

Fun Ride Applications for like Based Approach in Cloud and Big Data

G. Kumaresan, A. Samyurai, R. Vignesh and S. Dhivya

Department of Computer Science and Engineering,
Valliammai Engineering College, Tamil Nadu, India

Abstract: Massive growth in services is emerging in the world. There is an impact in extracting information from requested services. Clustering and Collaborative Filtering (ClubCF) is a technique used to extract information in a stimulated time. Cluster is used to group the similar set of services in the same cluster and is unique in nature. Filtration is done based on user request and behavior. It produces all the recommended services needed by the user. But these services will not be based on user interest. To satisfy this Like-base approach is used, where it produces only the user interested services and what they needed. These are implemented based on their user interest and their behavior.

Key words: Bigdata • Clustering • Collaborative Filtering • Like-based approach

INTRODUCTION

Big data has become wide in the world. The Big Data means large volume, Complex, growing data sets with multiplesources. This Big Data appears due to the creation of huge amount of data everyday. The impact leads to extract data will be difficult within an tolerable time.

In the existing system it provides all the services to the users. To extract the user relevant data is the fundamental challenge in bigdata. The major drawback is to produce the recommended services to users is the difficult one since it is the big data applications. It may also lead to a sparsity problem. In this system no techniques are used to provide recommended services to users.

To satisfy the user recommendations Clustering and Collaborative Filtering techniques is used. It is used to reduce the online execution time of collaborative filtering. This may lead the users to make decisions easily from so many services that are provided [1]. Clustering is a task of discovering a set of homogenous group of objects. Data clustering is based on the similarity or dissimilarity between the data. This measure makes the clustering technique meaningful. The high quality of clustering is the technique that is used to obtain the high intra-cluster similarity and low inter-cluster similarity. So Clustering is the step which is used to club the similar set of services

into same clusters which may contain similar like-minded users and Collaborative Filtering techniques are used to filter out the data that are needed by the user. These are done based on the user behavior and their categories [2].

In this paper we proposed an effective technique to provide only the recommended services and the users interested services. The effective approach or technique used is Like Based Approach. These techniques are implemented in the FunRide applications where it will display all the user recommended services.

Cloud computing has been the newly arising concept in the recent years and computing has shifted to the clouds. Cloud Computing is the use of computing resources such as applications and peripherals that are delivered through the network that is internet. Cloud Computing is the universal collection of data which extends over the world. This cloud contain rich functionalities and it contain the number of services. Cloud cannot handle the Big Data applications. Big Data is the collection of Complex growing data sets. Big Data is originated since we are creating large amount of data every day. The data produced everyday is estimated in the form of zettabytes. Big Data is Pervasive and it contains structured and unstructured data, In Structured data, data can be easily analyzed and categorized. In unstructured data, data cannot be easily categorized or analyzed. To do this it relies mainly on the keywords which allows user to filter out the data that are needed

for the user. In Big Data, Data Mining Concepts are used [3]. Data Mining is the concept of analyzing data from different perspectives and summarizing it into useful information. Data Mining is used to look through the big data set to retrieve the relevant information. The Fundamental challenge in Big data is to extract and explore the data. semantic analysis text mining method is used for evaluating and finding the related services [4].

The main disadvantage in this is difficult to capture, manage and handle data within tolerable time and also difficult to provide the recommended services to the customers. The data cannot be processed effectively. To process effectively and to provide the recommended services to users clubcf approach was used. It provides the recommended services to the customers based on their categories.

The rest of the paper is organized as follows. In Section 2, we discuss the related work which are needed to be carried out and the service creation in section 3, clustering in section 4, fun ride application in section 5, algorithm are explained clearly at Section 6. AtSection 7 comparison between existing and proposed in dealt, Finally at section 8.the conclusion and the advantages of our proposed scheme are explained.

Related Work: The existing system had proposed the Clustering and Collaborative Filtering techniques [5]. Clustering is the process of separating the set of objects into clusters. Clustering is done based on the Similarity and Dissimilarity. Similarity Clustering is done based on the high intra-cluster and low inter-cluster similarity. Dissimilarity is done based on the low intra-cluster and high inter-cluster similarity. Hierarchical Algorithm is the best algorithm used for clustering. It is used to represent in tree structure known as Dendrogram [6]. Clustering is done to produce the recommended services easily based on the user behavior. In this Clustering stage porter stemmer algorithm is applied where it is used to create the stem words by suffixing the English words [7]. There are many stemming algorithms like Lovin stemmer, Palice stemmer, Dawson stemmer which are used for different languages. Of all these the strength and accuracy of the porter stemmer algorithm was best and it was applied in this scheme [8]. By using these stemmed services they are clustered under one clusters. Based on the user’s behavior they are clustered and are stored in the database.

The next scheme or technique used is Collaborative Filtering which is applied to produce the user recommended services. Based on the user categories or

behavior the recommended services are filtered out and are displayed to them. This Collaborative filtering technique is done based on the user behavior [9], though it is based on user behavior it produces all the recommended services that are not interested by the user.

Our proposed schemes add to produce only the user recommended service and interested services to the user. These are implemented in the FunRide applications where it is used to display the recommended services to the users.

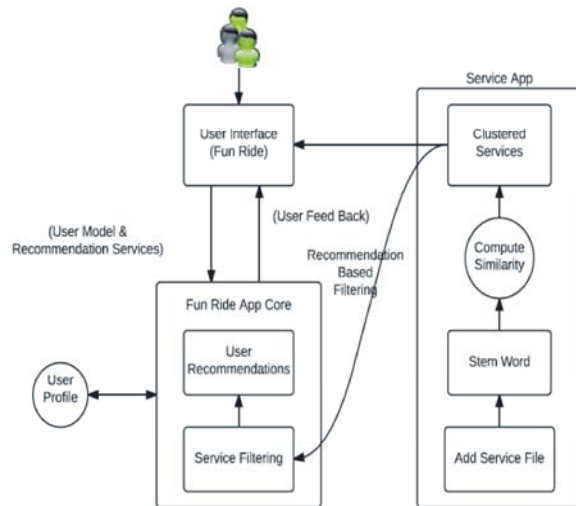


Fig. 1: Architecture of Filtering Recommendations

This architecture shows the overall process that is done in the project. Admin used to login with the help of username and password and they are used to create the service files. First they used to add the service files and then they are used to generate the stem words using porter stemmer algorithm [8].

Once the stem words are generated then similarity is computed between them. After Computing the similarity they are clustered based on the similarity. These are all done in the admin side. They will keep the clustered services and are stored in the database called bigtable [10]. While computing the clubcf the services are retrieved from this bigtable.

In the user side, User has to first register by filling their details. Once they are registered they can login and view their profile. In their profile they view the Recommended services that are filtered in the admin side.

Service Creation: In the Service Creation stage, admin use to log in with the credential and create a service file. Multiple service file can be created with a unique in nature which is used for clustering of large data set. The huge

volume of data is branched with these service file and cluster index. These data are stored in a data table called big table. Once service file is created successfully, they can add the service file to the database.

On adding the service file, tags are generated for key search of data during service request. These tags are mapped with the service file and post image to be browsed and created to post in the service application. Created service files can be viewed and stem words are generated from the tags using Porter Stemmer algorithm.

Service Name: Home and Kitchen Appliances

Keyword: Appliances

Tags: Refrigerators, Air Conditioners, Furnishing, crockery's,

Stem Words: Refrigerator, Air Conditioner, Furnish, Crockery

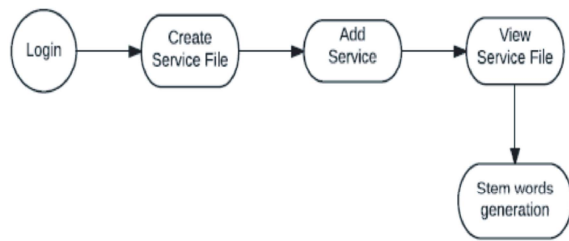


Fig. 2: Creation of service

In above scenario the stem words are generated based on the tags by removing the suffix of the English word. These are done based on the grammatical English words.

Clustering Stage: In the Clustering Stage, admin used to cluster the service by computing the similarity between the services of the stem words. They used to compute the characteristic similarity and the functional similarity of the stem words. The matrix D is created based on the character similarity. The Services are clustered based on the character similarity using the Agglomerative Hierarchical Algorithm [6]. By computing the similarity between the services they are clustered based on the stem words. Cluster contain the similar set of services in one cluster until all the clusters are grouped together into one cluster.

Services: Mechanical Engineering, Aeronautical Engineering, Software Engineering, Computer Science Engineering.

Clustered Services Name: Engineering

Funride Application: The Fun Ride App is the application where user has to register with their details. In this the Gender and the Designation are the mandatory fields. Once they filled all the details the user get registered successfully. After Successful registration user can login with their credentials. After login user can view their profile which contains the recommended services that are needed by the users. Then the user has to like the services which they like and based on that the services are filtered out using the Collaborative Filtering technique. First of all user get the services based on his/her category, After he press like on any of the advertising services user get the recommended services based on that services.

If the user is the new one he has to give the recommended services what he needed. If he is the active user he can view only recommended services what he needed. Then he can add any services what they needed. They can also Send Friend Request to their friends by searching with the help of email id or by their name. They can also chat with their friends and can also able to share the services or can post anything what's on their mind.

Once he finishes everything the user can logout the application. This is mainly used to provide the user recommended services based on their interest on those services.

Table 1: User Interested Services

USER	VIDEO	JEWELS	BOOKS	SOFTWARES
User 1	Like		Like	
User 2	Like	Like		
User 3			Like	Like
User 4	Like	Like	Like	Like

Agglomerative Hierarchical Algorithm: This Algorithm is used for the clustering the services. These are used to cluster the similar services until all the services are clustered into the same clusters. If there are n similar services then these n similar services are to be clustered into the same cluster.

Input: A set of recommendations $R = \{r_1, \dots, r_n\}$, b characteristic similarity matrix $A = [a_{ij}]_{n \times n}$, the number of required clusters N.

Output: Dendrogram_n for n = 1 to |r|.

1. $S_i = \{r_i\}, \forall i;$
2. $a_{ci,cj} = a_{ij}, \forall i,j;$
3. **for** $n=|r|$ **down to** N
4. $Dendrogram_n = \{S_1, \dots, S_n\};$
5. $m,o = \operatorname{argmax}_{i,j} a_{si,sj};$
6. $s_i = \operatorname{join}(s_m, s_o);$
7. **for each** $s_n \in r;$
8. **if** $s_n \neq s_m$ **and** $s_n \neq s_o$
9. $a_{si,sh} = \operatorname{Average}(a_{sm,sh}, a_{so,sh});$
10. **end if**
11. **end for**
12. $r = r - \{s_o\};$
13. **end for**

These are clustered and the output will be in the form of tree structure known as dendrogram which is nothing but the output of the algorithm.

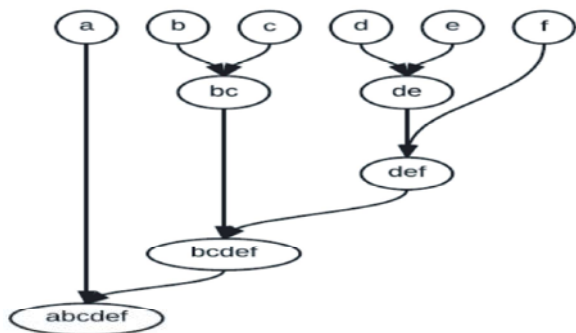


Fig. 3: Agglomerative Clustering Algorithm

Using this algorithm the services are clustered based on the stemwords and they contain the like-minded services.

RESULT

The analysis shows the comparison graph between processing rate and number of services in Figure 3. With the help of Fun Ride Application and the Like based approach the number of services will be available within the elapsed time. So the user can get their recommended services within the elapsed time thereby the performance of the transaction time will be less and the services will be provided easily. Since the number of services is made into cluster, the computation of rating similarity time will be tremendously reduced and by creating service file in the cluster the most likelihood data can be got.

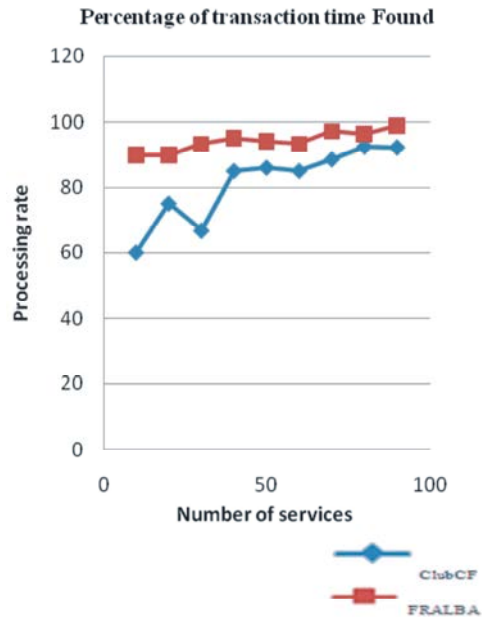


Fig. 3: Comparison graph for Processing rate and Number of Services

The processing rate for each likelihood queries are increased and the number of services recommended by user also fetched easily by means of Fun Ride Application thereby increasing the efficiency of data, increased throughput, latency and the response time and scalability can be greatly improved.

CONCLUSION

The Like Based approach is the best where only the recommended services and user interested services will be provided to the users. In their profile only their services can be viewed. This Fun Ride application also ensures data privacy and security and can be outsourced via a Multi Cloud Environment and utilizes the Big Data Approach for service creation when there is an unstructured data and grouping the structured data as well.

REFERENCES

1. Catherine, M.R. and E.B. Edwin, 2013. A Survey on Recent Trends in Cloud Computing and its Application for Multimedia, International Journal of Advanced Research in Computer Engineering & Technology, 2(1): 304-309.

2. Wu, X., X. Zhu, G.Q. Wu, *et al.*, 2014. Data mining with big data, *IEEE Trans. on Knowledge and Data Engineering*, 26(1): 97-107.
3. Liu, Z., P. Li, Y. Zheng, *et al.*, 2009. Clustering to find exemplar terms for keyphrase extraction, in *Proc. 2009 Conf. on Empirical Methods in Natural Language Processing*, pp: 257-266.
4. Chang, F., J. Dean, S. mawat, *et al.*, 2008. Bigtable: A distributed storage system for structured data, *ACM Trans. on Computer Systems*, 26(2): 1-39.
5. Club CF: A Clustering-based Collaborative Filtering Approach for Big Data Application, *IEEE Transactions on Emerging Topics in Computing* 2014.
6. Thilagavathi, G., D. Srivaishnavi, N. Aparna, *et al.*, 2013. A Survey on Efficient Hierarchical Algorithm used in Clustering, *International Journal of Engineering*, 2(9).
7. Gupta, V., and G.S. Lehal, 2013. A Survey of Common Stemming Techniques and Existing Stemmers for Indian Languages, *Journal of Emerging Technologies in Web Intelligence*, 5(2): 157-161.
8. Sandeep, R.S., C. Vinay and S.M. Hemant, 2013. Strength and Accuracy Analysis of Affix Removal Stemming Algorithms, *International Journal of Computer Science and Information Technologies*, 4(2): 265-269.
9. Bellogín, A., I. Cantador, F. Díez, *et al.*, 2013. An empirical comparison of social, collaborative filtering and hybrid recommenders, *ACM Trans. on Intelligent Systems and Technology*, 4(1): 1-37.
10. Beyer, M.A. and D. Laney, 2012. The importance of 'big data': A definition, *Gartner, Tech. Rep.*, 2012.