International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Keyword based Service Recommendation System

Ms.Reshma Nikam[1], Mr.Ranjit Mandale[2], Mr.Arbaj Mulla[3], Prof.Mr. V.B.Ingale[4]

[1]Student of CSE, in Adarsh Institute of Technology And Research Center ,vita, Maharashtra, India
[2]Student of CSE, in Adarsh Institute of Technology And Research Center ,vita, Maharashtra, India
[3]Student of CSE, in Adarsh Institute of Technology And Research Center ,vita, Maharashtra, India
[4]Assistant Professor, Dept. of Computer Science and Engineering, Adarsh Institute of Technology And Research Center, vita, Maharashtra, India

---

Abstract - *Big data refers to traditional enterprise data (includes customer information, web store transactions etc.), machine-generated /sensor data (includes call detail records, weblogs, equipment logs etc.) and social data (includes customer feedback streams, micro-blogging sites like Twitter, social media platforms like Facebook) [1]. The big data frameworks include the Google MapReduce, Hadoop , Twister, Hadoop++ and Spark etc. which are used to process big data[2]. As the number of users and their online information has been increasing rapidly, it produces the big data analysis problem for recommender systems.*

Recommender systems are commonly used techniques which provide user with his interested information. recommender system aims to help users for finding their interesting items. Items can be of any type, such as movies, hotels, restaurants, books, news, place and so on. Recommendation methods are mainly classified into collaborative filtering (CF), content based (CB), and hybrid methods [3].There have been many works on recommender systems like analyzing existing recommender system, developing new recommender system. Without considering previous users reviews and user's interests, many traditional recommender systems provide the same recommendation list to users. While processing huge amount data, scalability problem arises. The existing system (KASR) [4] is used to resolve above issues.

***Key Words*: Big data, keywords, recommendation system, thesaurus, user preference**

1. INTRODUCTION

Amount of data increasingly rapidly for analysing and storing such type of data big data is arrived .The growth of productivity and he innovations done in the field of big data worth viewing. Big data refers to data sets whose size is beyond the capability of current technology to manage and process the data within an elapsed time[1].The big data problem has huge impact on the recommender systems. The recommender systems and its growing popularity with large amount of data have yielded the service recommender system into big data environment[2][3].In most traditional hotel recommender system , the services offered and the recommendation list provided for the user are same .The recommendation list given to the user should be distinct for users with distinct preferences so, in order to provide the recommendation list of the user , considering the preferences of users is important[4].The reviews of different users have visited earlier are considered. If the reviews of the users contain the keyword which are similar to the requirements of the current user the name of hotel is recommended to the user[5].

To provide appropriate recommendation to the user, service recommender system is a valuable tool. From last 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. In traditional service recommender systems often suffer from scalability and inefficiency problems when processing or analyzing such large-scale data. The existing service recommender systems fails to meet users 'personalized requirements' because of there is presence of the same 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.

2. LITERATURE REVIEW

[1] has presented the concept of Big Data and importance of big data in enterprise. This paper has defined big data infrastructure requirements and also mentioned that for easily analyzing data, you need the right tools to capture and organize a wide variety of data types from different sources. Yi Cai, Ho-fung Leung, Qing Li, Huaqing Min, Jie Tang, and Juanzi Li [3] have proposed a novel typicality-based collaborative filtering recommendation method named TyCo which selects neighbors of users by calculating user's similarity based on their typicality degrees. Typicality means user typicality vector which indicate the user'spreference on each kind of items. Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen[4] have proposed a Keyword-Aware Service Recommendation method, named KASR which provides a personalized service recommendation list and most applicable recommending service to the users. Here, keywords are used to indicate user's preferences.

3. METHODOLOGY

3.1 Introduction

Keyword- based service recommendation method keywords are used to indicate both of user preferences and the quality of candidate services. A user-based CF algorithm is adopted to generate appropriate recommendations. 1. Read all the reviews and these reviews are collected online, stored in file physically. 2. A list of keywords that determine the interests according to that domain are extracted and store it as keyword interest set. 3. Now each keyword from the list is read and its corresponding related words are obtained.(online Thesaurus) 4. The active user who needs the recommendation system is provided with the keyword list and then user shall select the interest. 5. The interest of previous users shall be extracted from his/her reviews so as to create a keyword interest set using two algorithms namely preprocess and keyword extraction.6. Filter the positive and negative interests of previous users. 7. Search into new data structure if keyword is found in multiple domain of domain thesaurus.8. Measure the similarity computation using approximate and exact similarity computation algorithms. 9. Implement approximate and exact similarity computation algorithm on MapReduce. 10. Calculate the personalized rating with the help of similarity computation. 11. Generate recommendation and provide to active user.

3.1.1 Keyword service list:

Prepare a list of keyword that determines the interest according to respective domain, known as keyword service list, which can be denoted as =1,2…,, is the number of keywords in the keyword service list[4] .

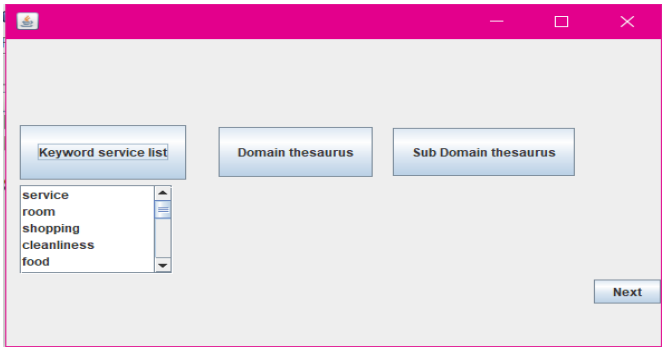

Fig -2: Selecting keyword

3.1.2 Domain thesaurus:

Domain thesaurus is the reference work for keyword service list. It lists words clustered by their similarity of meaning, including related and different words and antonyms [4].

3.1.2 Sub domain thesaurus:

This is the reference work for domain thesaurus that search into dictionary if keyword is found in multiple domains of domain. thesaurus.

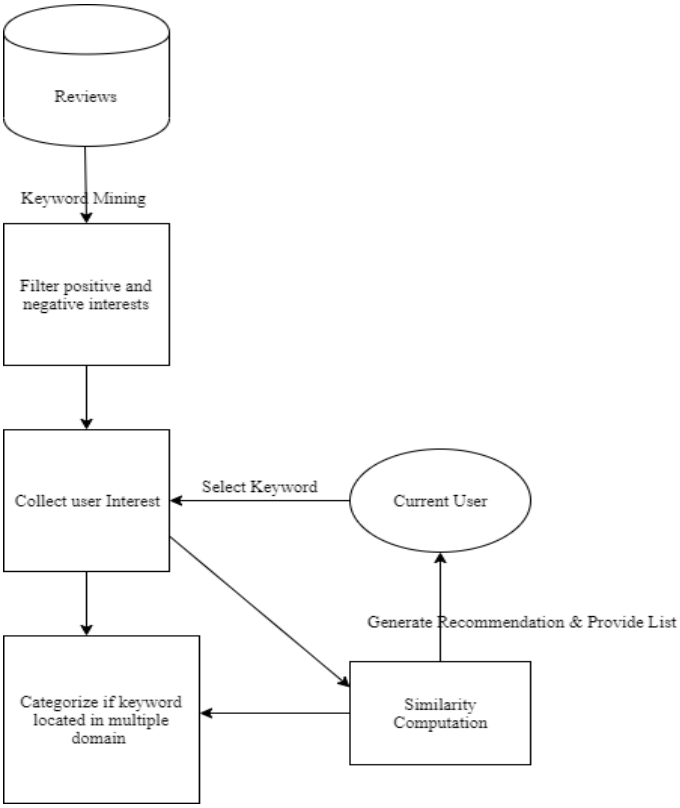

Fig -1: System Architecture

3.1 Proposed Work

In this project, we propose a keyword-based service recommendation method,. In this, keywords 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 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. 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 Collect user interest of previous user and active user.

In this module, interests of active and previous user are collected into respective keyword interest sets.

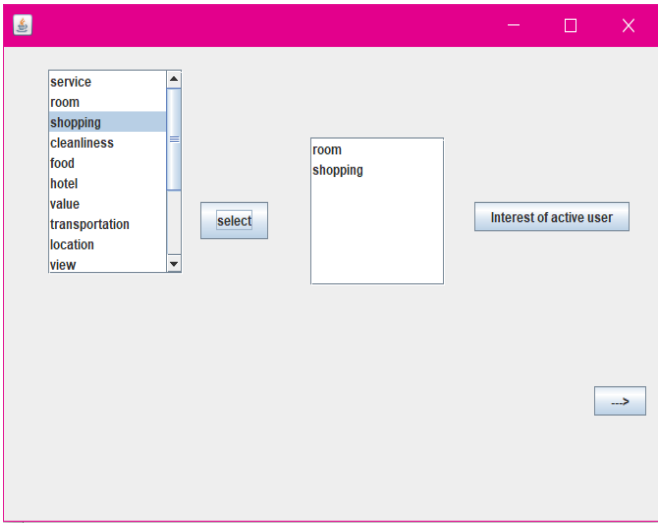

Fig -3: User Interest

Prepare keyword interest set for active user:

An active user shall give his/her interest by selecting keyword from keyword service list. The keyword interest set for active user shall be denoted as AIK. keyword selected from keyword service list, l is the number of selected keywords. For the active user a matrix is generated using the analytic hierarchy process (AHP)[5]. Each keyword is placed in a matrix form and the relative importance is obtained from Saaty scale table[6].The weight of keyword in the keyword interest set of active user will be calculated as,

$$wi = \frac{1}{m}\sum_{j=1}^{m}\frac{a_{ij}}{\sum_{k=1}^{m} a_{kj}}$$

Prepare keyword interest set for previous user:

The keyword interest set for previous user shall be extracted from his/her reviews so as to create a keyword interest set for previous user (PIK). PIK = {p1, p2, ... ph} Where,($1 \le \le h$)is the keyword extracted from the review, h is the number of extracted keywords. As shown in figure 2, two algorithm should be used to extract keywords, preprocess and keyword extraction.

• Preprocess: For improving the quality of keyword in reviews, HTML tags and stop words from reviews should be removed using Porter Stemmer algorithm (keyword stripping)[6].

• Keyword extraction: In this step, each review will be converted into a corresponding keyword set according to the keyword service list, domain thesaurus and sub domain thesaurus.

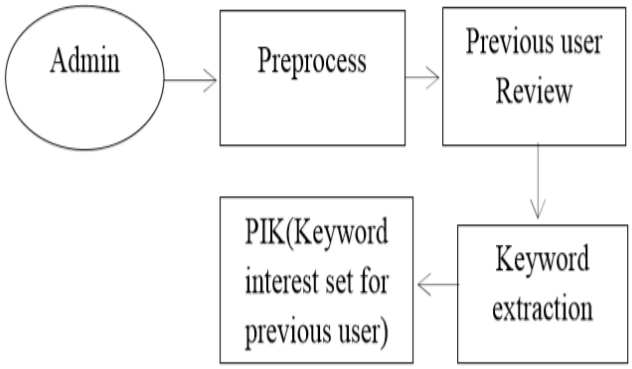

Fig -4: Generate PIK

If a multiple times keyword available in a review, the times of repetitions will be noted and based on that importance value will be assigned. The weight of the keyword in keyword interest set of previous user will be calculated by the term frequency-inverse document frequency (TF-IDF) approach [7].The TF-IDF weight of the keyword in the keyword interest set of user u' calculated as,

$$wpk = TF \times IDF = \frac{N_{pk}}{\sum_g N_{pk}} \times \log\frac{|R'|}{|r' : pk \in r'|}$$

Where, Npk (represents the number of occurrences of keyword pk in all the keyword sets of reviews commented by the same user u', g is the number of keywords in keyword interest set of user u' ,|R'| is the total number of reviews commented by user u' , and |r:pk:r'| is the number of reviews where keyword pk appears.

3.3.2 Filtering interest and Similarity computation.

For filtering the interest, feature-based opinion mining and summarization approach [8] shall be used. This approach will identify and extract features of an object or topic from each sentence of review and then determines whether the opinions about the features are positive or negative. Similarity computation step will be used to identify the reviews of previous users (PIK) who have similar tastes to an active user by finding neighborhoods of the active user (AIK) based on the similarity of their interest [4]. For similarity computation two algorithms shall be used,

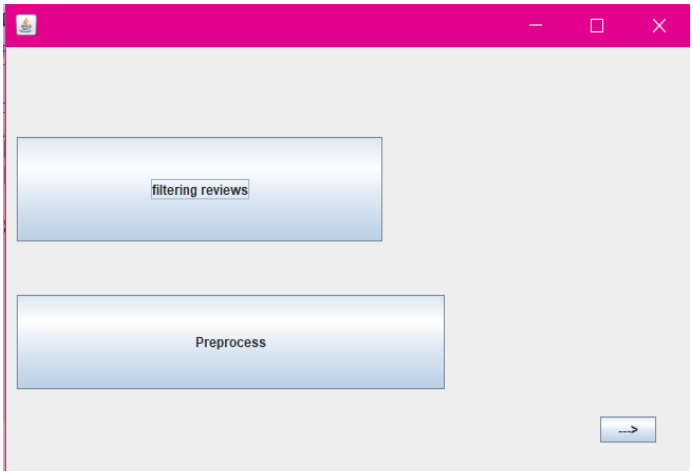

Fig -5: Filtering reviews

Approximate similarity computation method (ASC): the ASC method shall be used for comparing the similarity and variety of sample sets, Jaccard coefficient, is applied in the approximate similarity computation.

$$sim(AIK, PIK) = Jaccad(AIK, PIK) = \frac{|AIK \cap PIK|}{|AIK \cap PIK|}$$

Exact similarity computation method (ESC): A cosine-based approach will be applied in the exact similarity computation, which is similar to the vector space model (VSM) in information retrieval [9][10]. The similarity based on the cosine based approach is defined as follows:

$$sim(AIK, PIK) = \cos(Wa, Wp) = \frac{W_a \cdot W_p}{\|W_a\| \times \|W_p\|} = \frac{\sum_{i=1}^{n} W_{a.i} \times W_{p.i}}{\sqrt{\sum_{i=1}^{n} W_{a.i}} \sqrt{\sum_{i=1}^{n} W_{p.i}}}$$

Where Wa and Wp are respectively the interest weight vectors of the active user and a previous user. Wa.i is the ith dimension of Wa and represents the weight of the keyword Ki in keyword interest set AIK. Wp.i is the vith dimension of Wa and represents the weight of the keyword Ki in keyword interest set PIK.

3.3.3 Calculate personalized ratings and Generate recommendation interest.

Based on the similarity of the active user and previous users, further filtering will be conducted. 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].

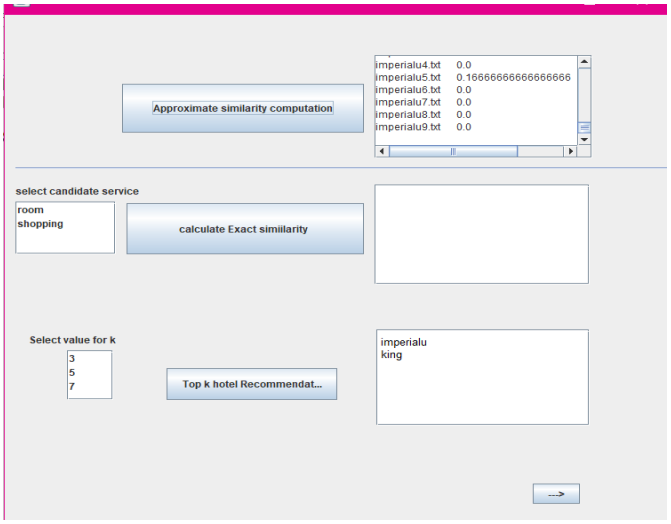

Fig -6: Recommending Top K Hotels

4. CONCLUSION

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 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 according to the keyword-candidate list and domain thesaurus. Our method aims at presenting a personalized service recommendation list and recommending the most appropriate service to the users.

REFERENCES

[1] An Oracle White Paper June 2013 “Oracle: Big Data for the Enterprise”.

[2] IRINA NEAGA, YUQIUGE HAO, “TOWARDS BIG DATA MINING AND DISCOVERY” ,Short Research Papers on Knowledge, Innovation and Enterprise.

[3] Yi Cai, Ho-fung Leung, Qing Li, Senior Member, IEEE, Huaqing Min, Jie Tang, and Juanzi Li,” Typicality-Based Collaborative Filtering Recommendation” IEEE Trans. Knowledge and Data Eng., vol. 26, no. 3, March 2014.

[4] Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen "KASR: A KeywordAware Service Recommendation Method on MapReduce for Big Data Applications," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no.12, December 2014.

[5] Thomas L. Saaty, “How to make a decision: The Analytic Hierarchy Process” European journal of operational research 48, 1990.

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

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

[8] F.Chang, J.Dean, SGhemawat,andWCHsie,"Bigtable: A distributed storage system for structured data"ACM Transactions on Computer Systems,Vol. 26, No. 2

(9), 2008. [4] W. Dou, X. Zhang, J. Li,J. Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications,"IEEE Transactions on Parallel and Distributed Systems,2013.

[10] 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.

[11] M. Bjelica, "Towards TV Recommender System Experiments with User Modeling," IEEE Transactions on Consumer Electronics, Vol. 56, No.3, pp. 1763-1769, 2010.

[12] 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

[13] F. Chang, J. Dean, S. Ghemawat, and W.C. Hsieh, "Big table: ADistributed Storage System for Structured Data