IJMTST | Volume: 2 | Issue: 05 | May 2016 | ISSN: 2455-3778

User Preferences Based Recommendation System for Services using Mapreduce Approach

Chaithra G V [1] | Nagarathna [2]

[1] PG Scholar, Department of Computer Science and Engineering, PES College of Engineering, Mandya, Karnataka, India

[2] Associate Professor, Department of Computer Science and Engineering, PES College of Engineering, Mandya, Karnataka, India

ABSTRACT

Service recommendations based on the user preferences using keyword aware service recommendation system simply called as KASR. Here the keyword shows the preference of the user. Based on the keyword service, recommendations are provided for the user. For this process we use a user-based collaborative filtering algorithm. To improve the efficiency of this process we implement KASR in Hadoop environment which is a open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs. To improve the efficiency and scalability of the KASR we proposed the combined preferences using rank boosting algorithm. In the rank boosting algorithm, it gets the input as combined preferences, based on the preferences it process the similarities with the reviews of the existing users then it provides the ranking to the services. Based on the ranking provided to the services we generate the output recommendations with high similarity matching results as the recommendation list to the end users for their combined preferences.

KEYWORDS:KASR, Hadoop, RankBoosting, user-based collaborative filtering algorithm

I. INTRODUCTION

In recent years, the 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 datasets 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 tendency 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.

Propose a keyword aware service recommendation method, named KASR. In this method, keywords are used to indicate both of users' preferences and the quality of candidate services. A user based CF algorithm is adopted to generate appropriate recommendations. KASR aims at calculating a personalized rating of each candidate service for a user, and then presenting a personalized service recommendation list and recommending the most appropriate services.

The amount of information and items got extremely huge, leading to an information overload. It became a big problem to find what the user is actually looking for. For doing a right decision, customers still encounter a very time consuming process in visiting a flood of online retailers, and get worthless information by themselves.

Sometimes the contents of Web documents that customers browse have nothing to do with those that they require indeed.

The system is user friendly and more accurate than the other related works. It gives more personalized recommendations and makes more profits for commercial sites. Therefore, the system becomes an interesting and successful recommender system taking the advantages of ground truth theory and application area.

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 main disadvantage is the 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.

To provide the user preferences based recommendation services, A keyword aware service recommendation system was proposed. 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. 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.

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 pretence of stirring Recommender Systems researchers towards a more user oriented perspective, indicating that people are not truly satisfied by existing Recommender Systems. To overcome the problems in the existing recommendation system, here we propose a combined preference based rank boosting algorithm. It improves the scalability and efficiency, when 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 obtain users' preferences.

II. Literature Survey

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. This proposed work improves the performance of simple Multi-rating Recommender Systems. It provides flexibility to examine every user individually. The main disadvantage is 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.

Zibin Zheng. QoS rankings provide valuable information for making optimal cloud service selection from a set of functionally equivalent service candidates. To obtain QoS values, real-world invocations on the service candidates are usually required. To avoid the time-consuming and expensive real-world service invocations, this paper proposes a QoS ranking prediction framework for cloud services by taking advantage

of the past service usage experiences of other consumers. Our proposed framework requires no additional invocations of cloud services when making QoS ranking prediction. Two personalized QoS ranking prediction approaches are proposed to predict the QoS rankings directly. In this proposed work, accuracy for rank prediction is high. The CloudRank2 approach obtains the best prediction accuracy for both response time and throughput. The disadvantage is the 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.

Faustino Sánchez, María Alduán, Federico Álvarez. This paper describes a recommender system for sport videos, transmitted over the Internet and/or broadcast, in the context of large-scale events, which has been tested for the Olympic Games. The recommender is based on audiovisual consumption and does not depend on the number of users, running only on the client side. This avoids the concurrence, computation and privacy problems of central server approaches in scenarios with a large number of users, such as the Olympic Games. The system has been designed to take advantage of the information available in the videos, which is used along with the implicit information of the user and the modeling of his/her audiovisual content consumption. The system is thus transparent to the user, who does not need to take any specific action. The Advantage is The system is transparent to the user, who does not need to take any specification. Important characteristic is that the system can produce recommendations for both live and recorded events. The disadvantage is the restrictions require that the recommender system runs only on the client side, therefore collaborative filtering or other social techniques cannot be used. The problems of our approach will appear in an uncontrolled scenario, because our system needs specific attribute modeling.

Milan Bjelica. In this paper, we analyze recommender system design under the broadcast scenario, where uplink connection to the network center is not available. We put special emphasis on user modeling algorithm that would be able to efficiently learn the user's interests. Our proposal applies the elements of machine learning and pattern recognition, as well as the information retrieval theory, like vector spaces and cluster hypothesis. The derived algorithm is computationally simple, while experimental results show high acceptance ratio of the proposed recommendations. The advantage is The best tested strategy achieved success rate of 87%. The success rate proves the quality of the proposed user modeling and program recommendation procedure. The disadvantge is it doesn't deal with the heterogeneous area of user interests. The overspecialization of these systems is elsewhere considered as their serious limitation.

Wanchun Dou, Xuyun Zhang, Jianxun Liu, and Jinjun Chen. This paper describes a Privacy-Aware CrossCloud Service Composition for Big Data Applications. Cloud computing promises a scalable infrastructure for processing big data applications such as medical data analysis. Cross-cloud service composition provides a concrete approach capable for large-scale big data processing. However, the complexity of potential compositions of cloud services calls for new composition and aggregation methods, especially when some private clouds refuse to disclose all details of their service transaction records due to business privacy concerns in crosscloud scenarios. Moreover, the credibility of cross-clouds and on-line service compositions will become suspicion, if a cloud fails to deliver its services according to its "promised" quality. In view of these challenges, we propose a privacy-aware crosscloud service composition method, named *HireSome*-II (History record-based Service optimization method) based on its previous basic version *HireSome*-I. In our method, to enhance the credibility of a composition plan, the evaluation of a service is promoted by some of its QoS history records, rather than its advertised QoS values. Besides, the *k*-means algorithm is introduced into our method as a data filtering tool to select representative history records. The advantage is it significantly reduces the time complexity of developing a cross-cloud service composition plan. The quality delivered by service providers does not change over time. The disadvantage is in the proposed approach it doesn't give a best solution to the privacy. There may be a chance of not-preserving the privacy.

III. System Design

The design phase includes Loading and preprocessing of dataset, HDFS upload and categorization, Mapper and Reducer process Similarity Measurement, Prediction of recommendation list modules. Figure 1 depicts the system architecture of KASR along with Rank

Boosting algorithm and Figure 2 depicts the flow of the proposed system.

Loading and preprocessing of dataset

In this module, we first load the dataset. The dataset consist of previous user information who have already access the data, movie information, rating information after loading of the data. After analyzing process, we view the information present in the dataset. After loading the data the data is preprocessed. In the preprocessing step, we remove the null value, missing tuples etc. In the next section we explain about categorization of the movie information.

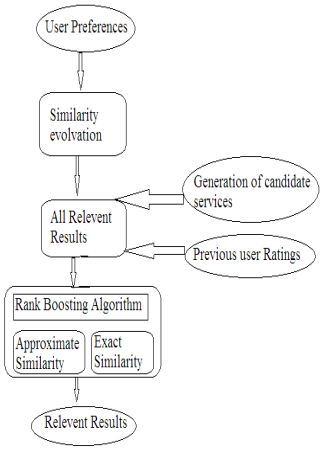

Fig 1 : KASR with Rank Boosting Algorithm

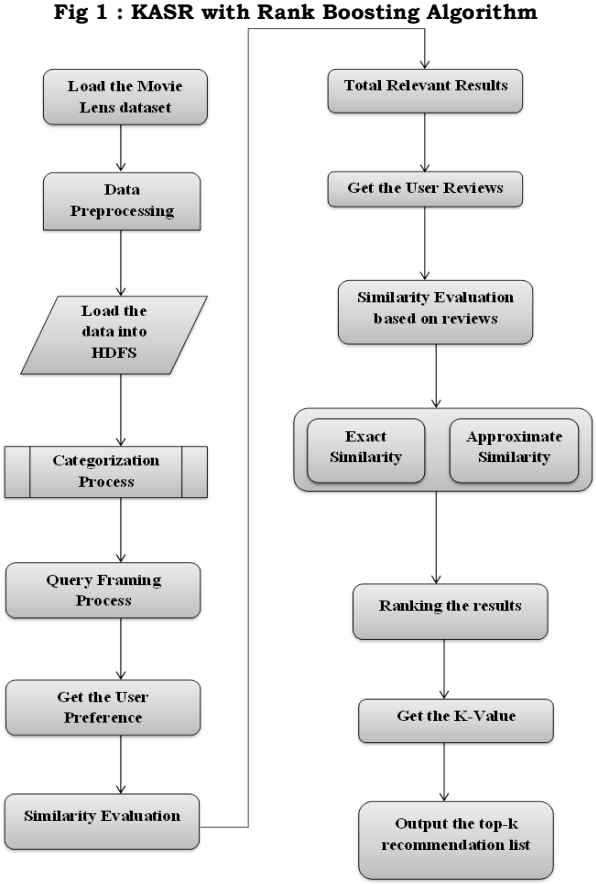

Fig.2 Flow of the proposed system

HDFS upload and categorization

After preprocessing, the modified datasets will be uploaded to HDFS (Hadoop Distributed File System), which is a highly fault tolerant distributed file system designed to run commodity hardware. HDFS will provide high throughput access to application data and provides efficient processing of large datasets. Categorization process (discuss later in the section experimental results) classify the movies according to its categories like Horror, Animation, Children, and Comedy etc. Movies belonging to each category will be identified and will be arranged according to the category to which they belong.

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 (Hoadoop). The process is performed by splitting the preferences i.e.it is done using mapper process. After the processing, the results are aggregated. The similarity analysis is carried out on this aggregated data which is explained in the next section.

Similarity measurement

Measure the similarity between the preference of the current user and that of previous users who have already given their reviews and ratings for movies.Two methods are used to calculating the similarity Aproximate similarity computation method and Exact similarity computation method.

a) Approximate similarity Measurement Jaccard coefficient is used to measure the Approximate similarity. Jaccard coefficient is used calculate the Similarities between the preferences of the current and previous users are computed with the help of following equation.

Sim(APK,PPK)=Jaccard(APK,PPK)=

$$\frac{|APK \cap PPK|}{|APK \cup PPK|}$$

b) Exact Similarity Measurement Accurate similarity measurement uses a cosine based approach. In this approach the preference of the current and previous users will be transformed into n-dimensional weight vector.

Prediction of recommendation list.

After mapreduce process execution and similarity calculation, 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.

IV. Algorithm Descriptions

Parameter used in the algorithm are as follows
APK = Preference keyword of the active user
PPK_j = The preference keyword set of a previous user
WS = Candidate services.
$\hat{R}$ = Used to store the remaining preference keyword sets of previous users.
K = Number given by the user
$\bar{r}$ = average ratings of the candidate service.

Algorithm 1 KASR

Input: The preference keywords of the current user APK
The candidate services
The threshold δ in the filtering phase
The number k
Output: The services with top-k highest ratings

1:For each service $ws_i \in WS$
2: $\hat{R} = \Phi$, sum=0,r=0
3: For each review R_j of service ws_i
4: Process the review into preference keyword set PPK_j
5: If $PPK_j \cap APK = \Phi$ then
6: Insert PPK_j into $\hat{R}$
7: End if
8: End for
9: For each keyword set $PPK_j \in \hat{R}$
10: $Sim(APK,PPK_j) = SIM(APK,PPK_j)$
11: If $SIM(APK,PPK_j) < \delta$ then
12: Remove PPK_j from $\hat{R}$
13: Else sum=sum+1 , r= r + r_j
14: End if
15: End for
16: $\bar{r}$ = r/sum
17: get pr_i by formula
18:end for
19:sort the service according to the personalized ratings pr_i
20:return the top-K services with highest ratings

Algorithm 2 RankBoost

Given: Initial distribution D over x × x.
Initialize : $D_1 = D$.
For t = 1,….,T:

- Train weak learner using distribution D_t .
- Get weak rating $h_t : x \rightarrow R$.
- Choose $\alpha_t \in R$.
- Update $D_{t=1(x0\ ,\ x1)} =$

$$\frac{Dt(\chi 0,\chi 1)\exp\left(\alpha t\left(ht(\chi 0)-ht(\chi 1)\right)\right)}{Zt}$$

Where Zt is a normalization factor

Output of the final ranking $H(x)=\sum_{t=1}^{T} \alpha t$ ht(x)

Rank Boost maintains a distribution Dt.Rank Boost chooses Dt to emphasize different parts of the training data. A high weight assigned to a pair of instances indicates a great importance that the weak learner order that pair correctly. The boosting algorithm uses the weak rankings to update the distribution. The final ranking H is a weighted sum of the weak rankings. A ranking feature is nothing more than an ordering of the instances from most preferred to least preferred. Zt normalization factor. fi as a scoring function where higher scores are assigned to more preferred instances.

V. Conclusion

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 Map Reduce framework in Hadoop platform. Finally, the experimental results demonstrate that combined preferences based on rank boosting algorithm gives the better results than KASR; it also significantly improves the accuracy and scalability of service recommender systems over existing approaches.

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 of Netnews," In CSCW '94 Proceedings of the 1994 ACM conference on Computer supported cooperative work, pp. 175-186, 1994.

[12] R. Burke, "Hybrid Recommender Systems: Survey and Experi-ments," User Modeling and User-Adapted Interaction, Vol. 12, No.4, pp. 331-370, 2002.

[13] G. Adomavicius, and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the Stateof- the-Art and Possi-ble Extensions," IEEE Transactions on Knowledge and Data Engi-neering, Vol.17, No.6 pp. 734-749, 2005.

[14] D. Agrawal, S. Das, A. El Abbadi, "Big Data and cloud computing: new wine or just new bottles?" Proceedings of the VLDB Endow-ment, Vol. 3, No.1, pp. 1647-1648, 2010.

[15] J. Dean, and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," Communications of the ACM, Vol. 51, No.1, pp. 107-113, 2005.