighbor community with the same or similar preference type.

3.3. Construction of intelligent recommendation model

Traditionally, the search engine only uses keywords to mechanically match the content when searching the information of suppliers and demanders of technology. Users' choice of keywords affects the accuracy and recall rate of search results. At present, the user groups of technology innovation platform include enterprises, universities, government, scientific research institutions and other units or individuals, the technical qualities of these users are different, and they have different expressions on the same technology and use different professional terms. For example, the technical demander chooses the key words "fresh milk" and "fresh keeping", and the technology supplier chooses the technical terms " modified atmosphere method based on film storage " and "thermal exhaust method". In this case, if the search engine directly matches keywords, it will lead to the omission of recommendation results.

When the supply side and demand side of technology releases the project information of technology supply and demand, the demand may be expressed in natural language. In order to further improve the quality of recommendation, it is necessary to correlate with the semantic relations between the explicit and implicit keywords. However, a large number of keywords and complex keywords involve many technical industries and fields. These aspects also present challenges to the data processing ability of information recommendation.

In order to achieve good technology promotion, the information recommendation system must be able to help users quickly find the required information recommendation items in massive and complex information. This puts forward high requirements for the speed and accuracy of recommendation system. The information to be processed includes website information and user personality information. Due to the large amount of information, complex structure, difficult calculation and high requirement of computation speed, the information recommendation system needs to introduce big data processing technology to balance the contradiction between massive heterogeneous data and rapid response requirement.

The main task of intelligent recommendation is to automatically recommend the commodity items to target users through the nearest neighbor community. Let's suppose the target user to be recommended is m, user similarity is sim. The nearest neighbor users of user m in the whole user preference spaceU_{PS} are searched, and then the nearest neighbor sets$U_{PSmi} = \{U_{PS1}, U_{PS2}, \ldots, U_{PSk}\}$ which has the same or similar preference with$m$ are obtained. For the preference type i, the similarity$sim(m, U_{PS1})$ of $U_{PS1}$and $m$is the highest. The second is the similarity$sim(m, U_{PS2})$ of $U_{PS2}$and$m$, and so on. The weighted average of the preference information of each user in U_{PSmi} is used to predict the

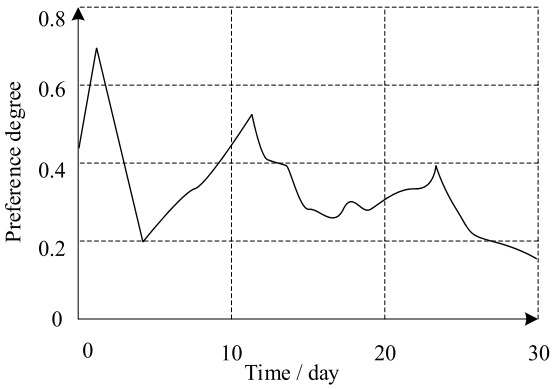

Fig. 5. Example of User Preference Sequence.

preference and demand of target user. The formula is:

$$P_{m,i} = \overline{W}_m + \frac{\sum_{u \in U_{PSmi}} sim(m,u) \times \left(W_{u,i} - \overline{W}_u\right)}{\sum_{u \in U_{PSmi}} |sim(m,u)|} \tag{9}$$

In the formula, $W_{u,i}$represents the weight of user$u$ for the preference type$i$. $sim(m,u)$ represents the similarity between user m and user u. $\overline{W}_m$and $\overline{W}_u$represent the average weight on all preference types. The fist projects with the highest prediction preference and demand are calculated by Formula (9). That is to say, Top – N will be recommended to the target users.

In the recommendation system on technology innovation platform, the intelligent information recommendation model is very important. The initial information recommendation model mainly includes the construction of user interest model, similarity (neighbor) user search and similarity matching. The contents are shown in Fig. 4.

The intelligent recommendation model refers to the establishment of personalized technology preference, industry and other related knowledge of the individual in technology supply and demand. As a professional collaborative innovation platform, the industry information and technology information of the technology innovation platform will become more important. Through accurate and true reflection for the preferences of different users, high-quality information can be recommended.

A good recommendation algorithm must ensure its recommendation accuracy. How to improve the recommendation accuracy is an important research direction in the recommendation algorithm. The recommendation algorithm with high accuracy can provide users with goods that meet their interests, increase users' satisfaction and improve the adhesion of users to the recommendation system.

4. Experimental analysis

Under the background of big data, the intelligent information recommendation algorithm is researched. In the experiment, it is necessary to standardize, store and recommend multiple data sources. In the experiment, the accuracy and effectiveness of the proposed algorithm are verified by comparing the algorithms in Reference [1], Reference [2] and Reference [3].

The user's preference order for Internet source data is shown in Fig. 5. Taking one month as the cycle, the user's preference for Internet data is higher on the second day of each month, so that the user's recommendation on that day can meet user's needs to a greater extent. By mining the users with the same preference, we can recommend users' preferred resource to achieve better recommendation effect. In addition to pushing multiple data sets to the user according to the recommendation degree of data source, if the user's preference is similar to other online retrieval users, the high satisfaction data sets retrieved by other users will be recommended when the minimum recommendation

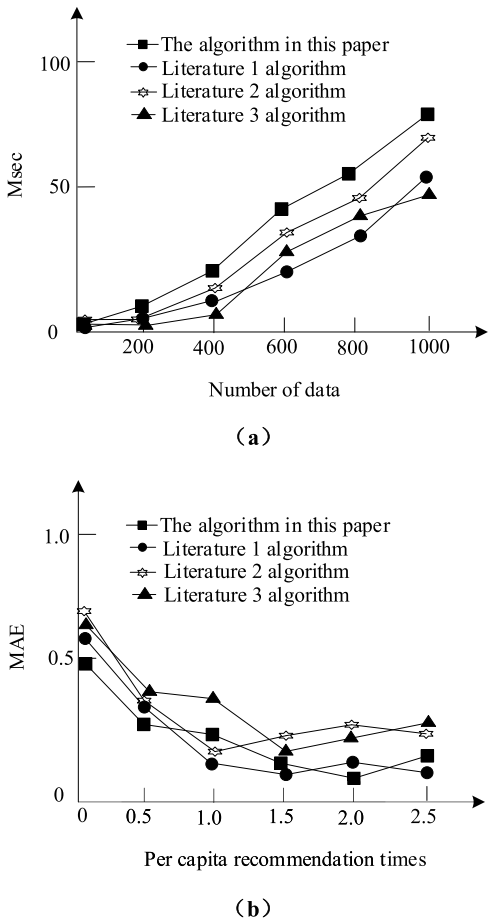

Fig. 6. Recommendation Effect of Single Data Source.

threshold is satisfied. In the recommendation system, we can give consideration to the inherent characteristics of data sources and the ever-changing interest.

The user preference sequence example is obtained through Fig. 5, and the experiments are carried out. The process is divided into two parts. The first part is to compare the effectiveness of different algorithms in unique data source when$PT(x) = 1$. The second part is to compare the different recommendation algorithms in the case of multiple data sources.

In the experiment, the mean absolute differenceM_{AE} is used as the standard to evaluate the recommendation quality. It evaluates the accuracy of the algorithm by calculating the deviation between the predicted user's interest in the recommended dataset and the user's real interest in the product. The smaller the valueM_{AE}, the more accurate the prediction, the higher the recommendation accuracy. The recommended values of the algorithm for k recommended dataset are $\{PA(1), PA(2), \ldots, PA(k)\}$, respectively, and the user's real interest values are$\{R_1, R_2, \ldots R_k\}$. The value of real interest measure is 0 or 1, which represents the low evaluation or high evaluation on the data set. The average absolute deviation is shown as follows:

$$M_{AE} = \frac{\sum_{i=1}^{n} |R_{ij} - PA_{ij}|}{n \times k} \tag{10}$$

The single data source uses *moverens* data set, which contains a total of 100,000 real records of 1682 movies rated by 943 anonymous users. The proposed algorithm applies all the data to the interest learning, and then updates the user interest continuously after the user's rating or evaluation information is updated. However, the algorithms in references can't mine the user's preference sequence of different data

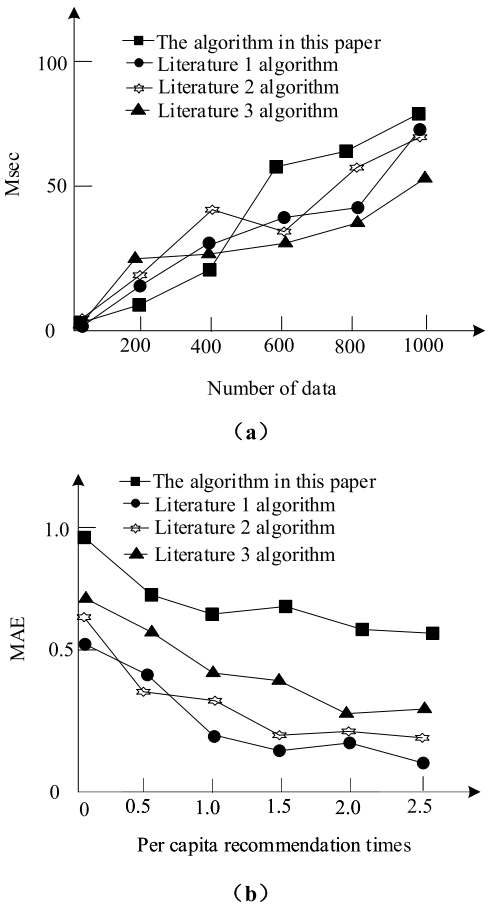

Fig. 7. Effect Recommendation of Multiple Data Source.

sources. In this case, the efficiency and accuracy of recommendation are shown in Fig. 6.

The experimental results show that the collaborative recommendation is completed by calculating the similarity of interests between different users, and the three algorithms in literatures complete the recommendation by calculating the data correlation in user interest data set, and increases the availability of data to improve the credibility of user evaluation. The details are shown in Fig. 6.

In the experiment, two kinds of data sets were adopted. One was two hundred comments and analysis articles of Internet users on a hot topic, and the credibility of the data set was set to 0.3. The other was one hundred expert's editorials of mainstream media on the topic, and the credibility of the data set was set to 0.8. The experiment established a unified full-text index for different data sets and recommended centrally. There was no pre-learning process in the two algorithms. It was necessary to carry out the self-learning after updating the user recommendation information. Under this condition, the recommendation efficiency and recommendation accuracy of different methods are shown in Fig. 7.

According to the analysis of experimental results, it can be concluded that the above algorithms can't effectively recommend the data sets with different quality. With the increase of the per capita recommendation times, the recommendation accuracy can't be improved rapidly. The proposed algorithm can greatly improve the accuracy of recommendation by mining user preference and analyzing the association between data sets and users.

5. Conclusions

Traditionally, the recommendation algorithm only uses the user's explicit preference information when calculating the user similarity, but ignores the user's implicit preference. Therefore, this article researches on the intelligent information recommendation algorithm under the background of big data. The user preference mining technology is used to obtain explicit and implicit user preference knowledge and calculate user similarity, and thus to realize the formation mechanism of nearest neighbor community based on user preference knowledge and the intelligent recommendation for user needs. Experimental results show that the algorithm achieves the desired results. The collaborative filtering recommendation based on user preference knowledge is the key to improve the accuracy and quality of recommendation result.

With the explosive growth of data, the research is deeply inadequate in the face of more massive data due to the limitation of its own ability. In view of some defects in this article, several research directions are proposed based on the knowledge level.

(1) The problem of data sparsity: at present, the amount of available data is increasing rapidly, but the problem of data sparsity is becoming more and more serious. The existing recommendation algorithms can't adapt to the data sparsity. How to solve the problem of data sparsity and improve the efficiency and accuracy of algorithm is the future research direction.
(2) The problem of cold start problem: at present, there is no good strategy for the cold start problem of recommendation algorithm, so how to solve the problem of the cold start problem of recommendation algorithm is also the future research direction.

Declaration of Competing Interest

There is no Conflict of interest in this research paper.

Acknowledgement

This work was supported by Science and Technology Project of Jiangxi Education Department under grant no. GJJ191001, GJJ191004, GJJ180978, and Teaching Reform Research Project of Jiangxi Education Department under grant no. JXJG-19–24–8.

References

[1] R. Ali, M. Afzal, M. Sadiq, M. Hussain, Knowledge-based reasoning and recommendation framework for intelligent decision making, Expert Systems 35 (2) (2018) 131–142.
[2] V. Sourabh, C.R. Chowdary, Peer recommendation in dynamic attributed graphs, Expert Syst Appl 120 (5) (2018) 335–345.
[3] H. Yu, B. Zhou, M. Deng, F. Hu, Tag recommendation method in folksonomy based on user tagging status, J Intell Inf Syst 50 (3) (2018) 479–500.
[4] Q. Zhang, L.T. Yang, Z. Chen, P. Li, A survey on deep learning for big data, Information Fusion 42 (2018) 146–157.
[5] C.G. Petra, N. Chiang, M. Anitescu, A structured quasi-newton algorithm for optimizing with incomplete Hessian information, Siam J Optimization 29 (2) (2019) 1048–1075.
[6] S.S. Sohail, J. Siddiqui, R. Ali, An OWA-based ranking approach for university books recommendation, Int J Intelligent Systems 33 (2) (2018) 396–416.
[7] D. Claudia, S. Matthias, NOMAD: the FAIR Concept for Big-Data-Driven Materials Science, MRS Bulletin 43 (09) (2018) 676–682.
[8] J.Q. Wang, X. Zhang, H.Y. Zhang, Hotel recommendation approach based on the online consumer reviews using interval neutrosophic linguistic numbers, J Intelligent & Fuzzy Systems 34 (1) (2018) 381–394.
[9] H. Liu, L. Yang, C. Ling, X. Wu, Collaborative social deep learning for celebrity recommendation, Intelligent Data Analysis 22 (6) (2018) 1375–1394.
[10] R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, X. Li, Efficient user profiling based intelligent travel recommender system for individual and group of users, Mobile networks & applications 24 (3) (2019) 1018–1033.
[11] Y. Qiao, Z. Xing, M. Fadlullah Z, J. Yang, Kato, Characterizing Flow, Application, and User Behavior in Mobile Networks: a Framework for Mobile Big Data, IEEE Wireless Communications 25 (1) (2018) 40–49.
[12] D. Li, A. Madden, C. Liu, Modelling online user behavior for medical knowledge learning, Industrial Management & Data Systems 118 (4) (2018) 889–911.
[13] S. Wang, D. Lo, B. Vasilescu, EnTagRec~(++): an enhanced tag recommendation system for software information sites, Empirical Software Engineering 23 (2) (2018) 800–832.

[14] F. Wang, X. Meng, Y. Zhang, C. Zhang, Mining user preferences of new locations on location-based social networks: a multidimensional cloud model approach, Wireless Networks 24 (1) (2018) 113–125.
[15] Y. Liu, T. Yang, T. Qi, An Attention-Based User Preference Matching Network for Recommender System, IEEE Access 8 (99) (2020) 41100–41107.
[16] L. Teddy, S. Siva, M. Eric, A. Shalaby, Subway user behaviour when affected by incidents in Toronto (SUBWAIT) survey — A joint revealed preference and stated preference survey with a trip planner tool, Canadian J Civil Engineering 45 (8) (2018) 623–633.
[17] R. Liu, J. Cao, K. Zhang, Gao Wenyu., J. Liang, L. Yang, When Privacy Meets Usability: unobtrusive Privacy Permission Recommendation System for Mobile Apps Based on Crowdsourcing, IEEE Transactions on Services Computing 11 (5) (2018) 864–878.
[18] J. Hu, J. Liang, Y. Kuang, H. Vasant, A user similarity-based top- N recommendation approach for mobile in-application advertising, Expert Syst Appl 111 (6) (2018) 51–60.
[19] A. Belesiotis, D. Skoutas, C. Efstathiades, Spatio-textual user matching and clustering based on set similarity joins, The VLDB journal 27 (3) (2018) 297–320.
[20] J. Choo, H. Kim, E. Clarkson, Z. Liu, C. Lee, F. Li, VisIRR: a Visual Analytics System for Information Retrieval and Recommendation for Large-Scale Document Data, ACM Trans Knowl Discov Data 12 (1) (2018) 1–20.
[21] M. Cheung, J. She, N. Wang, Characterizing user connections in social media through user-shared images, IEEE Transactions on Big Data 4 (4) (2018) 447–458.

Haoxiang Tang (1998-) received his Bachelor degree in environment engineering from Beijing Normal University in 2020. He received the admission to the MS in Information System program at the Johns Hopkins Carey Business School for Fall 2020. His-research interests include data mining, database development/application and expert system design.

Wei Yang (1985-) received his Master degree in software engineering from Yu Nan University in 2015. He is currently an associate professor in the center of collaboration and innovation of Jiangxi University of Technology. His-research interests include data mining, database development and application and expert system design

Susheng Zheng (1978-) received his Master degree in business administration from Guang Xi University in 2017. He is currently a lecturer of Jiangxi University of Technology. His-research interests include enterprise information management, data mining of enterprise massive data and business intelligence