International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 8 156 - 159

A Survey on Review Based Recommendation System

Ms. Sarika P.Khot
ME Student, Dept. of CSE,
D. Y. Patil College of Engineering & Technology,
Kolhapur, India.
khotsarika8@gmail.com

Dr. V. R. Ghorpade
Professor, Dept. of CSE
D. Y. Patil College of Engineering & Technology,
Kolhapur, India
line 4: e-mail: name@xyz.com

***Abstract*—** The advances in internet technology have resulted in the generation of huge amount of data called as Big Data. Recommendation system is a widely used technique for the filtering the huge amount of data and providing recommendations to users according to their interest. Without taking previous user interest into consideration, the traditional recommender system does not provide efficient solutions to the users. In this paper, we introduce recommender system to solve the above-described problems. The proposed recommender system will take into consideration previous user's interest and active user interest and by calculating similarity it will to provide recommendations to active user.

Keywords- *User's Interest; Hadoop; Big Data; Recommendation System; Domain Thesaurus.*

*****

I. INTRODUCTION

Recommender systems are commonly used techniques which provide user with his interested information. The 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.

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. For example, In the online shopping, we browse through products. The Recommendation system offer recommendations of products we might be interested in but without considering the perspective of business or consumer. A typical Recommendation system cannot perform well without sufficient data and big data supplies huge amount of user data such as previous purchases log, browsing history, and reviews for the Recommendation systems to provide effective recommendations. While processing huge amount data, scalability problem arises. The existing system (KASR) is used to resolve above issues. The existing system might suffer from accuracy and inefficiency problems. The proposed system resolves these problems by filtering positive user interest.

II. LITERATURE SURVEY

Yi Cai, Ho-fung Leung, Qing Li, Huaqing Min, Jie Tang, and Juanzi Li[1] 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. Typacality means user typicality vector which indicate the user's preference on each kind of items.

Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen [3] 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.

Thomas L. Saaty [4] have presented the analytical hierarchy process (AHP) process, a decision making approach. The AHP is used when final decision based on the calculation of a number of alternatives in terms of a number of criteria and criteria are expressed in different units.

B. Issac and W.J. Jap [5] have implemented Bayesian spam detection scheme with context matching using the Porter Stemmer algorithm. This helps to make the keyword search efficiently by considering stem word only.

Alisa Kongthon, Niran Angka wattanawit, Chatchawal Sangkeettrakarn [6] has provided a feature-based summary and analysis of a large number of customer reviews about objects like product, service, event etc. This approach is used to analyze customer's opinions about feature whether positive or negative opinion.

P. Castells, M. Fernandez, and D. Vallet [7] has proposed a novel based approach which includes an ontology-based scheme for the semiautomatic annotation of documents and a retrieval system. This scheme provides better search capabilities by qualitative improvement over keyword-based full-text search.

III. MOTIVATION

For many services like hotels, restaurants, movies, tourist places, we find numerous reviews which are updated online by the service users. If reviews of tourism agency considered then, we find that each review is based on the personal interest of a user. If a user John has interest for good service and good food provided by the agency then he writes mainly about that. Similarly if the user Bob has an interest for the number of tourist spots shown by the tour operator, he shall write only about that experience. Most of the recommender system do not consider the interest of the user and simply show the recommendations based on all factors. Considering the above features we can then design a system which will recommend the user, depending on the user interest.

IV. SYSTEM ARCHITECTURE

In the proposed system, reviews play a vital role. The reviews of previous user will be collected and stored into file. A list of keywords that determine the interests according to that domain will be extracted and store it as keyword service list. After that, each keyword from the list will be read and its domain thesaurus will be generated. Then, the active user who needs the recommendation will be provided with the keyword list. As shown in Figure1 the active/current user will select keyword from given keyword service list. The selected keywords from active user will be collected as active user interest. For collecting previous user interest, reviews of previous user will be filtered out and keywords will be extracted. After collecting previous and active user interest, similarity between users interest will be checked using approximate and exact similarity algorithms. When similar users found, personalized rating gets calculated. According to personalized ratings recommendation will be generated and it will be provided to active user.

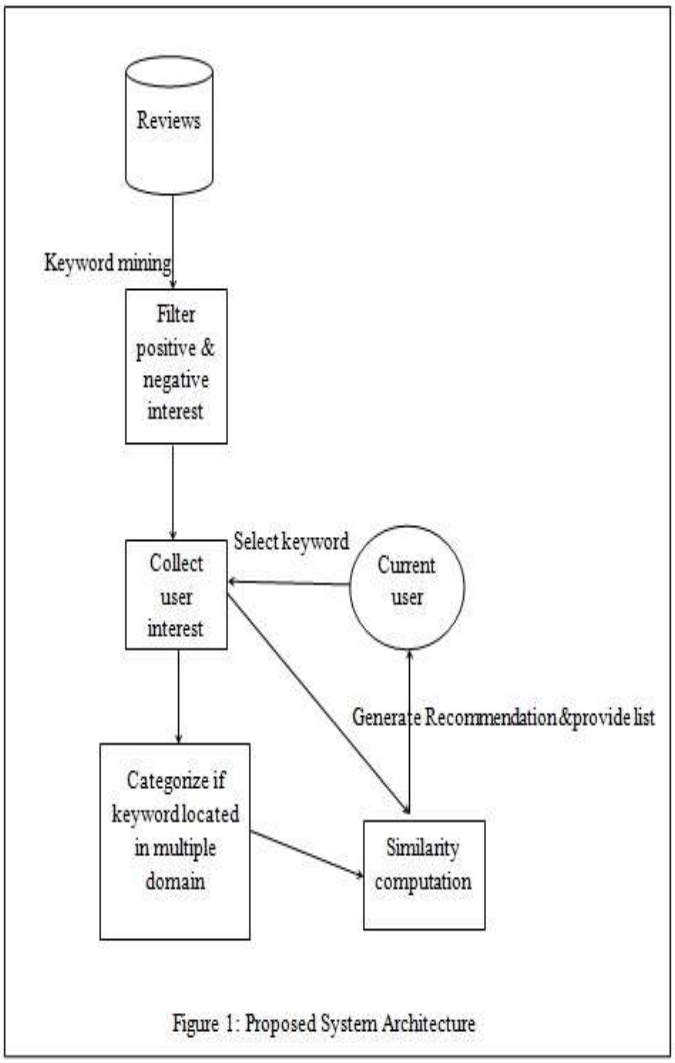

Figure 1. Roposed System Architecture.

Technical contribution of this paper:

- To provide recommender system for a service according to users interest.
- To improve the accuracy of existing recommender system, taking into account the previous user's interest and filtering positive user interest.

V. METHODOLOGY

The proposed system will be designed and implemented in the following modules:

A. *Data preparation Module*

The proposed system is based on reviews. Initially, these reviews are taken from online and stored into file. To help obtain a user's interest, three kinds of data structure gets prepared and introduced as follows:

a) 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 $K = \{k1, k2 \ldots, ki\}, i$ is the number of keywords in the keyword service list.

b) 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.

c) 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.

B. *Collect user interest of previous user and active user Module*

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

a) 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

$$AIK = \{a1, a2 \ldots, al\},$$

where $ak\ (1 \leq k \leq l)$ is k th 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).Each keyword is placed in a matrix form and the relative importance is obtained from saaty scale table. The weight of keyword in the keyword interest set of active user will be calculated as,

$$wi = \frac{1}{m} \sum_{j=1}^{m} \frac{aij}{\sum_{k=1}^{m} akj}$$

Where, aij the relative importance between two keywords, m is the number of the keywords in the keyword interest set of the active user.

b) Prepare keyword interest set for active user: An 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).

$$PIK = \{p1, p2, \ldots . ph\}$$

Where, $pk\ (1 \leq k \leq h)$ is the k th keyword extracted from the review, h is the number of extracted keywords.

a) 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).

b) 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.

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 thekeyword in keyword interest set of previous user will be calculated by the term frequency-inverse document frequency (TF-IDF) approach. The TF-IDF weight of the keyword pk in the keyword interest set of user u' calculated as,

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

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.

C. Filtering interest and Similarity Computation module

For filtering the interest, feature-based opinion mining and summarization approach 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. For similarity computation two algorithms shall be used,

a) 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

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

b) 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. The similarity based on the cosine based approach is defined as follows,

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

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 ith dimension of Wp and represents the weight of the keyword ki in keyword interest set PIK.

Calculate personalized ratings and Generate recommendation Module

Personalized rating shall be calculated by comparing similarity computation value with threshold value (δ). If $sim(AIK, PIK) < \delta$ then the keyword interest set of a previous user will be filtered out. When the set of most similar users will be found, the personalized ratings of each candidate service for the active user will be calculated. Lastly, a personalized service recommendation list will be provided to the user with the highest rating recommendation. To calculate the personalized rating pr of a service for the active user, weighted average approach will be used.

$$pr = ar + k \sum_{PIK \in R} sim(AIK, PIKj) \times (rj - ar)$$

$$k = \sum_{PIK \in R} sim(AIK, PIKj)$$

Where, $sim(AIK, PIKj)$ is the similarity of the keyword interest set of the active user AIK and the keyword interest set of a previous user $PIKj$, multiplier k denotes as a normalizing factor, R denotes the set of the remaining keyword interest sets of previous users after filtering, rj is the rating of the corresponding review of $PIKj$, and r is defined as the average ratings of the candidate service.

VI. CONCLUSION

In this paper, we have introduced a review based recommendation system. This recommendation system provides a service by considering active user interest and previous user interest. Here, keywords are used to indicate user's interest and a user based collaborative filtering algorithm is implemented to generate appropriate recommendations. More specifically, a keyword service list and domain thesaurus is provided to help to obtain user's interest. The active user gives his/her interest by selecting the keywords from the keyword service list, and the interest of the previous users can be extracted from their reviews corresponding to the keyword service list and domain thesaurus. The objective of this recommendation system is to provide a recommendation list and recommending the service to the active user. Moreover, to improve the accuracy of keyword aware service recommendation system (KASR), we have filtered positive and negative interest of the previous users from their reviews.

REFERENCES

[1] Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan , book ,"Collaborative Filtering Recommender Systems"

[2] Yi Cai, Ho-fung Leung, Qing Li, Senior Member, IEEE, Huaqing Min, Jie Tang, and

[3] 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 Keyword-Aware 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] B. Issac and W.J. Jap, "Implementing Spam Detection Using Bayesian and Porter Stemmer Keyword Stripping Approaches,"

Proc. IEEE Region 10 Conference. (TENCON '09), pp. 1-5, 2009.

[7] Alisa Kongthon, Niran Angkawattanawit, Chatchawal Sangkeettrakarn, Pornpimon Palingoon, Choochart Haruechaiyasak, "Using an Opinion Mining Approach to Exploit Web Content in Order to Improve Customer Relationship Management".

[8] Dunren Che1 ,Mejdl Safran , and Zhiyong Peng "From Big Data to Big Data Mining: Challenges, Issues, and Opportunities".

[9] An Oracle White Paper June 2013 "Oracle: Big Data for the Enterprise".

[10] Irina Neaga, Yuqiue Hao, "TOWARDS BIG DATA MINING AND DISCOVERY" , Short Research Papers on Knowledge, Innovation and Enterprise.

[11] Kaushik Pal "How Big Data is used in Recommendation Systems to change our lives",kDnuggets Home, tutorials, Oct 2015.

[12] Robin burke, Hybrid Recommender Systems: Survey and Experiments1" Department of Information Systems and Decision Sciences, California State University, Fullerton, CA 92834, USA.

[13] O'Reilly Media, Inc., Book, "Big Data Now"2013 edition.