International Journal of Current Trends in Engineering & Research (IJCTER)
e-ISSN 2455–1392 Volume 2 Issue 4, April 2016 pp. 1 - 8
Scientific Journal Impact Factor : 3.468
http://www.ijcter.com

A Keyword Aware Service Recommendation System

Pardeshi Gayatri[1], Pawar Vividha[2], Shivde Prerna[3], Pise Pooja[4]

[1,2,3,4] *Dept. Of IT., MET's Institute of Engineering, Bhujbal Knowledge City, Adgaon, Nashik (Ms), India*

Abstract—Service recommender system is most common in recent year and useful in variety of applications. It is provide appropriate recommendation to the user. In the last year the amount of user, services and online data has grown rapidly , the big data analysis the problem for service recommender system. Traditional recommender system was suffer from scalability and inefficiency problem when analysing large amount of data. Existing recommender system provide same rating and ranking of the services for different user but they does not consider users preferences and therefore they are fail to meet users specific requirement. In this paper, we propose the process of developing travel recommender system for address the above challenges. The aim of Keyword Aware Service Recommender System (KASR) is provide service recommendation list and give appropriate services recommendation to the user. It gives appropriate recommendation by collaborative filtering algorithm and keyword are used for indicating users preferences. To increase scalability and inefficiency,

KASR is implemented on Hadoop by using Big Data and for parallel processing Map-Reduce paradigm is use. Finally, large experiment is conducted by using Real-World dataset , that produce result for KASR significantly improve accuracy and scalability of service recommender system over existing recommender system.

Keywords—Personalized Recommendation, Hadoop, Keyword, MapReduce, Preferences, Bigdata, Domain Thesaurus

I. INTRODUCTION

In Recent year the large amount of data has been grown rapidly that data handling is very difficult so the best solution to handle that data in systematic manner we can use "Big Data"[1].

"Big Data': It is collection of large datasets that cannot be processed using traditional computing techniques. Big Data involves the data produced by various devices like social media, Stock exchange data, power grid data, transport data, search engine data.[2]Thus BigData includes huge volume, high velocity and extensible variety of data. With the help of Big Data we can improve the performance of service Recommendation system in large amount of data.

Recommendation system is a subclass of information filtering system that seek to predict the 'rating' or 'preferences' of that user. Recommender system typically produce appropriate recommendation to the user of what exactly user want. There are mainly two ways to produce Recommendations-Collaborative Filtering and Content-Based Filtering.

Collaborative Filtering approach building a model from a user's past behaviour (items previously purchased or selected ,ratings of that item)as well as similar decision made by other users[11]. Content-based Filtering approach utilizes a series of discrete characteristics of an item in order to recommend additional items with similar properties. Service Recommendation system is a valuable tool for the user to provide appropriate recommendation to user.

Examples of such practical applications include CDs, books, web pages and various other products now use recommender systems.

II. MOTIVATION

Now a day's number of customers, services and other online information is growing Rapidly so use of Service recommendation system on big data application is difficult for solving critical challenges. Moreover, in most existing service recommender systems, such as Tourism systems and tourism guides, the ratings of services and the service recommendation lists presented to users are the same. So user's different preferences and users personalized requirements are not considered.

To improve the accuracy and scalability of service recommender systems over existing approaches is the main Motivation of KASR Method and we have implemented it on a MapReduce framework in Hadoop platform.

III. PRELIMINARY KNOWLEDGE

3.1 Recommender System And Collabrative Filtering

Recommendation system is a sub class of information filtering system that seek to predict the 'rating'[14][15] or 'preference' that a user would give to an item. Recommender system have extremely common in recent year, and are applied in a variety of applications. The most popular once are probably movies, music, news, books, research article ,search queries, social tag and product in general. Current recommendation method can be classified into three main category :Content based, collaborative and hybrid recommendation approach. Content based filtering approach utilise a series of discrete characteristics of an item in order to recommend additional item with similar properties. Collaborating filtering approach building a model from a users passed behaviours and as well as similar decisions made by other users. This model then use to predict item that the user may have an interest in .Hybrid approach combines content-based and collaborative method in several different way.

In collaborative algorithm classify Item-based and User-based collaborative filtering. In item based, collaborative filtering is a process of filtering for information or patterns using techniques involving collaboration among multiple agent, viewpoint, data sources, etc. In user-based, collaborative filtering method is making of automatic prediction about the interest of a user by collecting preference or taste information from many users. The underlying assumption of the collaborative filtering approach is that if a person A has same opinion as a person B on an issue ,A is more likely to have B's opinion on a different issue x than two have the opinion on x of a person chosen randomly. And we use user based collaborative algorithm to deal with our problem.

3.2 Cloud Computing and MapReduce

Cloud computing is also known as 'on-demand computing', is a kind of internet based computing ,where shared resources ,data and information are provided to computers and other devices on-demand. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and processes there data in third party data centres.

There are several Cloud computing tools are available: Hadoop, Mahout, MapReduce of google[4], dynamo of amazon.com, the dryad of Microsoft and naptune of ask.com, etc. Hadoop is open source software framework for distributed storage and distributed processing of very large datasets on computer clusters built from commodity hardware. MapReduce is a programming model and associated implementation for processing and generating large data set with parallel ,Distributed algorithm on a cluster .

"Everything as a service" has a creating a big service so 'servicelization' is a way of offering social networking services, Big data analytics and internet services. Thus cloud computing tool is use to improve the scalability and efficiency use of service recommendation methods in the "Big Data" environment.

IV. KEYWORD AWARE SERVICE RECOMMENDATION SYSTEM

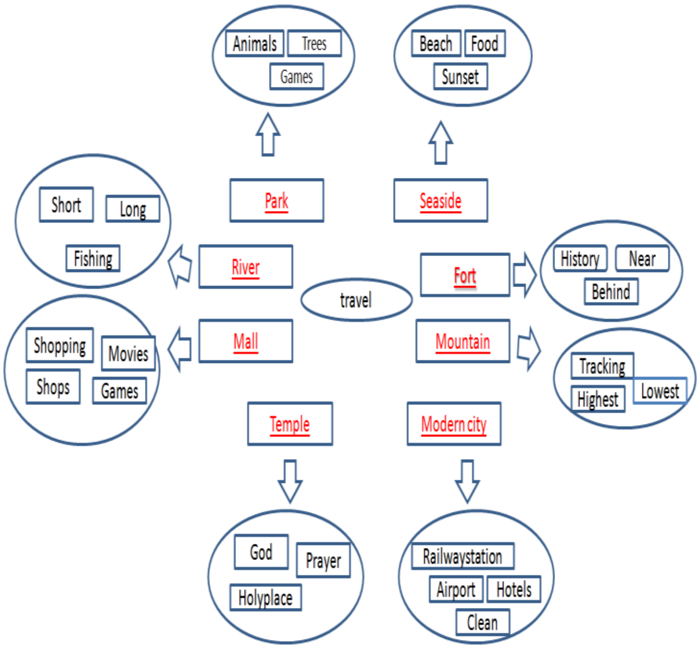

Fig.1 An example of a simple domain thesaurus of travel system

Definition 1 (Keyword-Aware Service Recommendation method, KASR) : In this paper, we propose a keyword- aware service recommendation system, named KASR. In this system, keywords are used to indicate users' preferences and the quality of candidate services. A user based CF algorithm is used to generate appropriate recommendations for the user. The main aim of KASR is 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 to her/him.

Moreover, to improve the efficiency and scalability of our recommendation system in "Big Data" environment, we implement it in a MapReduce framework on Hadoop by splitting the proposed algorithm into multiple MapReduce stages. Table 1 summarizes the basic symbols and notations used in this paper.

TABLE 1 Basic symbols and notations

Symbol	**Definition**
K	The keyword-candidate list, *K={k1, k2, …,kn}*
APK	The preference keyword set of the active user
PPK	The preference keyword set of a previous user
Sim(APK,PPK)	The similarity between *APK* and *PPK*
$\boldsymbol{W_P}$	A preference weight vector
$\mathbf{W}_{AP}$	The preference weight vector of the active user
$\mathbf{W}_{PP}$	The preference weight vector of the previous user

4.1 Keyword-candidate List and Domain Thesaurus: In our system, two data structures, "keyword-candidate list" and "Domain thesaurus", are introduced to help obtain users' preferences.

Definition 2 (Keyword-candidate list):

The keyword candidate-list is a set of keywords about users' preferences and multi-criteria of the candidate services, this can be denoted as K={k1,k2,k3…,kn} n is the number of the keywords in the keyword-candidate list. An example of a simple keyword-candidate list of the travel system is described in Table 2. Keywords in the keyword candidate list can be a word or multiple words related with the quality criteria of candidate services.

In this paper, the preferences of passive(previous) users will be extracted from their reviews for candidate services and formalized into a keyword set. Usually, since some of words in reviews can not exactly match the corresponding keywords in the keyword-candidate list which characterize the same aspects as the words. The corresponding keywords should be extracted as well. In this paper, we assume that domain thesauruses are built to support the keyword extraction, and different domain thesauruses are built for different service domains.

TABLE 2 Keyword-candidate list of Travel system

NO.	**Keyword**	**NO.**	**Keyword**	**NO.**	**Keyword**
1	River	6	City	11	Zoo
2	Seaside	7	Fort	12	College
3	Park	8	Creative place	13	Mountain
4	Temple	9	Mall	14	School
5	Museum	10	Hotel		

Definition 3 (Domain Thesaurus):

A domain thesaurus is a related work of the keyword-candidate list that lists words grouped together according to the similarity of keyword meaning, including related and contrasting words[16][17]. An example of a simple domain thesaurus of travel system is shown in Fig.1, the red colour words 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

4.2 A Keyword-aware Service Recommendation Method:

The main steps of KASR are depicted in Fig. 3, which are described in detail as follows.

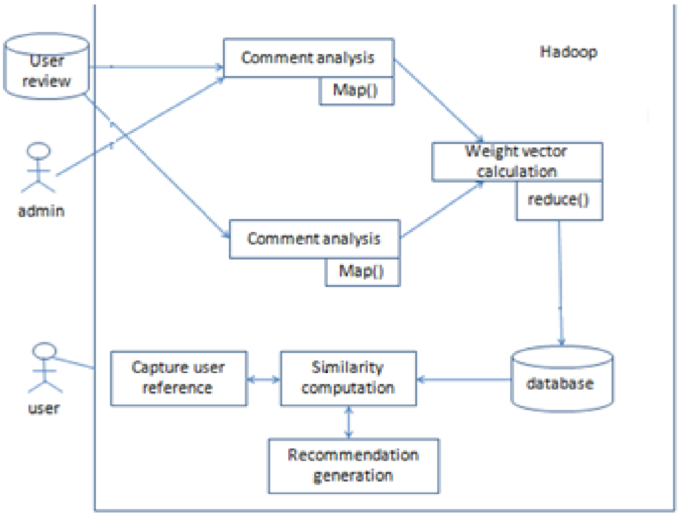

Fig.2: KASR main steps

(1) Capture user preferences by a keyword aware approach:

There are two types of user Active User and passive User so preferences of active users and passive users are captured and corresponding preference keyword sets are generated respectively. an active user is one user who needs recommendation.

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

- *Preferences of previous users.* The preferences of a previous user for a candidate service are extracted from his/her reviews for the service according to the keyword-candidate list and domain thesaurus.

(2) Similarity computation:

Two similarity computation algorithms are proposed in this method i.e approximate similarity computation and exact similarity computation.

a. Approximate Similarity Computation:

To calculate similarity and diversity of the sample sets Jaccard coefficient method is applied in the approximate similarity computation. Jaccard coefficient measures asymmetric information on binary (and non-binary) variables when negative value give no information then Jaccard Coefficient is used. This method shows similarity between active user preferences and previous user preferences. Jaccard coefficient described as follows:

$$\text{Sim}(APK,PPK)=$$

$$\text{Jaccard}(APK,PPK)=\frac{|APK\cap PPK|}{|APK\cup PPK|} \qquad [1]$$

Algorithm 1: SIM-ASC (Approximate Similarity Computation)

Input: The preference keyword set of the active user *APK* The preference keyword set of a previous user *PPKj*
Output: The similarity of *APK* and *PPKj*, $sim_{ASC}(APK,PPK_j)$

1. $sim_{ASC}(APK,PPK_j)=\frac{|APK \cap PPKj|}{|APK \cup PPKj|}$

2.**return** the similarity of *APK* and *PPKj*, $sim_{ASC}(APK,PPK_j)$

b. Exact Similarity Computation: A cosine based approach in the exact similarity computation, which is similar to Vector Space Model in information retrieval [12][13].

Algorithm 2: SIM-ESC (Exact Similarity Computation)

Input: The preference keyword set of the active user *APK* The preference keyword set of a previous user *PPKj* **Output:** The similarity of *APK* and *PPKj*, $sim_{ESC}(APK,PPK_j)$.
1: for each keyword *ki* in the keyword-candidate list .
2.if $K_i \in APK$ **then**
3. get $\vec{W}_{AP,}\, i$ by formula (2)
4.else $W_{AP,}$ i=0
5.end if
6.if $k_i \in \vec{PPK_j}$ **then**
7.get Wpp_j ,i by formula(5)
***8*.else** Wpp_j ,*i*=0
***9*.End if**
***10*.end for**
11.get $sim_{ESC}(APK,UPK_j)$ by formula(6).
12.return the similarity of APK and PPK_j, $sim_{ESC}(APK,UPK_j)$

(3) Calculating personalised Ratings and generate Recommendations:

Based on the similarity computation the filtering of Active user and Passive user will be calculated. Once the set of most similar users are found the personalised rating is for each candidate service of the Active user is calculated. Finally a personalised service recommendation list will be display to the end user and services with highest amount of rating that will be recommend to him/her.

Algorithm 3:Basic Algorithm of KASR

Input: The preference keyword set of the active user *APK*
The candidate services WS={ws_1,ws_2.....ws_n}
The threshold δ in the deep filtering phase.
The number k
Output: The services with the Top k highest rating
{tws_1,tws_2.....tws_k}
1.for each service ws_i € WS
2.$\hat{R}$ =ϕ, sum=0,r=0.
3: for each review *Rj* of service *wsi*
4: process the review into a preference keyword set *PPKj* .
5.if $PPKj \cap APK \neq \phi$ **then**
6. insert *PPKj* into $\hat{R}$
7. end if
8.end for
9.for each keyword set *PPK j* € $\hat{R}$
10. *sim* (*APK*,PPK_j)= *sim* (*APK*,PPK_j)
11.if *sim* (*APK*,PPK_j) < ϕ **then**
12. remove *PPKj* from $\hat{R}$
13.else *sum = sum+1,r=r+r*
14. end if
15: end for
16: $\bar{r}$ = r/sum
17. get *pri* by formula (7)
18.end for
19. sort the services according to the personalized ratings pr_i.
20. return the services with the Top-k highest ratings {tws_1,tws_2.....tws_k}

V. RELATED WORK

There have been number of recommender systems developed in both industry and academia. In[1], the authors purpose "How do your data grow?" In recent year Big data is grown rapidly. We need to analyse such a large amount of data. Here the author define how this data is grow rapidly in the systems. In[7], the author purpose new Recommendation Techniques for Multicriteria Rating Systems. Which will provide rating to the system.In [11], Adomavicius and Tuzhilin give an overview of the field of recommender systems and describe the current generation of recommendation methods. They also describe various limitations of current service recommendation methods, and discuss possible extensions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications. Most existing service recommender systems are only based on a single numerical rating to represent a service's utility as a whole [12]. In [10], the authors propose a Bayesian-inference-based recommendation system for online social networks. They show that the proposed Bayesian-inference-based recommendation is better than the existing trust-based recommendations and is comparable to Collaborative Filtering recommendation. In fact, evaluating a service through multiple criteria and taking into account of user feedback can help to make more effective recommendations for the users. In [1], the authors purpose "How do your data grow?" that means increasing of the Big Data in recent year is clear.

VI. CONCLUSION

In this paper, we proposed Keyword Aware Service Recommendation. In this KASR System keywords indicate Preference of user and collaborative filtering algorithms used to generate appropriate recommendation. more specifically , a keyword Candidate List and Domain Thesaurus helps to obtain preferences by using keyword Candidate list. The active user gives his/her preferences from the passive user Reviews, preferences can be extracted for services. our method aim is to provide Personalised service recommendation list and recommend appropriate service to the user.

Moreover, for improving scalability &efficiency problem we use "big data" Environment and we implement this system on Hadoop platform by using MapReduce framework. finally it provides appropriate recommendation over existing approach.

REFERENCES

[1] C. Lynch, Big Data: How do your data grow? Nature, Vol. 455, No. 7209, pp. 28-29, 2008
[2] J. Manyika, M. Chui, B. Brown, et al, Big Data: The next frontier for innovation, competition, and productivity, 2011:
[3] Multicriteria user model in recommender system issue No.02 - March/April(2011vol.26) pp: 64-76 KleanthiLakiotaki:
[4] MapReduce:simplified data processing on large clusters: Jeffrey Dean Communications of the ACM - 50th anniversary issue: 1958 - 2008 CACM Home page archive .
[5] H. Liang, J. Hogan, and Y. Xu, Parallel User Profiling Based on Folksonomy for Large Scaled Recommender Systems: An Implementation of Cascading MapReduce, In Proceedings of the IEEE International Conference on Data Mining Workshops, pp. 156-161, 2010:
[6] Y. Pan, L. Lee, Performance analysis for lattice-based speech indexing approaches using words and subword units, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, No. 6, pp. 1562-1574, 2010.
[7] G. Adomavicius, and Y. Kwon, New Recommendation Techniques for Multicriteria Rating Systems, IEEE Intelligent Systems, Vol. 22, No. 3, pp. 48-55, 2007.
[8] Y. Jin, M. Hu, H. Singh, D. Rule, M. Berlyant, and Z. Xie, MyS-pace Video Recommendation with Map-Reduce on Qizmt, Proceedings of the 2010 IEEE Fourth International Conference on Se-mantic Computing, pp.126-133, 2010.
[9] 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. 478481, 2010.
[10] X. Yang, Y. Guo, Y. Liu, "Bayesian-inference based recommendation in online social networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 4, pp. 642-651, 2013.
[11] 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.
[12] P. Castells, M. Fernandez, and D. Vallet, "An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval," IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No.2, pp. 261-272, February 2007.
[13] Y. Zhu, Y. Hu,"Enhancing search performance on Gnutella-like P2P systems," IEEE Transactions on Parallel and Distributed Systems, Vol. 17, No. 12, pp. 1482-1495, 2006.
[14] 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 Computing System, pp. 194-201, 1995.
[15] 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.
[16] H. Schütze and J. O. Pedersen, "A cooccurrence based thesaurus and two applications to information retrieval," Information Processing & Management, Vol. 33, No. 3, pp. 307-318, 1997.
[17] Y. Jing and W. Croft, "An association thesaurus for information retrieval," Proceedings of RIAO, Vol. 94, No. 1994, pp.146-160, 1994.

