ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

Personalized Service Recommendation System Using Data Analytics

Honeytta Kunjachan, Hareesh MJ, Sreedevi KM
Computer Science Department, KTU University
Kerala, India
brijithoney@gmail.com
sreedevikunnathmana@gmail.com
hareeshmjoseph@gmail.com

Abstract—Recommendation systems plays a vital role in today's Scenario. Recommendations help online users to purchase or to find their choices more easily. In last two decades the online customers, services has been increasing rapidly and this creates a great problem in big data analysis. Traditional Recommendation Systems often give same recommendations to different users without considering personal interest and they suffer from scalability when processing large scale data. In this paper we implemented a personalized service recommendation system on hadoop platform to improve the scalability and to enhance the efficiency.

Key words: Recommendations, Map Reduce, Big Data, keyword, Similarity measure.

I. INTRODUCTION

Big data refers to massive amount of data. The working requirements of big data is exactly the same as for data sets of any size. However the scalability, the processing speed, the characteristics that has to be dealt at each stage makes it unique. The main goal of big data refers to handle huge quantity of heterogeneous data that could not be handled using conventional methods. 4 V's of Big data are

1) Volume - Volume presents the most important challenge to many conventional IT structures. Many companies may have large amount of data in the form of logs but they do not have the technology or capacity to process massive amount of data. Ability to process chunks of data is the main benefit of big data analytics.
2) Velocity - It refers to the speed at which data is gener-ated.
3) Variety - Big data can handle any kinds of data, structured, semi-structured or unstructured. Structured refers to relational data. Semi-structured can be XML documents and unstructured can be social media data.
4) Veracity - Veracity refers to quality of data being pro-cessed[4].

Big data also has high impacts on service recommendation systems. With increasing number of alternative solutions, recommending services can be considered as a research issue. Recommendations can be of any type like dress materials, electronic gadgets, books etc. Recommending services can be either based on collaborative filtering or content based filter-ing. Some Recommendation systems also use combinations of both, known as Hybrid Recommendation systems.

Collaborative Recommendation System - In collaborative filtering approach services are recommended to each user, considering the users with similar taste preferred in the past. It is based on the key note that customers who liked in the past will like in future and they will prefer similar kinds of items as in the past. For instance, if user L and user M have a similar purchase history and user L has recently bought an item that user M has not yet purchased, the basic strategy is to recommend this item to user M. Collaborative filtering can be classified into memory based and model based.

Item based Approach - Item based collaborative filtering is an example for model based CF. In this method rating matrix is used to calculate the similarities between items and based on this similarity users preference for an item that is not rated by him/her is calculated.

User based Approach - In user based CF, similarity between two customers is considered. User plays the important role here. Users with the same taste are mapped into one. Recommendations are given to each user based on the items purchased by the other users in the same group with similar prefer-ences[2][7].

Content based Recommendation System - In content based filtering method services are recommended similar to those the user preferred in the past. In this method a profile for each user is created which has the information regarding his/her taste which is based on how a user rates these items. The items which are highly rated by the users are recommended[3].

II PROPOSED WORK

Hybrid Recommendation System - It is a combination of both collaborative and content based Approach. It helps to overcome Cold-Start Problem[2][6].

In conventional methods the services provided to each customer is the same. Different users may have different preferences. For example one may prefer a hotel with wide variety of delicious food in a rural area and while the other may prefer a hotel which is next to a shopping mall or town because he/she may be more interested in shopping. Traditional Hotel recommendation systems recommend the same hotel to all the users without considering his/her personal interest. In this hotel recom-mendation system we implemented a method in which user preferences are given the prime importance and the hotels are recommended to each user based on these ratings. The first section gives an introduction about big data, Recommendation systems and motivation of the project. Second section deals with preliminary knowledge and third section deals with the proposed work.

II. PRELIMINARIES

A. Map Reduce

It is a parallel programming framework which can handle huge amount of data,in parallel. In map reduce, each task is called the job and it usually divides the input data set into chunks of data and gives to different machines running in parallel. The output from the map phase is sorted, which is then given as the input to the reduce tasks. Depending on different applications the number of map tasks and reducer tasks can vary. It also takes care of scheduling and monitoring of the tasks. The input and output of the map reduce is a <key, value> pair. For example if we have a input dataset with 3 rows{Deer, Bear, River}, {Car, Car, River}, {Deer, Car, Bear}. In the first step the dataset is splitted and given to different nodes. Suppose First node receives {Deer, Bear, River} second node receives {Car, Car, River} and the third node receives {Deer, Car, Bear}. In the mapping phase each word is converted to a <key,value> pair. <Deer,1 >, <Bear,1 >, <River,1 >,<Car,1 >,<Car,1 >,<River,1 >,<Deer,1 >,<Car,1 >, <Bear,1 >. In the shuffling process datas are ordered in alphabetical order. In the reduction phase Bear is assigned 2, Car is assigned 3, Deer is assigned 2 and river is assigned 2. Final result produces the following <key, value> pairs <Bear,2 >,<Car,3 >, <Deer,2 >,<River,2 >[8].

B. Porter Stemmer Algorithm

Porter Stemmer algorithm is used for removing the suf-fixes from English words. This plays an important role in Information retrieval. For instance Distribute, Distribution, Distributed, Distributions, Distributed. By porter stemmer al-gorithm ed,ing,ion,ions are removed and all are converted into its stem word. stemming helps in reducing the size of the data as well as its complexity.

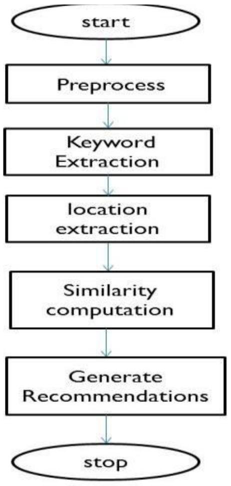
A. Algorithm

```
Input:Keywords of active user,Location referred by
user,Preferences of active user
Output:Top K Recommended Hotels with highest
Ratings Obtain the keywords of Active user(Ak)
while  Each Keyword do
        if Ak is present in the review of the
        previous user(Pk) with same location then
             Select the review;
        else
             loop for next review;
        end
       Compute the similarity of each preference
        rating of Active user with the previous user
        rating.
     If  Rating Similarity is greater than
         threshold then Sort the hotels in
         descending order;
     else
         Discard the hotel;
     end
```

Provide top K Recommendations end

Algorithm 1: Algorithm for Hotel Recommendation

B. Flow Chart



C. Methodology

1) Preprocess: Data set is a hotel data set consisting of the reviews from various customers. First HTML tags and stop words are removed from the reviews collected from the data set. Porter Stemmer algorithm is used to extract the stem of the item.

2) Keyword Extraction: In this step we consider two kinds of users. One is the active user who needs the recommendation and other is the previous user whose reviews are collected to provide the recommendations.

Preferences provided by Active User : An active user can give his/her preferences about the services. An active user can give preference's regarding service, cleanliness, overall quality of the hotel, Rooms, cleanliness regarding the location where the hotel is located, Sleep Quality, food provided, Customer service. Ratings are given in Table 1:

TABLE I RATINGS

Rating	Importance
1	very poor
2	poor
3	ok
4	good
5	excellent

Preferences of the Previous Users: The preference of a previous user for a service is extracted from the reviews. In this phase each review will be converted into a corresponding keyword set according to the Main keywords list and domain thesaurus.

Main keyword list - Each review is converted into a set of keyword candidate list. It is a set of keywords about user preferences. Usually since few words in reviews does not exactly match the corresponding keywords present in the main keyword list which represents the same aspects as the words. The corresponding keywords should be also extracted[1].

Domain thesaurus - It is actually a reference work of the main keyword list. In this list the words the same meaning are grouped into one category. For instance Transportation can be considered as a main keyword, then items like subway, bus, stop all those comes under its thesaurus.Similarly mall, shopping, store, market all come under the category of Shopping[1].Fig 1 describes the thesaurus for fitness. All the keywords Yoga, Gym, Spa, Fitness, Swimming, Pool can be Incorporated under the Main Keyword Fitness.

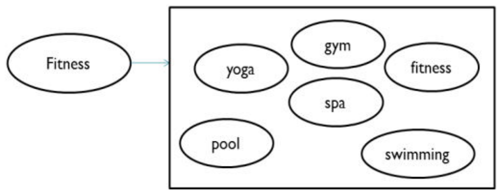


Fig. 1. Domain thesaurus for Fitness

3) Location Extraction: In location Extraction active user can select the location in which he/she wants the recom-mendation(Fig.2). The location given by the active user is mapped with the location of the previous user to provide the recommendation of hotels in that area.

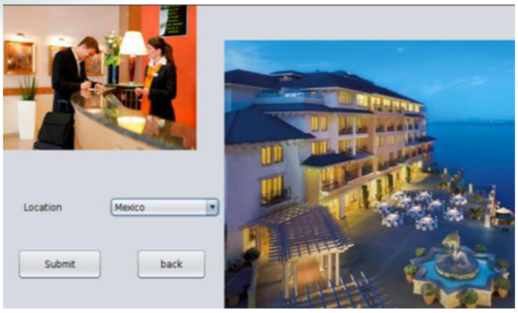

Location Mexico

Fig. 2. Location Extraction From Active User

4) Similarity Computation: In this step, users with the similar tastes are identified. Before the similarity is computed if there is no similarity in active users preferences with the previous users those reviews are initially filtered out. This is done by the intercept concept in the set theory. Initially we check the preference keyword set of the active user with the preference keyword sets of the previous users. If the intersection is an empty set then those reviews are not considered. Jaccard coefficient is used for the similarity computation method.

Jaccard coefficient is also known as Jaccard index. It is a measurement of asymmetric information on binary variables as well as non-binary variables and it is useful when negative values gives no information. It is represented as follows[1].

$$similarity(Ak, Pk) = Jaccard(Ak, Pk) \qquad (1)$$

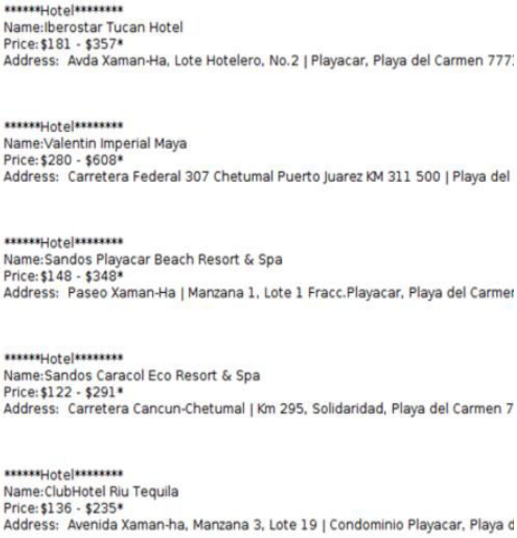
$$Jaccard(Ak, Pk) = \frac{Ak \cap Pk}{Ak \cup Pk} \qquad (2)$$

where Ak is the preference of the active user and Pk is the preference of the previous user.

5) Generating Recommendations: In this phase further fil-tering is done. Based on a threshold value again the recom-mended hotels are filtered. A threshold^1 ° value is taken into consideration and it is verified similarity(Ak,Pk) is greater than the threshold . Only those hotels whose similarity is greater than the threshold value is recommended to the user. The list of hotels are again sorted according to their similarity value such that the hotels with highest similarity appears first. Finally, list of highest rating hotels are provided to the users which satisfy their personal interest.

IV. RESULTS

To improve the scalability and efficiency this is implemented in hadoop platform. In the first step each keyword of the active user is compared with the previous users and the similarity is computed and the hotels are sorted according to their ratings.



```
******Hotel********
Name:Iberostar Tucan Hotel
Price:$181 - $357*
Address:  Avda Xaman-Ha, Lote Hotelero, No.2 | Playacar, Playa del Carmen 777

******Hotel********
Name:Valentin Imperial Maya
Price:$280 - $608*
Address:  Carretera Federal 307 Chetumal Puerto Juarez KM 311 500 | Playa del

******Hotel********
Name:Sandos Playacar Beach Resort & Spa
Price:$148 - $348*
Address:  Paseo Xaman-Ha | Manzana 1, Lote 1 Fracc.Playacar, Playa del Carmer

******Hotel********
Name:Sandos Caracol Eco Resort & Spa
Price:$122 - $291*
Address:  Carretera Cancun-Chetumal | Km 295, Solidaridad, Playa del Carmen 7

******Hotel********
Name:ClubHotel Riu Tequila
Price:$136 - $235*
Address:  Avenida Xaman-ha, Manzana 3, Lote 19 | Condominio Playacar, Playa d
```

Fig. 3. List of Hotels Recommended to Active User

V. CONCLUSION

In this paper we proposed a new method to find the recommendations for hotels based on their location, keywords and ratings. Collaborative filtering method is adopted to provides the recommendations. Active user can give his/her preferences and based on these preferences, reviews are filtered out and the similarity measure is computed based on the ratings. Moreover to improve the efficiency and scalabilty we have implemented in hadoop map reduce framework.

VI. FUTURE WORK

In future we would like to extend this algorithm to different clusters and evaluate the time and memory requirements.

Future work can also include how to distinguish between positive and negative preferences of the users and to make the recommendation more accurate. We would also like to implement this using Apache spark and also to try in HPC platform.

REFERENCES

[1] Shunmei Meng, Wanchun Dou, Xuyun Zhang,Jinjun Chen "KASR: A keyword Aware service Recommendation method on Map Reduce for BigData Applications", IEEE transactions on parallel and Distributed Systems, vol. 25, no . 12, 2014.

[2] S.Pandya, J.Shah, N.Joshi, H.Ghayvat,S.C.MUkhopadhyay, M.H Yap "A

Novel Hybrid based Recommendation System based on Clustering and Association Mining", International Conference on Sensing Technology, 2016 IEEE.

[3] Q.Wang, W.Cao and Y.Liu, "A Novel Clustering based Collaborative Filtering Recommendation system Algorithm", An advanced Technologies, Embedded and Multimedia for Human-centeric Computing,Springer journal 2014

[4] Kamalpreet Singh, Ravinder Kaur, "Hadoop Addressing Challenges of BigData," 2014 IEEE

[5] E.Sivaraman, Dr.R.Manickachezia "High Performance and Fault Tolerant Distributed File System for Big Data Storage and Processing using Hadoop " International Conference on Intelligent Computing Applications, 2014 IEEE

[6] "Multi criteria User Modeling in Recommender Systems",2011 IEEE

[7] M. Balabanovic and Y. Shoham, âA˘ IJFab: Content-Based, Collaborative Recommendationâ˘A˙I, Comm. ACM, vol. 40, no. 3, pp. 66-72,1997.

[8] Hamoud Alshammari, Jeongkyu Lee and Hassan Bajwa " H2Hadoop: Improving Hadoop Performance using the Metadata of Related Jobs" IEEE Transactions on Cloud Computing, 2015