I J C T A, 9(14) 2016, pp. 6667-6673

KULSR: Keyword and User Locality-based Service Recommendation Method on Map Reduce for Big Data Applications

Ujjwal Sinha Mahapatra* **and R. Annie Uthra***

ABSTRACT

Advertising nowadays became a regular trend in the marketing world, and a recommendation system deals with the task of providing an appropriate advertisement to a customer. Its challenge is not only limited to providing recommendations but also it has to deal with analyzing a large set of data on a scalable and scalability issues. The service recommendation systems cannot be made by considering only the ratings and other attributes of services but to give personalized recommendations to users it also has to consider the attributes of users. In this paper a Keyword and User Location-based Service Recommendation system is proposed, named KULSR, it will deal with the above challenges and provide more personalized service recommendations to users. Keywords and location of users were used as analyzing criteria of deciding proper service to be recommended. KULSR is implemented on Hadoop distributed computing platform using the MapReduce parallel processing technique. Finally, experiments are done on data sets and obtained results show that the system is more accurate than KASR method [1] of service recommendation.

Keywords: Recommender System; reviews; preference; keyword; location based; big data; MapReduce; Hadoop

I. INTRODUCTION

Recommendation systems [2] has a vital role in marketing area in online retail business. These systems have information matching and filtering techniques so that some particular set of items (some products or some services or some set of information) which are associated with users are suggested to the users by applying some association rule of statistical methods between those users' and previous users' attributes. So, a user-based collaborative filtering method is implemented, i.e. the recommendations are per-user-based rather than per-item based [3]. A user-based collaborative filtering [4] method is used to find similar users and those item sets are suggested to the users that are having highest ratings by the similar users.

The challenge also exists in implementing the system in the big data environment, similar to any other system the big data also affects the implementation of service recommendation systems. As the number of services is increasing day by day, the task of providing relevant services to users is turning into a research issue. Since 1992, much research work is done to provide an appropriate approach for recommendation systems.

In this paper keyword [5] extraction method and matching of location done to increase the accuracy of a service recommendation system. This technique can refine final recommendations which are to be suggested to the users.

II. RELATED WORK

Many approaches for building recommendation systems has been given till now which can be classified as Collaborative filtering [6], Content-based methods and Hybrid methods. In [7], The different methods of

* Department of Computer Science and Engineering, SRM University, Chennai, India, *E-mails: usmahapatra@gmail.com; annie.u@ktr.srmuniv.ac.in*

recommendation systems are explained and their performance is evaluated, and in the paper limitations of recommendation systems are also discussed. Recommendation systems can be implemented in Big Data environment using Hadoop and MapReduce technique can be used for scalable applications. For some local datasets, My SQL database is used.

Keyword extraction and string matching can be done using string handling in Java, and in [1] many recommendation systems are compared with their KASR method implementing on review dataset of users related to hotel related data.

Finally, comparing with KASR method user location is also considered in KULSR method which makes the recommendations more personalized. Scalability features in KULSR on MapReduce method is provided by Hadoop infrastructure.

III. KEYWORD AND LOCATION-BASED SERVICE RECOMMENDATION SYSTEM

For the experiment to be performed the service taken into consideration is of hotels, the recommendation of candidate services (some set of hotels) is done on the dataset of reviews of previous users.

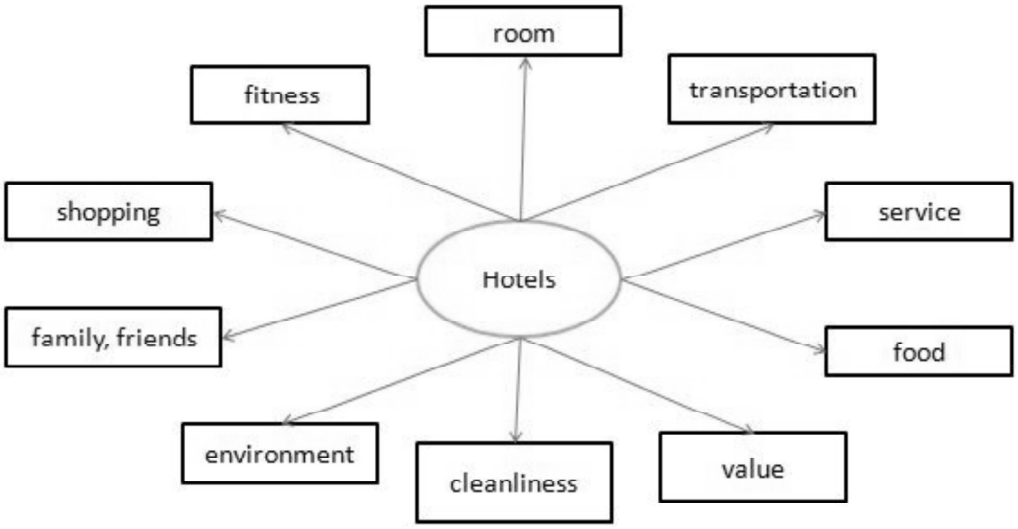

Figure 1: Keywords used in the experiment

Some terms in the previous users' reviews which belongs to particular domain were termed as related keywords.

The main steps of KULSR are depicted in Fig.2, whose details are mentioned as follows:

1. Preprocessing of review dataset: Before keyword extraction, stop words or any useless words like can, able, for, is, are, etc. were removed from the dataset which contains previous users' reviews.
2. Capture keywords from users: An active user has to select the keyword from a given list, and previous users keywords were determined by the number of related keywords present in the list of reviews submitted by previous users. Location of the active user is also taken here to do filtering in the last phase.

   To implement the algorithm efficiently in Big Data environment it is executed in a MapReduce framework, to accomplish this, the review dataset of previous users was saved in Hadoop distributed file system.
3. Similarity computation: To calculate similarity, the association between previous users' and active users is calculated by using Jaccard coefficient, it matches the domain of active users with the

previous users' entered reviews and ultimately gives recommendations for users of the same domain. The basic symbols used in this paper are described in Table I.

Table 1
Symbols used in the experiment

Symbols	*Definition*
APK	The review set of active user
PPK	The review set of previous user
Km	Matched keywords
Tk	Total number of keywords
CSi	Candidate services
LOCa	Location of active user
LOCp	Location of previous user in particular row in previous users detail dataset (PPK)

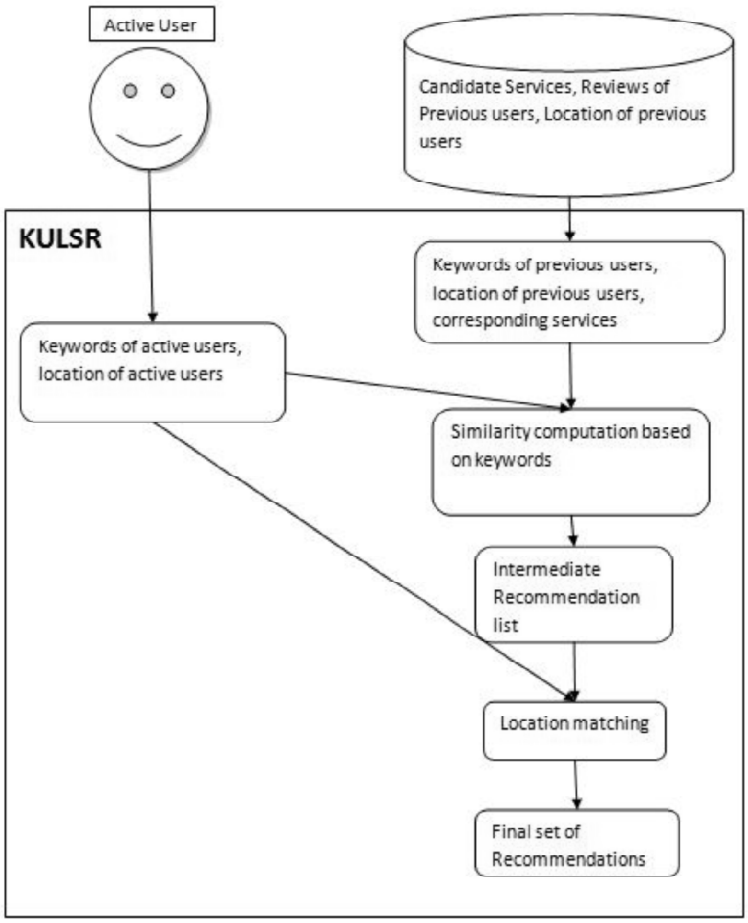

Figure 2: Steps of KULSR

Similarity (APK,PPK) = |APK^PPK|÷|APK∨PPK|

For this, the matched number of keywords is divided by a total number of keywords in a review of previous user and the active user. The calculated value for each of the previous users is calculated, and the intermediate recommendation list is generated by Jaccard coefficient (those Candidate services are included which are having the value of coefficient greater than a threshold value).

4. Location matching: Those candidate services (List of hotels) of previous users are considered which has the user locality value same as the active user.

Algorithm 1: Algorithm of KULSR method

Input: The Review data set of the active user APK
Candidate services CS i
Location of active user LOC a
The review dataset of previous user PPK i
Location dataset of previous users LOC p

Output: Recommendation set for active user

```
for each review from PPKi
    km=0,Tk=0
        If value of keywords in APK matches with keywords in each review
            Then keywords matched: Km= matched keywords
        End if
    Total keywords: Tk =Tk+Km
    Similarity value=Km/Tk
        If similarity value > threshold value
            Insert those CSi in Recommendation set
        End if
End for loop
for each element in recommendation set
    If LOC a equals LOC p
        Insert CSi into Output Recommendation set
    End if
End for loop
If Output Recommendation system not equal to 'null'
    Then display Output Recommendation set
    else
    display Recommendation set
End if
```

IV. EXPERIMENTAL EVALUATION

In this section, outcomes of KULSR method (with location) were compared with KASR method (without considering user location). The results were evaluated by considering nearly a constant value of number of keywords, in most of the cases 3 keywords (domains) are selected, the results can be shown in the table below

Then graphs are plotted for each condition. The graph shown below represents the recommendation results using KULSR method. Calculation of centroids of two outputs is done be applying k-means clustering algorithm for a single cluster, and we obtain the centroid values as follows.

Table 2
Experimental results

	chosen_domain	*without_location*	*with_location*
1	3	29	3
2	3	18	13
3	3	25	20
4	3	3	2
5	3	27	21
6	3	2	1
7	3	5	4
8	4	2	2
9	2	27	24
10	2	16	10
11	2	21	15

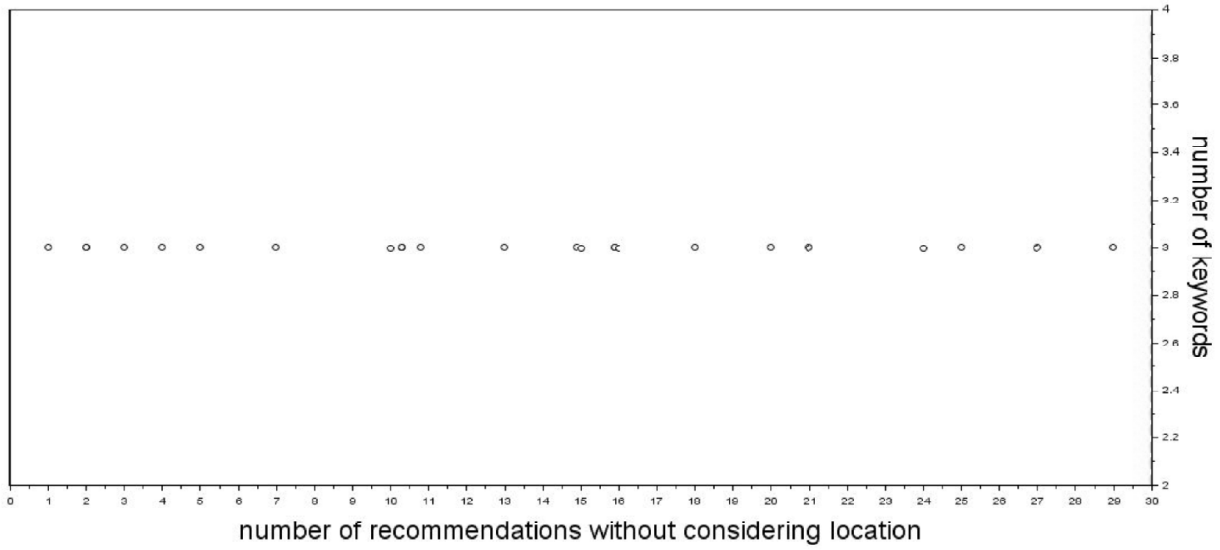

Figure 3: Graph plotted for outputs of KASR Method

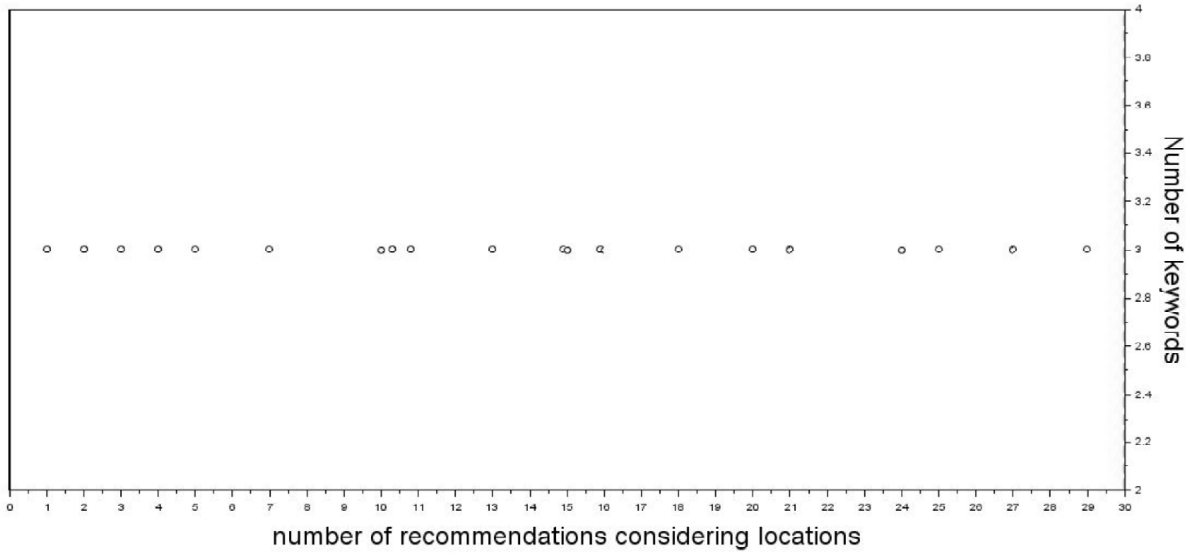

Figure 4: Graph plotted for outputs of KULSR Method

In the figures, the x-axis denotes the number of recommendations by keeping the y-axis nearly constant which shows the number of keywords which were opted by users i.e. those are the selected number of preferences by the active user.

Experimental results are evaluated using The Scilab editor.Metric used for comparing both the methods is to calculate the number of recommendations after applying both the algorithms and then comparing the mean of the density distribution in both the methods. In KULSR method, the centroid appears nearer to the origin as if it is compared to the KASR method, which is sufficient to show the accuracy increased by KULSR method.

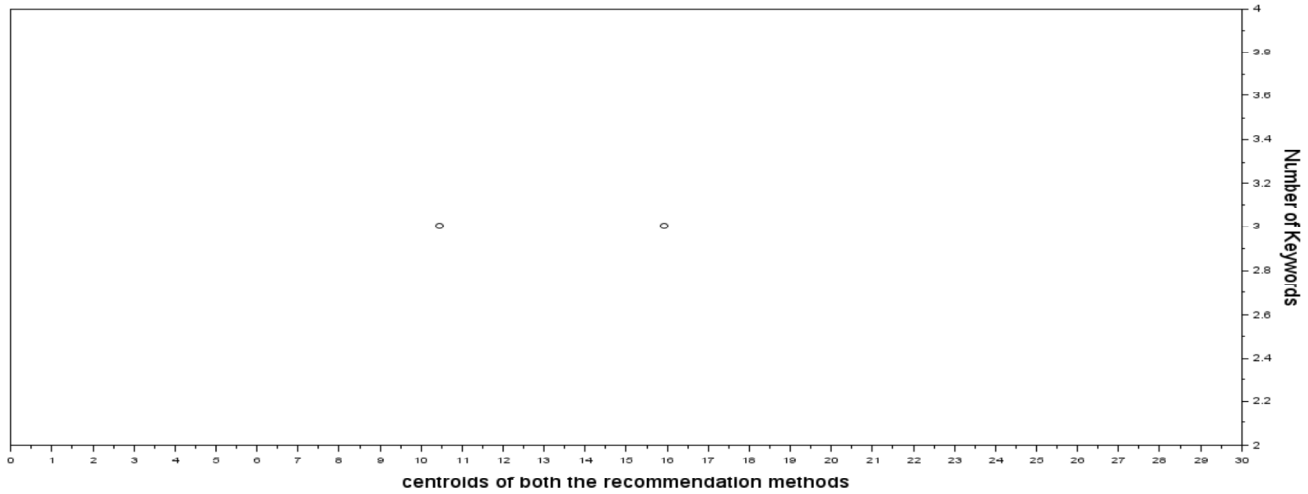

Figure 5: Graph to show average difference between results of two output set

The centroid obtained for the number of recommendations using KASR method =15.90909, and centroid for the number ofrecommendations after applying KULSR method= 10.45455, so it can be stated that the proposed system is more efficient than KASR by a factor of 15.90909-10.45455 =5.45454.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a service recommendation system which is based on keywords and user location, named KULSR. In KULSR, keywords are the domains of user and user location is used to implement a user-based collaborative filtering algorithm which is meant for providing appropriate recommendations. This paper demonstrates that how KASR method can be modified to provide more accurate results after considering both the factors keywords and user location. Similar to any research regarding recommendation system this system also aims to provide more appropriate recommendation system. Finally, the experimental results demonstrate that KULSR improves the accuracy of KASR method.

In the future, we can categorize particular users by age groups to enhance the recommendation system.

REFERENCES

[1] Shunmei Meng, Wanchun Dou, and Xuyun Zhang and Jinchun Chen, “KASR: A Keyword-Aware Service RecommendationMethod on MapReduce for Big Data Applications, vol. 25, No. 12, December 2014.

[2] G.Linden, B. Smith, and J. York, “ Amazon.com Recommendations: Item-to-Item Collaborative Filtering, “ IEEE Internet Computing, vol.7, no.1, January 2003.

[3] M. Deshpande and G. Karypis,”Item-Based Top-n Recommendation”, ACM Transaction.Information System, vol. 22, no.1, pp.143-177,2004

[4] J.S. Breese, D.Heckerman, and C. Kadie, “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” Proc. 14th annual conference, Uncertainty in Artificial Intelligence, pp. 43-52,1998.

[5] C. Basu, H. Hirsh, and W.Cohen, "Recommendation as Classification: Using Social and Content-Based Information in Recommendation", Recommender Systems. AAAI Press 1998.

[6] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl,"Grouplens: An Open Architecture for Collaborative Filtering of Netnews,"Proc. ACM Conference Computer Supported Coop. Work, pp. 175-186,1994.

[7] G. Adomavius, A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions".