A Novel Approach for Keyword-Aware Service Recommendation System using Collaborative- Filtering Algorithm to Implement in Hadoop Framework

Mr. M. Vengateshwaran M.E[1], T.S Janarthanan[2], S. Senbagam[3], S. Shanmugapriya[4]

Professor& Head, Department of Computer Science & Engineering, MASS Group of Institutions, Kumbakonam[1]

Student, Department of Computer Science & Engineering, MASS Group of Institutions, Kumbakonam[2,3,4]

Abstract: Service recommendation system is an one of the valuable tool and it provide a appropriate recommendation to the users. Now a days large number of customers, services and the other online information needs service recommendation system. It provides the recommendations about the services to the user. In the existing recommendation system, the ratings of services and the service recommendation lists presented to users are the same.In the existing system the major problem is scalability and inefficiency problems when processing or analyzing such large-scale data. Here we propose a Keyword-Aware Service Recommendation method. In the proposed system it recommends the services based on the user preferences. In this system keyword refers to the user preferences. In this system user based collaborative filtering algorithm is used to generate recommendations. To improve the efficiency here we implement the system in Hadoop. In addition to that to make services effective, here we are going to use rank boosting algorithm with combined preferences. To achieve the scalability and efficiency with help of mapreduce framework in an big data environment.

Keywords: Recommendation system, Hadoop, collaborative algorithm, preference, keyword, Mapreduce.

I. INTRODUCTION

In recent years, large amount of data in our world has been increasing explosively, and analyzing large data sets so called "Big Data" becomes a key basis of competition underpinning new waves of productivity growth, innovation,and consumer surplus.Then,what is "Big Data"?,Big data refers to data sets whose size is beyond the ability of current technology, method and theory to capture, manage,and process the data within a tolerable elapsed time.Today, Big Data management stands out as a challenge for IT companies. The solution to such a challenge is shifting increasingly from providing hardware to provisioning more manageable software solutions. Big data also brings new opportunities and critical challenges to industry and academia.Similar to most big data applications, the big data tendancy also poses heavy impacts on service recommender systems.With the growing number of alternative services, effectively recommending services that a user preferred has become an important research issue. Service recommender systems have been shown as valuable tools to help users deal with services overload and provide appropriate recommendations to them. Examples of such practical applications include CDs, books, webpages and various other products now use recommender systems.Over the last decade, there has been much research done both in industry and academia on developing new approaches for service recommender systems.

II. MOTIVATION

Motivated by these observations, in this paper, we address these challenges through the following contributions:
(1) A keyword-aware service recommendation method, named KASR, is proposed in this paper, which is based on a user-based Collaborative Filtering (CF) algorithm.
(2) In KASR, keywords extracted from reviews of previous users are used to indicate their preferences. Moreover, we implement it on a distributed computing platform, Hadoop, which uses MapReduce as its computing framework.

III.LITURATURE REVIEW

3.1 Title:Multi-Criteria User Modeling in Recommender Systems Kleanthi Lakiotaki, Nikolaos F. Matsatsinis
In parallel, Multiple Criteria Decision Analysis (MCDA) is a well established field of Decision Science that aims at

analyzing and modeling decision maker's value system, in order to support him/her in the decision making process. In this work, a hybrid framework that incorporates techniques from the field of MCDA, together with the Collaborative Filtering approach, is analyzed. The proposed methodology improves the performance of simple Multi-rating Recommender Systems as a result of two main causes; the creation of groups of user profiles prior to the application of Collaborative Filtering algorithm and the fact that these profiles are the result of a user modeling process, which is based on individual user's value system and exploits Multiple Criteria Decision Analysis techniques. Experiments in real user data prove the aforementioned statement.In the presented work, a hybrid methodological framework is proposed that combines techniques from the field of Multiple Criteria Decision Analysis and more specifically from the Disaggregation-Aggregation approach to model user's preferences, together with the Collaborative Filtering technique from the field of Recommender Systems, to identify the most preferred unknown items for every user.

The proposed methodology was demonstrated as a movie Recommender System and its performance was tested in real user data. Furthermore, comparison study with other single and multiple criteria collaborative filtering methodologies clearly proves that the creation of user profiles constitutes an integral part of a recommendation process. These profiles are formed by an especially designed to treat multiple criteria approach, named the Disaggregation-Aggregation approach, as demonstrated by the UTA* algorithm. A sophisticated MCDA method, like the aforementioned, applied to build user profiles and the subsequent clustering of these profiles enhances the performance of any collaborative filtering technique used to predict preferences and offers new insights for Recommender Systems researchers to penetrate into novel, multidisciplinary, user modeling methodologies.The key to more effective personalization services is the ability to develop a system able to understand not only what people like, but why they like it. In other words, an accurate modeling of user value system and an effective preference representation schema, will potentially lead to the design of a recommendation algorithm with increased performance. Such a system can understandhow users think about items, by considering the knowledge about the underlying attributes that attract users to choose this particular item and hence recognize preferences, not just patterns, ensuring a more sophisticated understanding of the user.

Advantage:
- This proposed work improves the performance of simple Multi-rating Recommender Systems.
- It provides flexibility to examine every user individually

Disasdvantage:
- It fails to compute a rating in the case of a single common item.
- The recommendation process as a decision problem and exploit techniques from Decision Theory

3.2 Title:QoS Ranking Prediction for Cloud Services-Zibin Zheng

QoS rankingsprovide valuable information for making optimal cloud service selection from a set of functionally equivalent service candidates. Toobtain QoS values, real-world invocations on the service candidates are usually required. To avoid the time-consuming and expensivereal-world service invocations, this paper proposes a QoS ranking prediction framework for cloud services by taking advantage of thepast service usage experiences of other consumers. Our proposed framework requires no additional invocations of cloud serviceswhen making QoS ranking prediction. Two personalized QoS ranking prediction approaches are proposed to predict the QoS rankingsdirectly.

Similar to traditional component-based systems, cloudapplications typically involve multiple cloud componentscommunicating with each other over application programminginterfaces, such as through web services. An example of cloud applications. Cloud application 1 is a tourism Website deployed in the cloud, providingvarious types of tourism services to customers. The businessprocess of this cloud application is composed by a number ofsoftware components, where each component fulfills aspecified functionality. To outsource part of business toother companies, some of these components invoke othercloud services.

The most straightforward approach of personalized cloudservice QoS ranking is to evaluate all the candidate servicesat the user-side and rank the services based on the observedQoS values. However, this approach is impractical in reality,since invocations of cloud services may be charged. Even ifthe invocations are free, executing a large number of serviceinvocations is time consuming and resource consuming, andsome service invocations may produce irreversible effects inthe real world. Moreover, when the number of candidateservices is large, it is difficult for the cloud applicationdesigners to evaluate all the cloud services efficiently.To attack this critical challenge, we propose a personalizedranking prediction framework, named CloudRank, topredict the QoS ranking of a set of cloud services withoutrequiring additional real-world service invocations from theintended users.

By taking advantage of the past usage experiences of otherusers, our ranking approach identifies and aggregates thepreferences between pairs of services to produce a rankingof services. We propose two ranking prediction algorithmsfor computing the service ranking based on the cloudapplication designer's preferences. Experimental resultsshow that our approaches outperform existing rating-basedapproaches and the traditional greedy method

Advantage:
- In our proposed work, accuracy for rank prediction is high.

- The CloudRank2 approach obtains the best prediction accuracy for both response time and throughput.

Disasdvantage:

- Our proposed work doesn't deals with the time aware-QOS rank prediction.
- The critical problem of personalized QoS ranking for cloud services and proposes a QoS ranking prediction framework to address the problem.

IV. EXISTING SYSTEM

In the existing traditional service recommender system, it only deals with the single numerical rating to represent a service's utility as a whole. In this system, they implement the process using the collaborative filtering algorithm. But the problem occurs in this system is the scalability problem. To solve the scalability problem they divide the dataset. But this method doesn't provide the favorable   scalability and efficiency if the amount of data grows. In this traditional service recommender system, it provides same ratings and rankings to all services which are all viewed by the users. In this approach they implement a large-scale video recommendation system. This recommendation system is implemented based on item based collaborative filtering algorithm.

The recommendation process starts with the specification of the initial set of ratings that is either explicitly provided by the users or is implicitly inferred by the system.To integrate multi-criteria rating information into recommender systems discussed in the previous section apply primarily to the similarity-based recommenders, such as traditional collaborative filtering techniques. In contrast, in this section we present a different approach that is not limited to any specific recommendation algorithm. Generally speaking, service composition is promoted in an open web environment. For a private the privacy and security are crucial issues in service access.It often leads to an awkard situation that some QoS information may be unavailable in cross-cloud composition evaluation. It is just the reason that although it is assumed that the history records can be obtained through some monitoring mechanism there is few general QoS dataset widely recruited for testing the performance and accuracy of history record-aware service composition as mentioned.

Performance of item-based prediction algorithms is of superior quality than user-based prediction algorithms. Performance of implicit rating based algorithms, in the sense that they have been defined for these tests, is of inferior quality than explicit rating based algorithms. Item-based algorithm, based on explicit ratings(CFIB-ER) seems to be very sensitive to sparsity levels. As sparsity reduces, the MAE of the algorithm decreases, which means that prediction accuracy is increased. CFIB-ER performs better than classic Collaborative Filtering prediction algorithm, CFUB-ER,for Sparsity levels close to 97.2%.CFUB-ER-CBincreases the accuracy precision of CFUB-ER,as it calculates the user average based only on the subset of items that belong to the same categories as the active item. However, this increment is insignificant if taking into consideration the extra computation needed to include the category information.

DISADVANTAGES

- Scalability and Inefficiency problems when processing or analyzing large scale data.
- Existing traditional recommender system fails to meet the user personalized requirements.
- In traditional recommender system, the ratings and rankings given the services are same.

V. PROPOSED SYSTEM

To provide the user preferences based recommendation services, here we propose a keyword aware service recommendation system. In this keywords are used to indicate both of users' preferences and the quality of candidate services.Evaluating a service through multiple criteria and taking into account of user feedback can help to make more effective recommendations for the users. Our proposed system is simply called as KASR. To implement the keyword service Based on the keyword service recommendations are provided for the user. For this process here we use a user-based collaborative filtering algorithm. The user-based collaborative filtering algorithm is used to provide the efficient recommendation list about the services to the users. To improve the efficiency of this process we implement this in hadoop environment.

Cloud computing can provide effective platforms to facilitate parallel computing, which has gained significant attention in recent years to process large volume of data. There are several cloud computing tools available, such as Hadoop, Mahout, MapReduce of Google, the Dynamo of Amazon.com, the Dryad of Microsoft and Neptune of Ask.com, etc. Among these tools, Hadoop is the most popular open source cloud computing platform inspired by MapReduce and Google File System papers, which supports MapReduce programming framework and mass data storage with good fault tolerance. MapReduce is a popular distributed implementation model proposed by Google, which is inspired by map and reduce operations in the Lisp programming language.

The majority of existing Recommender Systems obtains an overall numerical rating, as input information for the recommendation algorithm. This overall rating depends only on one single criterion that usually represents the overall preference of user u on item i. However, articles like underline the preference of stirring Recommender Systems

researchers towards a more user oriented perspective, indicating that people are not truly satisfied by existing Recommender Systems. The service recommender systems, users tend to be recommended the top services of the returned result list. The services in higher position, especially the first position, should be more satisfying than the services in lower position of the returned result list. To evaluate the quality of Top-K service recommendation list, MAP and DCG are used as performance evaluation metrics. And the higher MAP or DCG presents the higher quality of the predicted service recommendation list.

ADVANTAGES

- To improves the scalability and efficiency , KASR is implemented in Hadoop.
- KASR main aimed at presenting the personalized rating of each candidate service for a user.
- In KASR keywords are extracted from reviews of previous users are used to indicate their preferences.
- In KASR, keyword-candidate list and domain thesaurus are provided to help obtain users' preferences.

SYMBOLS FUNCTION

SYMBOLS	DEFINITION
K	The keyword-candidate list,k={k1,k2,..kn}
APK	The preference keyword set of active user
PPK	The preference keyword set of previous user
Sim(APK,PPK)	Similarity between APK and PPK
Wp	Preference weight vector
Wap	Preference weight vector of active user
Wpp	Preference weight vector of previous user

VI. PROPOSED SYSTEM ARCHITECTURE
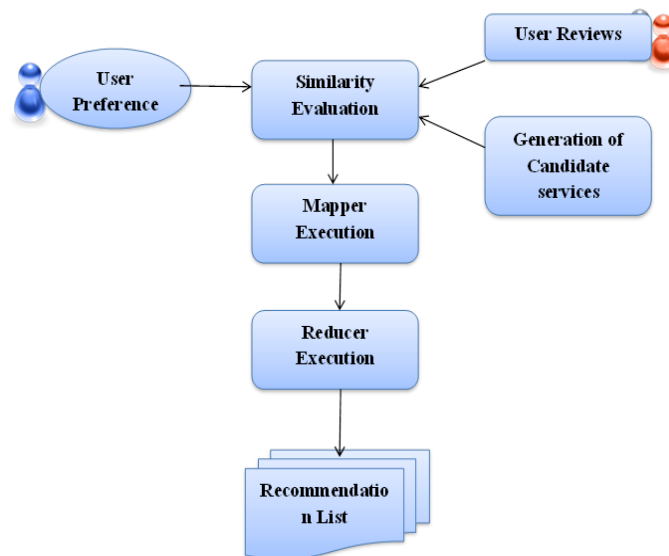
VII. SYSTEM MODULE IMPLEMENTATION

7.1 Loading And Preprocessing Of Data

In this module , we first load the data. After loading , we analyze the data. After analyzing process , we view the information present in the dataset. After that we start the preprocessing step. In the preprocessing step,we remove the null value,missing tuples etc. Here we are going to process three set of data.First one is user information.It consists of user id ,user age,profession,gender,zipcode etc.Second one is ratings information.The ratings data consists of user id,item id,ratings and timestamp.Third one is movie dataset.It consists of movieid,name,release date,imdb url category of the movie etc.We first load all the data into the hadoop distributed file system and preprocessing all the data.

7.2 Analysis Of User Reviews

After preprocessing , we view the cleaned or processed data. At the same time we analyze the user reviews. The user reviews contain the information about the place or hotels or transportation etc. Using the review we further continue

our process.In this process we are going to analyze the user reviews.i.e.the users who all are watch the movie information already and provide the ratings to thatmovies.In this processing we compare the relevant results to the ratings dataset and retreive all the ratings of the relevant results for an movie.

7.3 Mapper And Reducer Process

In this module we first collect the user preferences in the form of query model. We implement the query model to get the user request, here it is user preference. The user preference is processed using the map reduce mechanism. The process is performed by splitting the preferences i.e.it is done using mapper process. After the processing , the results are aggregated.In this process separation of ratings takes place.Based on the ratings the movies also categorized.The categorization take place based on the user preferences.The review processing and relevant results are analyzed based on the user preferences.

7.4 Prediction Of Recommendation List

After map reduce process execution ,we aggregate the result to generate the recommendation list.The recommendation list is generated using user collaborative filtering algorithm.This algorithm generates the output ,recommendation list.A keyword-aware service recommendation method, named in this paper, which is based on a user-based Collaborative Filtering algorithm.Keywords extracted from reviews of previous users are used to indicate their preferences.

7.5 Evaluation Process

The evaluation process takes place to predict the accuracy of the recommendation list.The evaluation is in the form of graphs.To evaluatethe performance of KASR in accuracy, we compare KASRwith other two well-known recommendation methods: Rank Boost algorithm using Pearson Correlation Coefficient(PCC) and item-based algorithm using PCC, which arecalled as UPCC and IPCC respectively.

VIII. EXPERIMENTAL RESULTS SHOTS
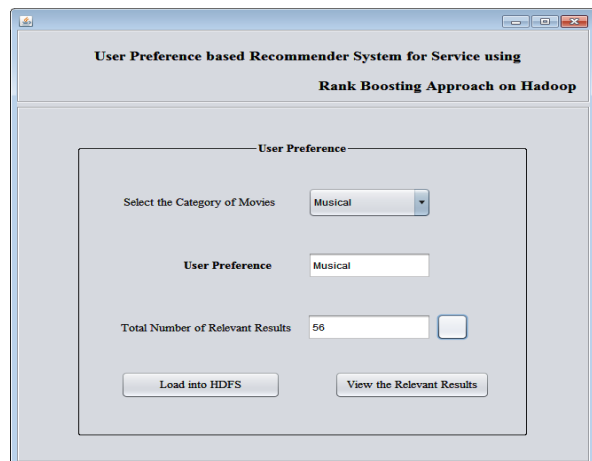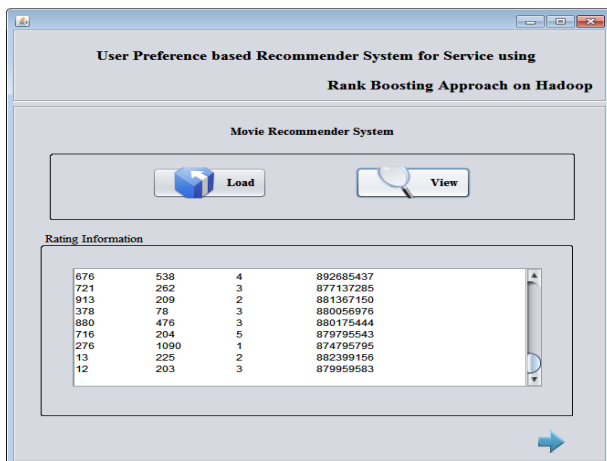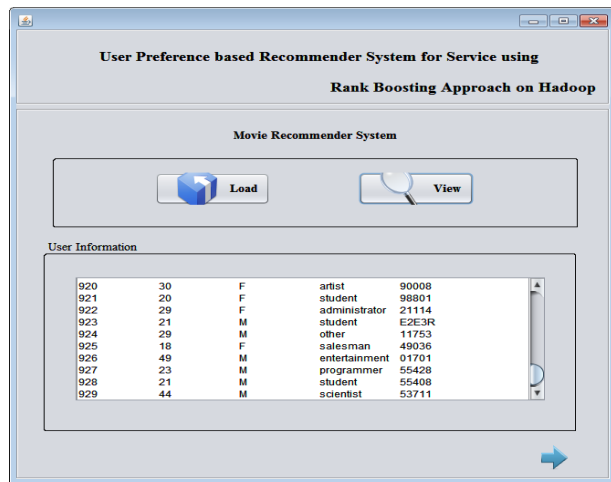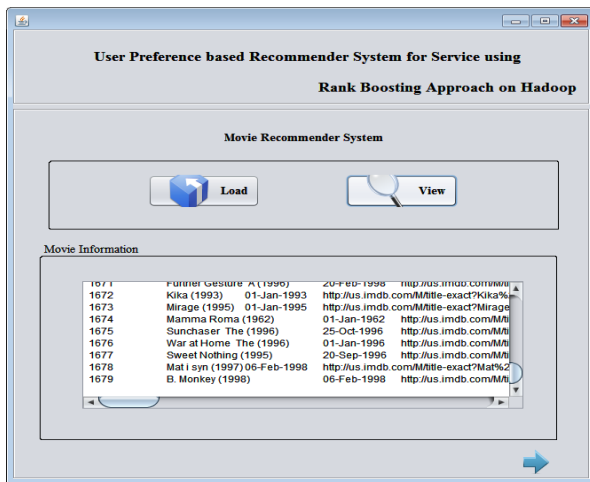
IX. CONCLUSION

Now a days , increasing large number of services ,on-line information etc rapidly growth.In KASR, key-words are used to indicate users' preferences, and a user-based Collaborative Filtering algorithm is adopted to generate appropriate recommendations. More specifically, a keyword-candidate list and domain thesaurus are provided to help obtain users' preferences. The active user gives his/her preferences by selecting the keywords from the keyword-candidate list, and the preferences of the previous users can be extracted from their reviews for services ac-cording to the keyword-candidate list and domain thesaurus. Our method aims at presenting a personalized service recommendation list and recommending the most appropriate service(s) to the users. Moreover, to improve the scalability and efficiency of KASR in "Big Data" environment, we have implemented it on a MapReduce framework in Hadoop platform. Finally, the experimental results demonstrate that KASR significantly improves the accuracy and scalability of service recommender systems over existing approaches. The user profile is mapped into the domain ontology to get initial recommended items. These items are mapped into UDPO to infer user potential interests and semantic similarity. Our experimental results on a real-world dataset confirm that, when available, multi-criteria ratings can be successfully leveraged to improve recommendation accuracy. We expect that the proposed approaches will be useful in other application domains as well, where they will be able to predict overall ratings more accurately by utilizing the available multi-criteria rating information. This work described how explicit ratings can be utilized in order to implicitly obtain user's preference to specific categories. A number of prediction algorithms have been designed and implemented, based on either user or item similarity and have been thoroughly evaluated according to their statistical and decision-support accuracy performance.

REFERENCES

[1]   J. Manyika, M. Chui, B. Brown, et al, "Big Data: The next frontier for innovation, competition, and productivity," 2011.
[2]   C. Lynch, "Big Data: How do your data grow?" Nature, Vol. 455, No. 7209, p. 28-29, 2008.
[3]   F. Chang, J. Dean, S. Ghemawat, and W. C. Hsieh, "Bigtable: A distributed storage system for structured data," ACM Transactions on Computer Systems,Vol. 26, No. 2 (4) , 2008.
[4]   W. Dou, X. Zhang, J. Liu, J. Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applica-tions," IEEE Transactions on Parallel and Distributed Systems, 2013.
[5]   G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, Vol. 7, No.1, pp. 76-80, 2003.
[6]   M. Bjelica, "Towards TV Recommender System Experiments with User Modeling," IEEE Transactions on Consumer Electronics, Vol. 56, No.3, pp. 1763-1769, 2010.
[7]   M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sport Videos Based on User Audiovisual Consumption," IEEE Transactions on Multimedia, Vol. 14, No.6, pp. 1546-1557, 2013.
[8]   Y. Chen, A. Cheng and W. Hsu, "Travel Recommendation by Min-ing People Attributes and Travel Group Types From Community-Contributed Photos".IEEE Transactions on Multimedia, Vol. 25, No.6, pp. 1283-1295, 2012.
[9]   Z. Zheng, X Wu, Y Zhang, M Lyu, and J Wang, "QoS Ranking Pre-diction for Cloud Services," IEEE Transactions on Parallel and Dis-tributed Systems, Vol. 24, No. 6, pp. 1213-1222, 2013.
[10] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and Evaluating Choices in a Virtual Community of Use," In CHI '95 Proceedings of the SIGCHI Conference on Human Factors in Com-puting System, pp. 194-201, 1995. [11] P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering .

BIOGRAPHIES

Mr. M. Vengateshwaran B.E., M.E., Asst. Professor & Head in CSE, Mass Group of Institutions, Kumbakonam. Specialization: Bigdata, Datamining, IR, Database, SE etc,

T.S Janathanan, Student in Department of CSE, Mass Group of Institutions, Kumbakonam

S. Senbagam, Student in Department of CSE, Mass Group of Institutions, Kumbakonam

S. Shanmugapriya Student in Department of CSE, Mass Group of Institutions, Kumbakonam