

# KEYWORD-BASED SERVICE RECOMMENDATION SYSTEM USING MAPREDUCE FOR BIG DATA

Mr. M.P.Gaikwad<sup>1</sup>      Mr. C.S.Arage<sup>2</sup>  
Mr. Rahul Alagoudar<sup>3</sup>      Mr. Prem Khatri<sup>4</sup>

Department of Computer Science and Engineering

Sou. SushilaDanchandGhodawat Charitable Trust's Sanjay Ghodawat Group of Institutions Maharashtra

<sup>1</sup>[gaikwad.mp@sginstitute.in](mailto:gaikwad.mp@sginstitute.in)

<sup>2</sup>[Arage.cs@sginstitute.in](mailto:Arage.cs@sginstitute.in)

<sup>3</sup>[alagoudarr@gmail.com](mailto:alagoudarr@gmail.com)

<sup>4</sup>[khatriprem1000@gmail.com](mailto:khatriprem1000@gmail.com)

*Abstract - To offers suitable recommendation to the user, service recommender system is a important means. From previous few years, increased number of customer, services and online information has grown widely, so that big data analysis problem has been occurred for service recommendation system. The old-style service recommender systems often suffer from scalability and inefficiency problems when processing or analyzing such large-scale data. The present service recommender systems fails to meet users 'personalized requirements' because of their presence of the similar ratings and rankings of services to dissimilar users without considering diverse users' preferences.*

*In this paper, we propose a Keyword-Based Service Recommendation system, to resolve the above challenges. It delivers a personalized service by consideration of not only recommendation list but also recommending the most suitable services to the users effectively. To generate suitable recommendations, user's preference as well as user-based Collaborative Filtering algorithm is adopted.*

**Key Words:** Big data, big table, keywords, recommendation system, thesaurus, user's preferences

## 1. INTRODUCTION

The quantity of data in our world has been growing explosively and analyzing large data sets called "Big Data", become a key basis of rivalry underpinning new waves of productivity growth, innovation, and consumer surplus[1]. Big data is a broad term for data sets, so large or complex that traditional data processing applications are insufficient. Challenges include analysis, capture, data creation, search,

sharing, storing, transfer, visualization, and querying and information privacy. Today, Big Data management stands out as a challenge for IT industry companies. The solution to such a challenge is flowing increasingly from providing hardware to provisioning more manageable software solutions [2]. Big data also brings new opportunities and serious challenges to industry and academia [3] [4]. Similar to most big data applications, the big data tendency also poses hefty impacts on service recommender systems. With the increasing number of other services, effectively recommending services that users preferred have become an important research issue. Service recommender systems have been shown as valued tools to help users deal with services overload and offers appropriate recommendations to them. Example of such practical applications include CDs, books, web pages and various other products now use recommender systems[5][6][7]. Over the last decade, there has been much research done both in industry and academic world on developing new approaches for service recommender systems.

To offers suitable recommendation to the user, service recommender system is a valuable tool. From previous few years, increased number of customer, services and online information has grown extensively, so that big data analysis problem has been occurred for service recommendation system. The traditional service recommender systems often suffer from scalability and inefficiency problems when processing or analyzing such large-scale data. The present service recommender systems fails to meet users 'personalized requirements' because there is presence of the alike ratings and rankings of services to different users without considering diverse users' preferences. Motivated by these observations, in this paper, we address these challenges through the following contributions:

- A keyword-based service recommendation method is proposed in this paper, which is based on a user-

based Collaborative Filtering (CF) algorithm.

- In this, keywords extracted from reviews of previous users are used to point out their preferences [13].

Recommendations are offered as ranked list of things. In performing this ranking, recommender systems try to predict what the most appropriate products or services are, based on the user's preferences and constraints.

## 2. LITERATURE REVIEW

While processing or analyzing huge scale data the service recommendation system suffer from the scalability and inefficiency. The present existing recommender system offers rating and ranking of services without bothering of various users' preference list .Because of this reason they are failed to meet users personalized necessity. Moreover, in most existing service recommender systems, such as hotel reservation systems and hotel guides, the ratings of services and the service recommendation lists offered to users are the same. They have not considered users' diverse preferences, without meeting users' personalized requirements.

Recommendation carried out by different ways:

### 2.1 Content-based approaches.

Content based recommenders offer recommendations by comparing depiction of contents describing an item or a product to the representation of the content describing the interest of the user. They are sometimes denoted to as content based filtering [9]. Content based recommendation system is nothing but to get from the content of element. In most of the cases they are the words which depict the element. This recommendation system is able to introduce new element for the user. It recommending item along with explanation of element to the user. The main slogan of this method is to analyze the group of document of items, which is rated by the user and construct a model according to the user interest which is based on feature of the items. Information recovery and information filtering both are the sources of content based recommending approach. It offers only written information like as news webs and documents. No content-based recommendation system can offers appropriate suggestions if the analyzed content does not contain enough information to discriminate items the user likes from items the user does not like[9]. Some illustrations acquire only certain views of the content, but there are many others that would manipulate a user's experience.

### 2.2 Collaborative recommendation approaches.

Collaborative filtering approaches recommend services [5] is well known Favored by user in the past. Collaborative filtering method is most vital method in recommendation system, in which it finds a set of people who shares similar interest with you. These people could be

determined by the similar ranking on things. These people recognized by neighborhood of the current user. Collaborative filtering use's "Netflix". If any recommendation system gives rating on item then it might be using collaborative filtering method. First problem is that result of recommendation system will be hinge on presence of information. In the relationship mining, new stuffs not-yet-rated can be abandoned in the recommendation processes. Second difficulty is that, the collaborative filtering unable to cover the extreme case. Similarity decisions are unable to be established when the scales of the users are small or the users have unique taste. Update regularity is the third difficulty. If any new information of users has to be included in the recommendation processes in real time, data latency will rise the waiting time for the query result [5].

### 2.3 Big data.

Big data is nothing but one kind of the data set whose size is superior to that of the typical database which is used to capture, store, manage and analyze the data. Sweating of data is carried out in every industry and business so that it becomes significant factor of production. Now days, the use of big data will be responsible for competition and increase of individual firm [10]. Due to extensive use of digital technology, there is rapidly increase in the digital data in every organization .Digital network hold the large number of users, devices and sensors which are interconnected to each other and as time passes they are continuously increasing so that ability to produce, communicate, share and access the data has been revolutionized.

### 2.4 Big Table

Big table is a distributed storage system for running structured data that is designed to scale to a very large size. Many projects at Google, store data in big table, as well as web indexing, Google Earth, and Google Finance [8]. Data in the table is organized in three dimensions which is row, column and timestamp.

## 3. METHODOLOGY

### 3.1 Introduction

In Keyword- based service recommendation method, key words are used to point out both of users' preferences and the quality of candidate services. A user-based CF algorithm is adopted to generate suitable recommendations.

It aims at calculating a custom-made rating of each candidate service for a user, and then presenting a personalized service recommendation list and recommending the most appropriate services to user. Just think service as hotel reservation system. As shown in Fig 1 while reserving any hotel we are considering so many things like room service, quality of food, cleanliness of hotel and its environment, etc. The necessities of hotels are depending on customer, as person changes their requirement also changes. Some people more concern about cleanliness of hotel not bothering about value, transport facilities. But on the other hand some people concern in value. Rectangles contain main keywords and oval contain related words of that keyword.

The user may put the requisite as “he/she want mall near to hotel” instead of “shopping”. So that, we are applying domain treasures on the keyword. Because of that we can easily get the related words of keyword which will be our candidate list1.

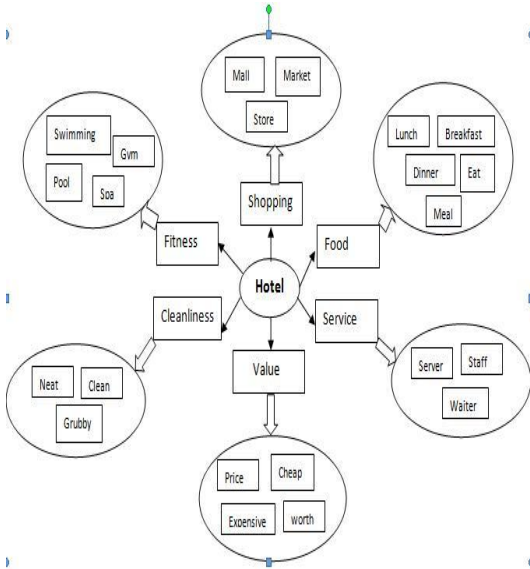


Fig -1: Hotel System

### 3.1.1 Keyword-candidate list:

The keyword-candidate list is a set of keywords about users’ preferences and multicriteria of the candidate services, which can be denoted as  $K = \{k_1; k_2; \dots; k_n\}$ ;  $n$  is the number of the keywords in the keyword-candidate list. An example of a simple keyword-candidate list of the hotel reservation system is described. Keywords in the keyword-candidate list can be a word or multiple words related with the quality criteria of candidate services [13].

### 3.1.2 Domain thesaurus:

A domain thesaurus is a reference work of the keyword-candidate list that lists words grouped together according to the similarity of keyword meaning, including related and contrasting words and antonyms[12][13]. An example of a simple domain thesaurus of hotel reservation system is shown in fig-1. The words in the rectangle connected to circle named Hotel are the keywords in the corresponding keyword-candidate list, and the words in the ovals are the related words of the keywords. Often, domain thesauruses are updated regularly to ensure the timeliness of the words.

### 3.2 Proposed Work

In this project, we propose a keyword-based service recommendation method. In this, keywords are used to indicate users’ preferences. Collaborative Filtering algorithm is adopted to produce appropriate recommendations. More specifically, a keyword- candidate list and domain thesaurus is offered to help obtain users’ preferences. The active user gives his/her preferences by selecting the keywords from the keyword candidate list, and the preference of the previous users can be extracted from

their reviews for services according to the keyword-candidate listing and domain thesaurus. Our method aims at presenting a personalized service recommendation list and recommending the most appropriate service to the users. Finally, the experimental results demonstrate that KBSR significantly improves the accuracy and scalability of service recommender systems over existing approaches.

### 3.3 Proposed Methodology

#### 3.3.1 Capture user preferences by a keyword-aware approach.

In this step, the preferences of active users and previous users are formalized into their consequent preference keyword sets respectively. In this project, an active user refers to a current user needs reference.

**Preferences of an active user:** An active user can give his/her preferences about candidate services by selecting keywords from a keyword-candidate list, which mirror the quality criteria of the services he/she is concerned about.

**Preferences of previous users:** The preferences of previous user for a candidate service are extracted from his/her reviews for the service according to the keyword-candidate list and domain thesaurus. And a review of the previous user will be formalized into the preference keyword set of him/her.

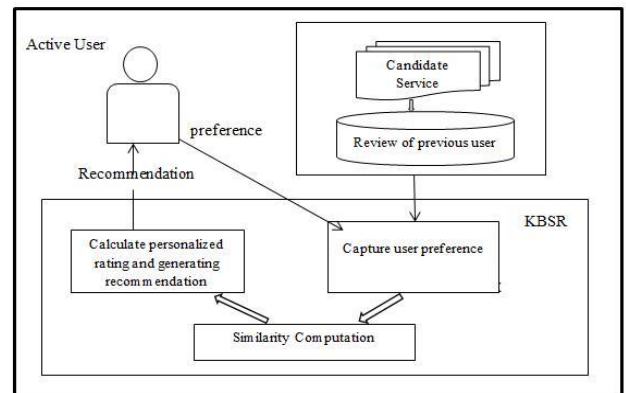


Fig -2: Main steps of KBSR System

The main steps of KBSR are depicted in Fig. 2, which are described in detail as follows: Our system offers candidate keyword list, so current users can select keyword like food, transport, hospitality etc. from given preference list as shown in Fig 2. Keyword selected by active user will go to the text area which will consider as active user preference list.

#### 3.3.2 Similarity computation.

The second step is to identify the reviews of previous users who have similar tastes to an active user by finding neighborhoods of the active user based on the resemblance of their preferences. Before similarity computation, the reviews unrelated to the active user’s preferences will be

filtered out by the intersection concept in set theory. We are using word net dictionary for finding synonymous of keyword which are presenting active user preference list. If the intersection of the preference keyword sets of the active user and a previous user is an blank set, then the preference keyword set of the previous user will be filtered out [14].

### 3.3.3 Calculate personalized ratings and generate recommendations.

Based on the resemblance of the active user and previous users, further filtering will be directed. Once the set of most similar users are found, the personalized ratings of each candidate service for the active user can be calculated. Finally, a personalized service recommendation list will be presented to the user and the service(s) with the highest rating(s) will be recommended to him/her [14]. .

## 4. CONCLUSIONS

We have proposed a keyword-based service recommendation method, named KBSR. In KBSR, keywords are used to indicate users' preferences, and a user based Collaborative Filtering algorithm is adopted to produce appropriate recommendations. More specifically, a keyword-candidate list and domain thesaurus are offered to help obtain users' preferences. The active user offers 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 the domain thesaurus. Our method aims at presenting a personalized service recommendation list and recommending the most appropriate service to the users.

## 5. REFERENCES

- [1] J. Manyika, M. Chui, B. Brown, et al, "Big Data: The next frontier for innovation, competition, and productivity," 2011. 1045-9219 (c) Citation information: DOI 10.1109/TPDS.2013.2297117, IEEE Transactions on Parallel and Distributed Systems 10.
- [2] C. Lynch, "Big Data: How do your data grow?" Nature, Vol. 455, No. 7209, pp. 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] MR.W. Dou, X. Zhang, J. Liu, J. Chen, "HireSome-II: Towards PrivacyAware Cross-Cloud Service Composition for Big Data Applications," IEEE Transactions on the Parallel and the Distributed Systems, 2013.
- [5] MR.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 the User Modeling," IEEE Transactions on Consumer Electronics, Vol. 56, No.3, pp. 1763-1769, 2010.
- [7] MR.M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for the Sport Videos Based on User Audiovisual Consumption " IEEE Transactions on Multimedia, Vol. 14, No.6, pp. 1546-1557, 2013.
- [8] MR.Y. Chen, A. Cheng and W. Hsu, "Travel Recommendation by Mining People Attributes and the Travel Group Types From CommunityContributed 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 Prediction for Cloud Services," IEEE Transactions on the Parallel and the Distributed Systems, Vol. 24, No. 6, pp. 1213-1222, 2013.
- [10] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, for "Recommending and Evaluating Choices in a Virtual Community of Use," In CHI '95 Proceedings of the SIGCHI Conference on the Human Factors in Computing 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 supportedcooperative work, pp. 175-186, 1994.
- [12] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," 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 Possible Extensions," IEEE Transactions on Knowledge and Data Engineering, 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 Endowment, 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.