© 2017 IJEDR | Volume 5, Issue 3 | ISSN: 2321-9939

Smart Service Recommendation System by Applying MapReduce Techniques on Big Data

[1]Pallavi R. Desai, [2]Amol R. Dhakne
[1] Student ME (Computer),
[2]Assistant Prof. & HOD of Computer Engg.[1]Department of Computer Engineering

***Abstract*—Now days, the amount of data in internet grown beyond the size of processing, this is Big Data. In the last few years, the information is growing vast, and separating this wide information, yielding the big data and analysis is major problem for recommender systems. Consequently, most of the traditional recommender systems frequently suffer from scalability and lack of accuracy problem and fails to meet personalized requirements. The purpose of Smart Service Recommender system is providing appropriate recommendations to users as per their interest and gives a recommendation list and recommending the most right items to the users. To improve its scalability, it is executed on Hadoop with MapReduce and Filtering algorithm is adopted to generate recommendations. Finally, general experiments are conducted on real-world data sets of movielens, and results demonstrate that Smart Service Recommendation System expressively recovers the accuracy and scalability of service recommender systems over existing approaches.**

***Index Terms*—Big Data, MapReduce, Hadoop, recommender system, preference, keyword.**

I. INTRODUCTION

The amount of structured, semi-structured and unstructured data has been growing very quickly in our world. These data is nothing but bid data. Study of big data is challenges for companies. To solve these challenges of big data we want to provide hardware and software solutions. With the increasing of data challenges may also increases. Every day, people are come with many options and preferences. Which item to buy? What to purchase? Where to stay? Which blog post to read? Which place gives better service? And so on. Each of every question has many different solutions. With the growing number of alternative services, and for these types of big data want to providing an effectively recommendation system that users can get that they want. Service recommender systems are one of the parts of information filtering systems that try to find to calculate the rating and preferences that a user would give an item. Recommender systems have been shown as important tools to help users deal with services overload and provide proper recommendations to them. Smart service recommendation system is an important application in the development of information processing technology. With the continuous quick growth and development of Internet technology, there has been an explosive growth of information on the Internet. Even if widely available user search engines have become the most effective to search the information on Internet, these tools fail to satisfy all user needs and preferences. Also, if we consider user rating then also customer was unable to get perfect information. Consequently, Smart Service recommendation system has been developed. Smart service recommendation system provides an automatic function that gives recommendation details by obtaining and examining user preferences, calculations based on the analysis and information are made prior to the user launching a search. To achieve the perfect recommendations we apply Jaccard coefficient for Similarity calculation of keywords. The core value of Smart Service recommendation system lies in its recommendation capability. The appropriate use of recommendation system algorithms that improve the accuracy and scalability of recommendations and algorithms those return results consistent with user interests.

.

II. RELATED WORK

The authors Shunmei Meng,Wanchun[1] proposes a Keyword-Aware service Recommendation method, to presenting a personalized recommendation list and recommending the most appropriate items to the users effectively. Specifically, keywords are used to indicate users' preferences, and a user-based Collaborative Filtering algorithm is adopted to generate appropriate recommendations. Finally, the experimental results demonstrate that KASR significantly improves the accuracy and scalability of service recommender systems over existing approaches.

The authors, X. Yang, Y. Guo, and Y. Liu [4], propose a Bayesian-inference-based recommendation system is developed for online social networks. They show that the proposed Bayesian-inference-based recommendation is better and comparable than the existing trust-based recommendations. They are also manages the recommendation quality and quantity. The Prior Distribution is specially used to overcome cold start problem and data sparseness problem.

In [9], Adomavicius and Tuzhilin give an overall of the structure of recommender systems and describe the current generation of recommendation methods. They also describe various drawback and limitations of current service recommendation methods, and discuss possible solutions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications.

Most existing service recommender systems are only use a single numerical rating to represent a services utility recommendation as a whole [4]. They are not considering the user's' preferences. In fact, analyzing services through multiple criteria, choices and considering users feedback can help to make more accurate and effective recommendations for the users.

The authors Z D Zhao and M. S. Shang of [7] implement a user-based CF algorithm on Hadoop. They solve the scalability and efficiency problem by dividing dataset into different domains. But their method doesn't have favourable of scalability and efficiency if the amount of data grows beyond the capacity. It proposes a parallel user profiling approach based on folksonomy information and implements a scalable recommender system by using MapReduce and Cascading techniques.

M. Hu, H. Singh, D. Rule, M. Berlyant, and Z. Xie Y. Jin [6] presenting a large scale video recommendation system implemented by using item-based CF algorithm. They implement their proposed approach in Qizmt, which is a .Net MapReduce framework, thus their system can work for large scale video sites.

The authors [2] proposed a trust-aware system for generating personalized user recommendations in social networks. Its foundations lie on a reputation mechanism that is mathematically formulated, comprising both local and collaborative rating formation. The proposed system provides users with personalized positive and/or negative recommendations that can be used to establish new trust/distrust connections in the social network.

The author [3] proposed location-aware recommender system, solve a problem untouched by traditional recommender systems by dealing with three types of location-based ratings: spatial ratings for non-spatial items, non-spatial ratings for spatial items, and spatial ratings for spatial items. All above techniques can be applied separately or in concert to support the various types of location-based ratings. It is efficient, scalable, and provides better quality recommendations than techniques used in traditional recommender systems.

III. Existing System

In the existing system they have proposed a keyword-aware service recommendation method, named KASR. In this, keywords are used to indicate users' preferences, and a user based Collaborative Filtering algorithm is adopted to make appropriate recommendations. Firstly they create a keyword- candidate list and domain thesaurus by using preferences of active as well as previous users and is 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 according to the keyword-candidate list and domain thesaurus. This system aims at presenting a personalized service recommendation list and recommending the most appropriate service(s) to the users. As compared with previous systems, these systems are failed to provide better recommendation to any particular item. That means traditional recommendation system has the problem of scalability and efficiency. And to improve the scalability and efficiency of KASR in "Big Data" environment, they have implemented it on a MapReduce framework in Hadoop platform.

IV. Overview of System

The project proposes a different method of Smart service for recommendation system by applying MapReduce technique with Hadoop. In which Keyword-preference List and Domain Clusters are maintained for particular system. Preferences are taken from reviews of past users. And match between users preference are searched out by keyword extraction method and similarity calculations by applying smart service approach. Then the keywords are classified according their weight and weights of reviews of similar users' preferences are then calculated.

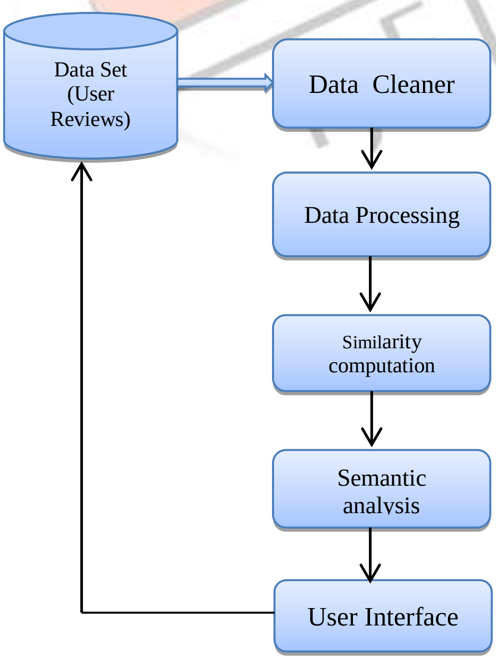

Figure 1: System Architecture of Smart Service Recommendation System

Then finally, recommendation list of top-k items is generated. In proposed model, keywords are used to show both of users' preferences and the quality of recommendation services. A Smart Service algorithm is assumed to give proper recommendations. Proposed system aims at calculating rating of each recommendation services for a user, and then presenting a recommendation list and recommending the most right services to them.

1. Capture user preferences by a keyword-aware approach
2. Similarity computation
3. Calculate personalized ratings and generate recommendations

a) Approximate similarity computation: A frequently used method for comparing the similarity and diversity of sample sets, Jaccard coefficient, is applied in the approximate similarity computation.
b) Exact similarity computation: A cosine-based approach is applied in the exact similarity computation, which is similar to the Vector Space Model (VSM) in information retrieval
c) Keyword aware service recommendation system: To calculate the similarities between active and previous user is used.

V. METHODOLOGY

Keyword- preference List and Domain Clusters: In this technique, two data structures, keyword preference list and specialized domain clusters, are introduced to obtain users' preferences. Keyword-preference list: The keyword preference list is a collection or set of keywords about users preferences and multi-criteria of the user services, which can be denoted as K=n where n is the number of the keywords in the keyword preference list. Domain clusters: A particular domain clusters are built to support the keyword retrieving from reviews of old users. A domain clusters is a group of reference of the keyword-preference list that lists words grouped together according to the similarity of keyword , including similar and different words and antonyms.

Keyword Preprocess: In this step, the preferences of current user and old users are grouped into their equivalent preference keyword sets respectively. First, HTML tags and stop words from the reviews keyword sets should be removed to avoid the quality of the extraction process of keyword in the next operation. And the Stemmer algorithm is used to reducing the keywords to get actual keyword or root. The ratio of preferences/choices of previous users and preference/choice of current user are calculated to observe accuracy in keyword-preference list. Its main use is as part of a term normalization process that is usually done when setting up Information Retrieval systems.

Keyword Extraction: In this stage, each review will be transformed into a corresponding keyword set according to the keyword-preference list and domain clusters. Each one keyword will be changed into a comparing keyword set by pivotal keyword applicant rundown and domain clusters. If the review contains a word in the domain clusters, then the corresponding keyword should be extracted into the preference keyword set of the user. For example, if the past user gives the review for a hotel. And it has the word Stop, which is corresponding to the keyword Transportation in the domain clusters, then keyword Transportation should be contained in the preference keyword set of the old user. The list contains film which referencing animation as well as action. If a keyword found many times more than once in reviews, the times of repetitions will be recorded. In this method, it is regarded that keywords appearing multiple times are more important. The number of recurrences will be used to calculate as the weight of the keyword in preference keyword set.

Similarity Calculation: The next step is to find the reviews of old users who have similar tastes to a current user by finding neighborhoods of the current user based on the similarity of their preference of keywords. Before similarity computation, the reviews separate to the current user's preferences will be filtered out. To this the intersection concept in set theory is used. If the intersection of the word preference sets of the current user and old user is an empty set, then the preference keyword set of the past user will be filtered out. Two similarity computation methods are introduced in our recommendation method:

1. Approximate similarity computation method.

2. Exact similarity computation method.

1) *Approximate Similarity Computation:*

The approximate similarity computation method is appropriate for to compute similarity between the preferences of the current user and old user for the case that the weights of the keywords in the preference keyword set are not used. It is often used method for comparing the similarity and variety of sample sets. The Jaccard coefficient is measurement of asymmetric information on binary (and non-binary) variables, and it is useful when non positive values give to information.

Jaccard coefficient is described as follows:

$$sim(AK, PK) = Jaccard(AK, PK) = \frac{|AK \cap PK|}{|AK \cup PK|} \qquad (1)$$

Where AK is the preference keyword set of the active user, PK is the preference keyword set of a previous user.

2) *Exact similarity computation:* The exact similarity computation method is appropriate for to compute similarity between the preferences of the active user and a previous user for the case that the weights of the keywords are available. A cosine-based approach is applied in the exact similarity computation, which is similar to the vector model in information retrieval.

Calculate the weight by the following function:

$$WKi = \frac{1}{n}\sum_{j=1}^{n}\frac{aij}{\sum_{k=1}^{n} akj} \qquad (2)$$

B. *Generate Personalized Recommendation List :* Based on the similarity of the current user and old users, further filtering will be conducted. Given a threshold δ, of sim(AK, PKj) < α, , the preference keyword set of a previous user PKj will be filtered out, otherwise PKj will be retained. The alpha given in two similarity calculation methods are different, which are both empirical values. Once the set of most similar users are found, the personalized ratings of each user service for the current user can be calculated. Finally, service recommendation list will be presented to the user and the service with the top rating will be recommended to users.

Algorithm of Smart Service Recommendation System

Input : The preference rating set of the active user AP K

The rating set RS = $\{rs_1; rs_2; \ldots; rs_N\}$

The Threshold δ in the filtering phase

The number Top-K

Output: The service with the Top-k highest rating $\{tws_1; tws_2; \ldots\ldots; tws_k\}$

```
for each rating RSi Є RS
R = ø , sum = 0, r =0
for each rating Rj of services rsi
 process the rating into a previous preference keyword set PPKj
if PPKj ∩ AP K 6=   Then
insert PPKj into R
end if
end for
for each keyword set PPKj Є R
SIM(APK; PPKj)=SIM (APK; PPKj)
SIM (APK; PPKj) can be SIM-ASC (APK; PPKj) or SIM-ESC (APK; PPKj)
if sim(APK; PPKj) < δ  then
remove PPKj from R
else sum = sum + 1 ,r = r + rj
end if
end for
r = r/sum
get prj
end for
sort the services according to the personalized ratings prj
return the services with the Top-K highest ratings {tws1; tws2; …… twsk}
```

VI. Experimental Result

In this section, experiments are designed and analysed to evaluate the accuracy and scalability of purposed personalize user-based recommendation system (SSSR). To evaluate the performance of purposed recommendation system (SSSR) in accuracy, SSSR compare with other. Three well-known recommendation methods:

- Keyword aware service recommendation system (KASR)
- User-based algorithm using Pearson Correlation Coefficient (UPCC)
- Item-based algorithm using Pearson Correlation Coefficient (IPCC)

Three metrics are used to evaluate the accuracy:

1. Mean absolute error (MAE)
2. Mean average precision (MAP)
3. Discounted cumulative gain (DCG).

1. COMPARISON OF SSSR WITH UPCC, IPCC AND KASR IN MAE

Fig. 2 shows the MAE values of SSSR, UPCC, IPCC and KASR. It could be found that the MAE value of proposed system (SSSR) is much lower than KASR, UPCC and IPCC. Thus proposed methods SSSR can provide more accurate predictions than the all of them.

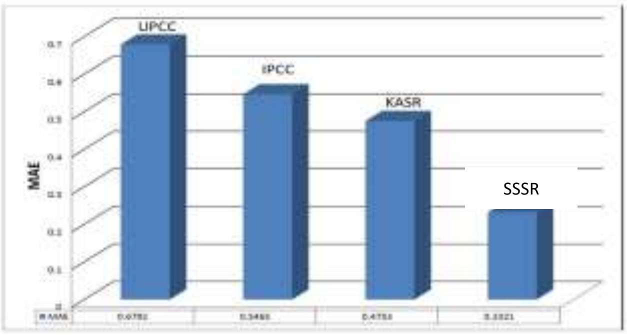

Figure 2: Comparison of UPCC, IPCC, KASR and SSSR in MAE.

2. COMPARISON OF SSSR WITH KASR, UPCC AND IPCC IN MAP AND DCG.

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.
Figs. 5.9 and 5.10, respectively show the DCG values and MAP values of Top-K (K = 30, 50, 70) recommendation list of SSSR, KASR, UPCC and IPCC.

From Figs. 3 and 4., it could be seen that the DCG values and MAP values of proposed recommendation system (SSSR) is comparatively higher than all of them.

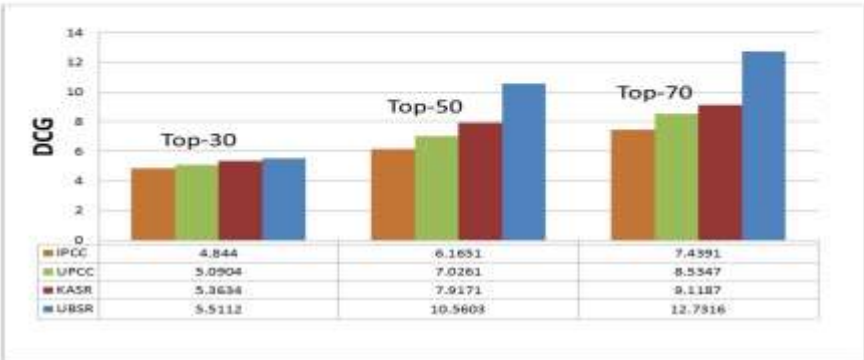

Figure 3: Comparison of UPCC, IPCC, KASR and SSSR in the DCG values of Top-K (K = 30, 50, 70) recommendation list.

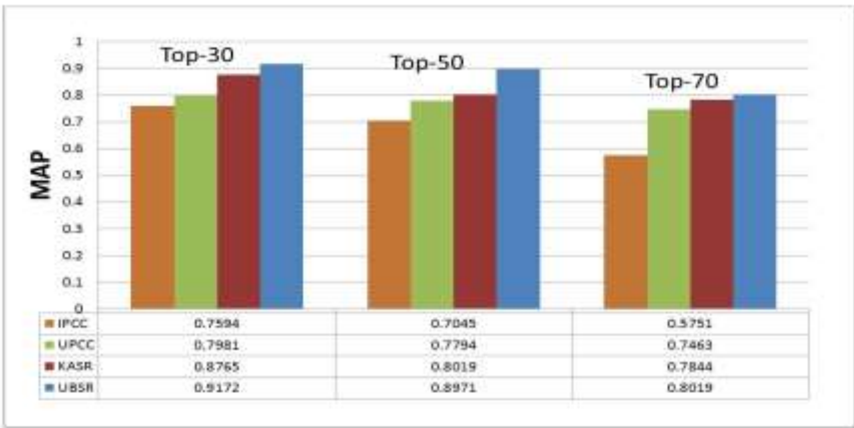

Figure 4: Comparison of UPCC, IPCC, KASR and SSSR in the MAP values of Top-K (K = 30, 50, 70) recommendation list.

It also could be found that the DCG values increase when Top-K increases, while the MAP values decrease when Top-K increases.

VII. ACKNOWLEDGMENT

I hereby take this opportunity to record my sincere thanks and heartily gratitude to Guide Prof. Shweta A. Joshi useful guidance and making available to me her intimate knowledge and experience in making **" Smart Service Recommendation System by applying MapReduce Technique on BigData"** as a project and i express my sincere gratitude towards people who have worked in this area.

VIII. CONCLUSION

The system titled "Smart Service Recommendation System by applying MapReduce Technique" overcomes the limitations of the traditional filtering algorithm such as scalability and poor accuracy .Moreover, to improve the scalability and accuracy in "Big Data" environment, this system executed it on a MapReduce framework in Hadoop platform.. The proposed system is more effective in terms of complexity. And the system gives more accurate results or recommendations to the users. Finally, the results express that SSSR extensively improves the accuracy and scalability of recommender systems than existing systems. Comparing with existing system, SSSR generate more accurate recommendation in any case or of any item.

REFERENCES

[1] Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen, Senior Member, IEEE KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 12, DECEMBER 2014.

[2] Magdalini Eirinaki, Malamati D. Louta, Member, IEEE, and Iraklis Varlamis, Member,IEEE "A Trust-Aware System for Personalized User Recommendations in Social Networks"IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 44, NO. 4, APRIL 2014 409.

[3] Sarwat, Justin J. Levandoski, Ahmed Eldawy, and Mohamed F. Mokbel IEEE: An Efficient and Scalable Location-Aware Recommender System Mohamed TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 6, JUNE 2014.

[4] X. Yang, Y. Guo, and Y. Liu, Bayesian-Inference Based Recommendation in Online Social Networks, IEEE TRANS. PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO.4, PP. 642-651, APR. 2013.

[5] An Effective Recommendation Framework for Personal Learning Environments Using a Learner Preference Tree and a GA Mojtaba Salehi, Isa Nakhai Kamalabadi, and Mohammad B. Ghaznavi Ghoushchi, Member, IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES,VOL. 6, NO. 4, OCTOBER-DECEMBER,2013.

[6] M. Hu, H. Singh, D. Rule, M. Berlyant, and Z. Xie Y. Jin,"MySpace Video Recommendation with Map-Reduce on Qizmt," Proeedings of the 2010 IEEE Fourth International Conference on Semantic Computing, pp. 126-133, 2010.

[7] Z D Zhao and M. S. Shang, "User Based Collaborative Filtering Recommendation Algorithms on Hadoop," In the third International Workshop on Knowledge Discovery and Data Mining, pp. 478-481, 2010.

[8] G. Adomavicius and Y. Kwon, New Recommendation Techniques for Multicriteria Rating Systems, IEEE Intelligent Systems, vol. 22, no. 3, pp. 48-55, May/June 2007.

[9] A. Tuzhilin and G. Adomavicius, "Toward the Next Generation of Recommender Systems: A Survey of the State of the Art and Possible Extensions," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 6, PP. 734-749, 2005.

[10] J. Manyika, M. Chui, B. Brown, et al, "Big Data: The next frontier for innovation, competition, and productivity," 2011.

[11] C. Lynch, "Big Data: How do your data grow?" Nature, Vol. 455, No. 7209, pp. 28-29, 2008.

[12] Y. Chen, A. Cheng, and W. Hsu, "Travel Recommendation by Mining People Attributes and Travel Group Types from Community- Contributed Photos," IEEE Trans. Multimedia, vol. 25, no. 6,pp. 1283-1295, Oct. 2013.

[13] C. Lynch, "Big Data: How do your data grow?" Nature, vol. 455, no. 7209, pp. 28-29, 2008.

[14] S. Ghemawat, H. Gobioff, and S. T. Leung, "The Google File System," The 19th ACM Symposium on Operating Systems Principles,pp.29-43,2003.

[15] L.Zhang, "Editorial: Big Services Era: Global Trends of Cloud Computing and Big Data". IEEE Transactions on Services Computing, Vol. 5, No. 4, pp. 467-468, 2012.

[16] Y. Pan and L. Lee, "Performance Analysis for Lattice-Based Speech Indexing Approaches Using Words and Subword Units," IEEE Trans. Audio, Speech, and Language Processing, vol. 18, no. 6, pp. 1562-1574, Aug. 2010.

[17] G. Kang, J. Liu, M. Tang, X. Liu, and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," Proc. IEEE 19th Int'l Conf. Web Services (ICWS), pp. 186-193, 2012.