RESEARCH ARTICLE OPEN ACCESS

Enhance Keyword Based Service Recommendation Using Big Data Analytics

Kavita Mandlik [1], Dr. Pramod S Nair [2], D.Srinivasa Rao [3]
Research Scholar [1], Professor [2], Associate Professor [3]
Department of Computer Science and Engineering
MITM, Indore
M.P - India

ABSTRACT

Big data technique not only offers the storage of large amount of data it also provides the ability to process or analyse the data. This domain of information processing is called the big data analytics. The big data analytics are used in various applications now in these days such as social media, e-commerce and others for fast information processing and delivery. The big data analytics is a combination of data mining techniques and big data capabilities for data processing and storage. In this presented work an application of big data mining is presented. This application is used for service recommendation. When a number of service providers for the same service are available in large amount the evaluation of service quality is required on basis of their previous clients. With this concept a hotel service recommendation model is proposed for design and implementation. This service recommendation work first based on the user provided attribute based search and then that is filtered and ranked according to the available user reviews. For computing the review scores the weighted technique based on the service quality and reviews are computed and demonstrated. The implementation of the proposed approach is provided on the basis of Hadoop and JAVA technology. Additionally the performance of the system is described in terms of search accuracy and the time and space complexity. The experimental results with the increasing amount of data indicate the performance of the system is consistent and acceptable for real world application and their utilization in different areas of applications.

Keywords:- Big Data, recommendation system, big data analytics, Keyword based search, Optimization.

I. INTRODUCTION

Service recommender systems have been shown as valuable tools for providing appropriate recommendations to users. In the last decade, the amount of customers, services and online information has grown rapidly, yielding the big data analysis problem for service recommender systems [1]. Currently, online searching process increases and people searches new information in the search process. Most of the search engine gives additional supporting information. Recommender system involves in this process and implements as service. Service recommender system gives additional information to the user but if information grows then these process become a critical one [2].

The explosive increase in the amount of data in our world challenges IT industry and academia. The large data-sets called big-data provides new opportunities in service recommendation. To analyze huge data-set, traditional recommendation services suffer problems of efficiency and scalability [3]. Hence, a recommender system analyzes on available data-set in order to make suggestions to consumers about services and products that he might be very interested. It guides the user with large number of possible options about interesting and useful services or products [4]. Therefore, in this project we present a new approach of service recommendation based on keyword knowledge which give ensure that efficiency and scalability of the developed recommendation system for end user applications.

II.PROPOSED WORK

This chapter provides the detailed description of the proposed system modelling for providing the appropriate service recommendation system. In order to describe the functional aspects of the proposed model and their working this chapter includes the system overview, methodology of system and the algorithm of the system.

A. System Overview

Recommendation system provides the faculty to the end user for selection of best and optimal content, service or products among various available services and products. That is a technique of data mining which is used with some kinds of predictive and associative pattern mining. In this presented work the service recommendation system is aimed too developed. Because, now in these days a number of service providers are exists for providing same or similar kinds of services with the different prices. Additionally the quality of the service is also a major concern on selection of the appropriate and actually desired services [5]. In order to understand the basic issue behind the proposed work let us an example: suppose an online user wants to visit in a place says "Indore". This user is in search of a hotel for their rest and

with some parking space and a restaurant within the hotel. In search process a number hotels are listed and some of them are very costly and some of them and not much appropriate according to the requirements. In addition of that the reviews about the required both services are also suitable to hire the hotels.

In this context a weighted technique is proposed for finding the recommendations of the top rated hotels and offered services by the hotel service providers. In this context a keyword filter based technique is used which works on the basis of weighted score based technique. Additionally only those service providers are suggested that high relevant and provides the actually valuable services. Thus a criteria about the based on user reviews and ratings are prepared that is termed here as the threshold and one the basis of computed scores the most suitable hotels and services are recommended. This section provides the overview about the proposed work and the next section includes the detailed working and functional aspects of the work.

B. Methodology

The proposed technique for finding appropriate ranked service and with the application of hotel service recommendation is demonstrated on figure 2.1. There different components which are utilized in this system is described in this section in detail.

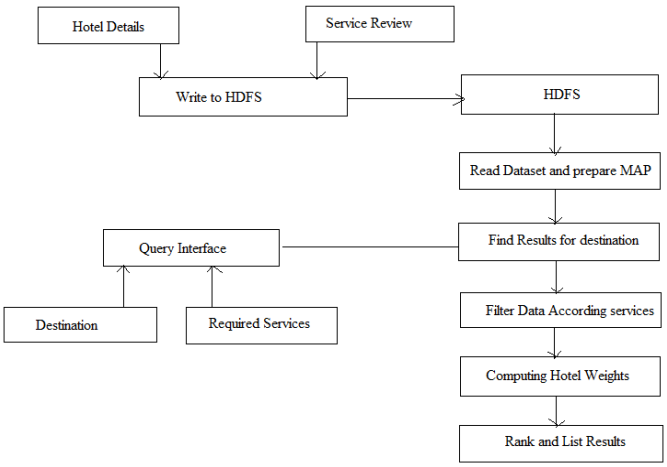

Figure 2.1 proposed system architecture

Dataset input: in order to work with any data model a data source is required which is used to process using the implemented algorithms. Thus a dataset using the observations on different hotel search websites are created for four different cities. Every city contains a number of different hotels thus a part of data set is created which contains the general information about the hotels such as the name, phone number, address and others. In addition of that a different set of data is also prepared which contains the list of services offered by the hotels and their reviews provided by the users.

Write to HDFS: the selected dataset from the local file system is needed to be place on the Hadoop file system. This enable the big data to process the data and utilize with the prepared search algorithm.

HDFS: basically the HDFS is also termed as Hadoop File System. That is developed using distributed file system environment. As compared to other kinds of distributed systems HDFS is designed for low-cost hardware resources. That is enabled for carrying very large amount of data and also allow to easy access. In order to store large amount of data files can be stored on multiple machines. That file system also makes applications for parallel processing.

Read dataset and prepare MAP: after placing data on the HDFS file system the data is enabled to read by the application. Therefore a separate program is written for reading the file and prepares a data structure that makes enable to perform the search query on that data.

User input: in the similar manner as the initial dataset is prepared and executed with the help of big data environment. That part of data is assumed as the data source for making the search. On the other hand the user input is required in this phase need to prepare query for finding the data from the initial input data. That query data is also prepared in two parts first the city which is needed to be visited by the end user and second the services which is required by the service provider.

Query interface: that is the key user interface which is designed to interact with the end user. Using this user interface the user can provide input the keywords and performs the search operation for finding the required outcomes.

Find results for destination: this phase utilizes the first user input and compute the hotels in the particular place. Therefore in order to compute the hotels according to the given place is computed using the following process:

Input: data structure D_n, User Place to visit P Output: List of hotels H_L
Process: 1. $for(i = 1; i < n; i++)$ a. $diff = \sqrt{D_i^2 - P^2}$ b. $if(diff == 0)$ i. $H_L.Add(D_i)$ c. $end\ if$ 2. End for 3. Return H_L

Table 2.1 city search process

Filter data according to services: now the filtered list of similar place which is returned in previous process H_L is used to refine the data. Thus the second user input is used in this

phase for finding the required keyword based hotels. The process of filtering data is given using the table 3.2.

Input: list of hotels H_L, list of services required S_R Output: filtered list of hotels based on services H_s
Process: 1. $for(i = 1; i \leq R; i++)$ a. $Ser = S_i$ b. $for(j = 1; j \leq L; j++)$ i. $if\left(H_j.contains(Ser)\right)$ 1. $H_s.Add(H_j)$ ii. $end\ if$ c. End for 2. End for

Table 2.2 filtering hotels

Computing hotels weight: in order to find the best service hotels among the available or shortlisted hotels we are considering the user reviews about the particular services. In the prepared data set the services are categorized in six parts. Thus in order to generate the scoring of the services we map these reviews according to the numerical values. The table 2.3 contains the available review and there mapping values

Excellent	5
Very good	4
Good	3
Ok	2
Poor	1
N/A	0

Table 2.3 mapping values

In addition of that for obtaining the weights among the 0-1 the normalization of mapping values are performed. For normalizing the values the following formula is used:

$$new\ value = \frac{current\ value - minimum\ value}{maximum\ value - minimum\ value}$$

Now need to encode the service using the mapping values which are shortlisted in the hotels list H_s according to the service availability. The encoding process is given as follows:

Input: shortlisted hotels H_s, service mapping M<key, value> Output: encoded service list E_s
Process: 1. $for(i = 1; i \leq s; i++)$ a. $if\left(H_s.contains(M.key)\right)$ i. $E_s = H_s.replace(Key, M.value)$ b. End if 2. End for

Table 2.4 encoding process

After encoding of the service reviews according to the mapping table the weight computation is performed thus in order to compute the weights the following formula is used:

$$W = \frac{1}{S}\sum_{i=1}^{S} E_i$$

The weights are associated with the hotels name according to the required services.

Rank and list results: after the computing of weights the hotels are ranked according to the obtained weight values and the higher weighted values are ranked first. In addition of that a threshold is also created for finding more precise values about the hotels review. This threshold value is 30% of the higher weight computed for the services.

III. RESULTS ANALYSIS

This chapter provides the details about the performance evaluation of the proposed service recommendation system. Thus this section includes the computed parameters by which the performance of the system is indicated.

A. Search Accuracy

The search accuracy is the measurement of correctness of the search outcomes produced using the given search engine. The search accuracy of the system is computed using the following formula:

$$search\ accuracy = \frac{actual\ correct\ results}{expected\ results} X100$$

Dataset Size	Search Accuracy (%)
100	82
200	87.5
300	88.3

500	89.42
700	92.46
1000	91.85
1200	92.77

Table 3.1 search accuracy

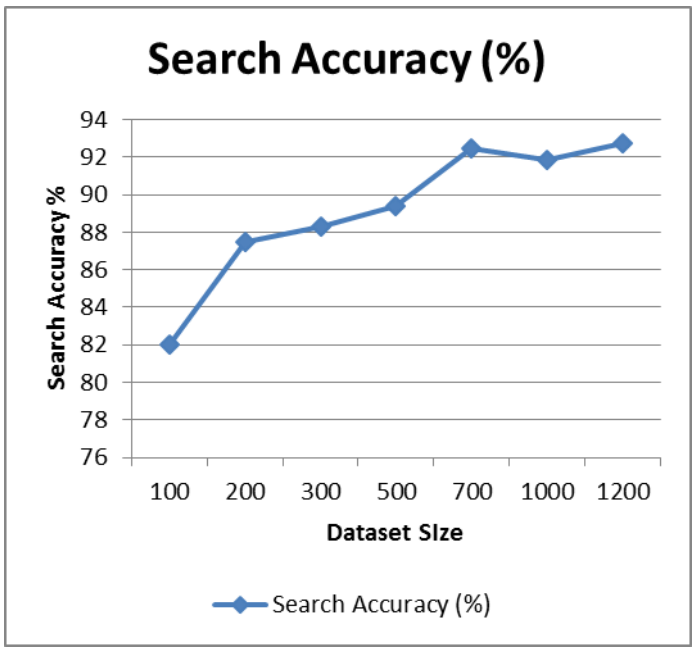

Figure 3.1 search accuracy

The table 3.1 and figure 3.1 contains the amount of search results generated by the proposed system for finding the user constraint and key word based search outcomes. The X axis indicates the total size of dataset and the Y axis shows the amount of correct outcomes obtained during user search. The accuracy of the search system is measured in percentage correct results over the total expected results. According to the obtained results the amount of correct results generation is increases with the amount of data, but after sometimes it provides the consistent results based on the different query supplied by the users. Therefore the proposed technique is suitable for accurate and consistent outcomes with large amount of data.

B. Memory Usages

The processes need an amount of main memory for execution of the current task. Additionally that is assigned dynamically according to the requirements of processes. The memory usages of the process or algorithm also termed as the memory consumption or the space complexity of algorithms. The memory requirements of the algorithm are computed using the following formula:

$$memory\ usage = total\ assigned\ memory - total\ free\ space$$

Dataset Size	Memory Usage (KB)
100	25371
200	25946
300	26514
500	27018
700	27847
1000	28814
1200	29462

Table 3.2 memory usages

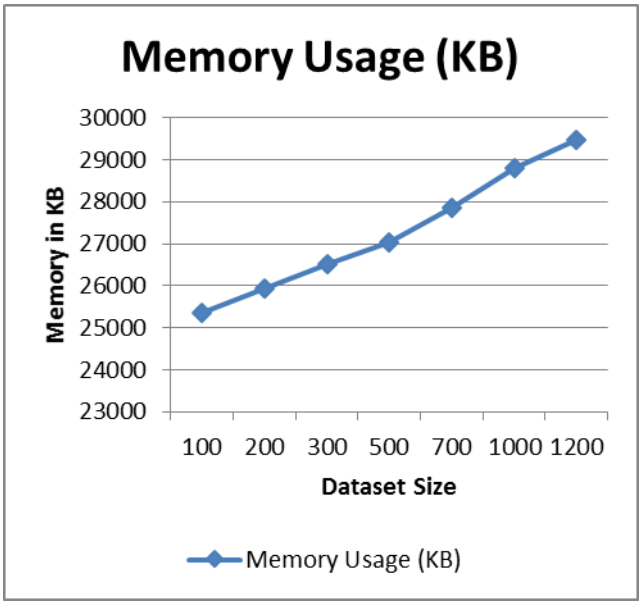

Figure 3.2 memory usages

The amount of space requirement of the proposed service recommendation model is demonstrated using table 3.2 and figure 3.2. The table provides the values of memory requirements in KB (kilobytes) and figure visualize the pattern of requirements on the basis of values. In this figure X axis includes the dataset size and the Y axis shows the corresponding memory usages. According to the obtained performance the requirements of the system is increases as the amount of data increases for processing. But according to the variations observed the amount of requirements are considerable according to the data size.

C. Time Consumption

The processes take an amount of time for processing the input data according to the algorithms evaluation. This time

requirement of the algorithm is termed as the time consumption of algorithm or the time complexity of the system. The time of the search processing is computed using the following formula:

$$time\ consumption = result\ genration\ time - keyword\ submission\ time$$

Dataset Size	Time consumption (MS)
100	28
200	42
300	70
500	116
700	143
1000	191
1200	237

Table 3.3 time consumption

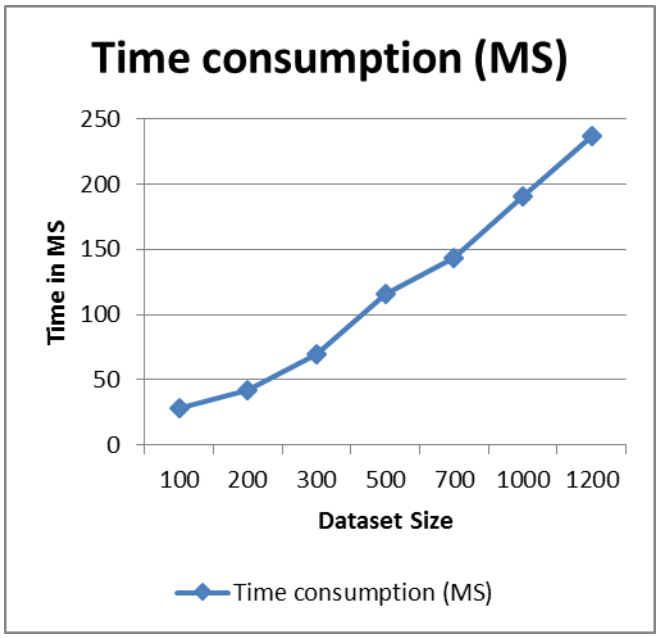

Figure 3.3 time consumption

The time requirement of the proposed service recommendation model is provided in figure 3.3 and table 3.3. Table includes the two attributes first the size of data for processing and second is the amount of time required for performing the evaluation. The time is measured here in terms of milliseconds (MS). The X axis of figure includes the dataset size and the Y axis provides the time requirements to visualize the graphs. According to the obtained performance the amount of time is increases with the amount of data increases for the analysis.

IV. CONCLUSION AND FUTURE WORK

This chapter provides the complete summary of the work involved in this presented work. in order to provide the summary the design methodology and experimental outcomes are considered. In addition of that possible future scope of the work is also included with the chapter.

A. Conclusion

The big data technology is a combination of hardware and software for improving the performance of the traditional computing. This technique is usages the concept of parallel and distributed computing concepts to deal with the large amount of data storage and processing. In addition of that when the large amount of data exist in different area of applications where fast processing and accurate information required we can implement the technique for automated data analytics. The data analytics domains allow us to implement the data mining and data analysis algorithms for discovering the required patterns from huge data sources. Therefore that is the effort of computational efforts and the big data capability for data analysis and obtaining required outcomes. In order to demonstrate the effectiveness of the big data analytics a service recommendation model is proposed in this work.

Now in these days for service requirements a number of service providers are available. Additionally they claim to provide low cost and interactive services to their clients. The evaluation of the services based on the reviews of the services offered by the service provider is the main aim of the proposed work. Therefore the proposed technique involves a simulation of the hotels recommendation based on the offered services and their client's review. Thus a three phase filter is implemented in this work. In first by the user visit the place the filtering on data is performed. In next the priority based on user keyword selection or service requirements the filtering is performed. Additionally in final phase the weighted process is used for finding the score of reviews by which the ranking of the hotel and service provider is sorted or ranked for results demonstration.

The implementation of the proposed work is performed on the basis of big data environment therefore the hosting of data is provided using the Hadoop and for implementation of the functional aspects of the proposed work the JAVA technology is used. After the implementation of the proposed service recommendation system the performance is computed and summarized in table 4.1.

S. No.	Parameters	Remark
1	Search accuracy	The highly precise results are produced and below a threshold level of reviews is not suggested to the end client additionally performance becomes

		consistent 90-94%
2	Time consumption	The amount of data increases the time of processing but that is acceptable
3	Memory usages	Memory is also increases with the amount of data produced for search but acceptable for accurate processing of data

Table 4.1 performance summary

According to the obtained performance the proposed technique is accurate and efficient for recommending the user defined services for the high quality. Additionally with low resource requirement and efficiency that execute the required task. Thus the proposed model is acceptable for the real world applications and recommendation system designs.

B. FUTURE WORK

The main aim of the proposed work is to enhance the traditional key word based service recommendation system design is achieved successfully. In near future the following future extension is possible.

1. The current system is works only limited amount of cities data in near future the dataset need to be increase
2. Currently the service review is used in terms of limited categories in near future the text reviews are used and based on sentiment analysis the reviews are produced
3. The proposed system currently works for a single domain of service recommendation in near future that is extended for multiple domain acceptance.

REFERENCES

[1] Meng, Shunmei, et al. "KASR: A Keyword-Aware Service Recommendation method on MapReduce for big data applications", IEEE Transactions on Parallel and Distributed Systems 25.12 (2014): 3221-3231.

[2] GK, Kiran Kumar Bathi, and N. Sheela, "Personalized Recommender System", PARIPEX-Indian Journal of Research 4.8 (2016)

[3] Priyank Jain and Manasi Gyanchandani, "A Survey on Big Data privacy using Hadoop Architecture", IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.2, February 2017

[4] Pu P, Chen L, Hu R, A user-centric evaluation framework for recommender systems. In: Proceedings of the fifth ACM conference on Recommender Systems (RecSys'11), ACM, New York, NY, USA; 2011. p. 57–164

[5] Chen, Yan-Ying, An-Jung Cheng, and Winston H. Hsu, "Travel recommendation by mining people attributes and travel group types from community-contributed photos", IEEE Transactions on Multimedia 15.6 (2013): pp. 1283-1295