ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016

# Clustering and Filtering Approach for searching Big Data Application Query

S. V. Phulari, Prasad P. Shah, Atul D. Kalpande, Vikas A. Pawar

***Abstract*— Cluster-based recommendation is best thought of as a variant on user-based recommendation. Instead of recommending items to users, items are recommended to clusters of similar users. This entails a preprocessing phase, in which all users are partitioned into clusters. Recommendations are then produced for each cluster, such that the recommended items are most interesting to the largest number of users. The upside of this approach is that recommendation is fast at runtime because almost everything is pre computed**

***Index Terms*— Big data Application, Cluster, Collaborative Filtering, Mashup.**

## I. INTRODUCTION

Document clustering has been used in many different areas of text mining and information retrieval. Initially it was used for improving the precision and recall in information retrieval systems and finding nearest neighbors of a document. Later it has also been used for organizing the results returned by a search engine and generating hierarchical clusters of documents. Initially we applied the K-Means and Agglomerative Hierarchical clustering methods on the data and found that the results were not very satisfactory and the main reason for this was the noise in the graph, created for the data. This provided us the motivation for trying a pre-processing of the graph to remove the extra edges. We applied a heuristic for removing the inter cluster edges and then applied the standard graph clustering methods to get much better results.

### *A. Overview*

Now a days we need to quickly go through large amounts of textual information to find out documents related to our interests and this document space is growing on a daily basis at an overwhelming rate. Now days it is common to store several million web - pages and hundreds of thousands of text files. Analyzing such huge quantities of data can be made easier if we can have a subset of words (Keywords) which can provide us with the main features, concept, theme etc of the document. Appropriate keywords can serve as a highly concise summary of a document and help us in easily organize documents and retrieve them based on their content. Keywords are used in academic articles to give an idea to the reader about the content of the article. In a textbook they are useful for the readers to identify and retain the main points in their mind about a particular section. As keywords represent the main theme of a text, they can be used as a measure of similarity for text clustering.

### *B. Brief Description*

Clustering based collaborative filtering approach contains two modules. First Clustering, in this services are clustered depend on similarity in Description, Functionality & Characteristics respectively. Second Collaborative filtering, in this, first rating similarity is computed & then predicted rating is given to the clustered services.

#### *a) clustering*

In step clustering first stem words are recognized by using Porter Stemmer Algorithm. Then similarity in services based on Description & Functionality is Computed By using Jaccard similarity coefficient (JSC). Characteristic similarity between two services is computed by using weighted sum of Description Similarity and Functionality Similarity. At last services are clustered using Agglomerative Hierarchical Clustering Algorithm.

#### *b) Collaborative filtering*

In this module rating similarity between two services is computed by using Pearson correlation coefficient (PCC).Then neighboring services are selected by using Constraint Formula. In last step all recommended services are ranked in non-ascending order according to their predicted ratings.

### *CLUSTERING*

Clustering is a major task in data analysis and data mining applications. It is the method of assigning a objects so that objects in the identical group are more related to each other than to those in other groups. Cluster is an ordered

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016
list of data which have the familiar characteristics. Cluster analysis can be done by finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters. Clustering is an unsupervised learning process. No super-vision means that there is no human expert who has assigned documents to classes. In clustering, it is the distribution and makeup of the data that will determine cluster membership. A good clustering method will produce high superiority clusters with high intra-class similarity and low inter-class similarity. The superiority of a clustering result depends on the similarity measure used by the method and its implementation. The superiority of a clustering technique is also calculated by its ability to find out some or all of the hidden patterns. Similarity of a cluster can be expressed by the distance function. In data mining, there are some requirements for clustering the data. Clustering based collaborative filtering approach mainly contains two types of clustering algorithms.

### a) Partitioned clustering

Partitioning clustering algorithm splits the data points into k partition, where each partition represents a cluster. The partition is done based on certain objective function. The cluster should exhibit two properties, these are (a) each group must contain at least one object (b) each object must belong to exactly one group. Partitioning methods relocate instances by moving them from one cluster to another, starting from an initial partitioning. Such methods typically require that the number of clusters will be pre-set by the user. Partitional clustering contains algorithms like K means clustering, K medoids clustering. But these Partitional algorithms have some limitations.

### b) Hierarchical Clustering

Hierarchical clustering is a technique of clustering which divide the similar dataset by constructing a hierarchy of clusters. This method is based on the connectivity approach based clustering algorithms. It uses the distance matrix criteria for clustering the data. It constructs clusters step by step. A hierarchical method creates a hierarchical decomposition of the given set of data objects. Tree of clusters is called as dendrograms. Every cluster node contains child clusters, sibling clusters partition the points Covered by their common parent. Hierarchical clustering is further divided in to two types.

Agglomerative: Agglomerative hierarchical clustering is a bottom-up clustering method. It starts by letting each object form its own cluster and iteratively merges cluster into larger and larger clusters, until all the objects are in a single cluster or certain termination condition is satisfied. The single cluster becomes the hierarchies root. For the merging step, it finds the two clusters that are closest to each other, and combines the two to form one cluster. Clustering based collaborative filtering approach uses agglomerative algorithm for clustering services.

### Divisive

It works in a similar way to agglomerative clustering but in the opposite direction. As it uses top down approach, this method starts with a single cluster containing all objects, and then successively splits resulting clusters until only clusters of individual objects remain.

## COLLABORATIVE FILTERING

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. Advantage of the collaborative filtering approach is that it does not rely on machine analyzable content. It is capable of accurately recommending complex items without requiring an understanding of the item itself. Collaborative Filtering assumes that people who agree in past will agree in future too and people will like the similar kinds if items they like in the past. Collaborative filtering contains two types of techniques, User based collaborative filtering and Item based collaborative filtering.

### a) User Based Collaborative Filtering

User-based collaborative filtering predicts a user's interest in an item which is based on rating information from similar user profiles. User based CF assumes that a good way to find a certain user's interesting item is to find other users who have a similar interest. This type of technique first tries to find the user's neighbors based on user similarities and then combine the neighbor users' rating scores.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016

**b) Item Based Collaborative Filtering**

Item based collaborative filtering technique also applies same idea like user based CF but instead of similarity between users it uses similarity between items. The rating of an item by a user can be predicted by averaging the ratings of other similar items rated by user.

### *C. Purpose*

- We proposed a Agglomerative Hierarchal Clustering or Hierarchal Agglomerative Clustering
- Clustering are such techniques that can reduce the data size by a large factor by grouping similar services together.
- A cluster contains some similar services just like a club contains some like-minded users. This is another reason besides abbreviation that we call this approach ClubCF.
- This approach is enacted around two stages. In the first stage, the available services are divided into small-scale clusters, in logic, for further processing. At the second stage, a collaborative filtering algorithm is imposed on one of the clusters.

### *D. Scope*

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

### *E. Applying software engineering approach*

***Incremental Model***

The product is decomposed into a number of components, each of which are designed and built separately (Termed as builds). Each component is delivered to the client when it is complete. This allows partial utilization of product and avoids a long development time. It also creates a large initial capital outlay with the subsequent long wait avoided. This model of development also helps ease the traumatic effect of introducing completely new system all at once. There are, overall, few problems with this model.

The incremental build model is a method of software development where the model is designed, implemented and tested incrementally (a little more is added each time) until the product is finished. It involves both development and maintenance. The product is defined as finished when it satisfies all of its requirements. This model combines the elements of the waterfall model with the iterative philosophy of prototyping. The product is decomposed into a number of components, each of which are designed and built separately (termed as builds). Each component is delivered to the client when it is complete. This Customizable Indoor Location and Navigation System Based on Bluetooth allow partial utilization of product and avoid a long development time. It also creates a large initial capital outlay with the subsequent long wait avoided. This model of development also helps ease the traumatic effect of introducing completely new system all at once. There are, overall, few problems with this model.

***Benefits-***

1. Any faulty piece of software can be identified easily as very few changes are done after every iteration.
2. It is easier to test and debug as testing and debugging can be performed after each iteration.
3. This model does not affect anyone's business values because they provide core of the software which customer needs, which will indeed help that person to keep run his business.
4. After establishing an overall architecture, system is developed and delivered in increments.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016

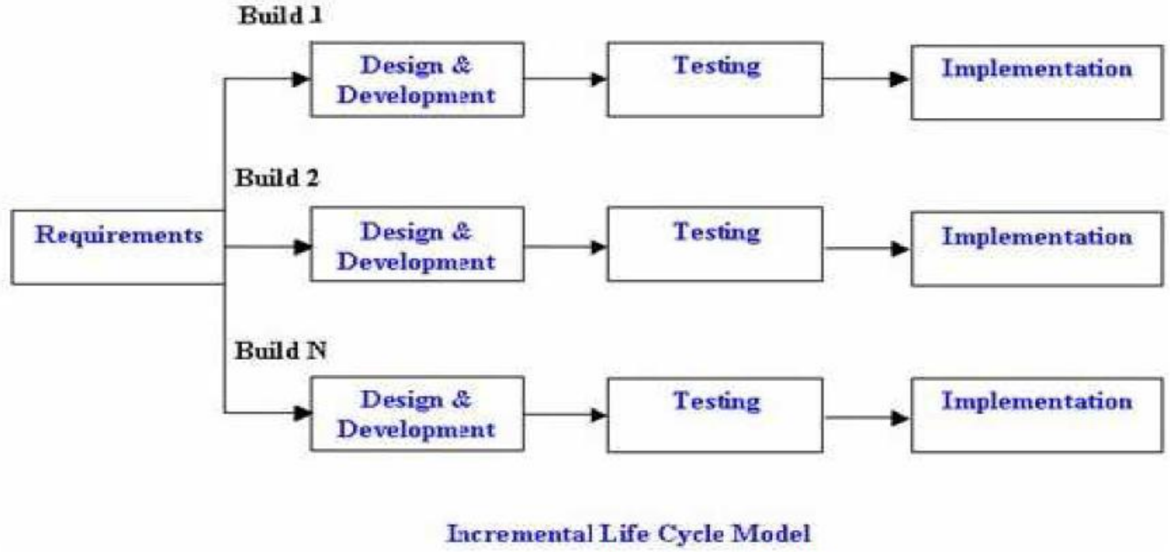



Figure 1: Incremental Model

***F. Steps***

***Deployment of Clustering Stage***

***Step 1.1: Stem Words***

Different developers may use different-form words to describe similar services. Using these words directly may influence the measurement of description similarity. Therefore, description words should be uniformed before further usage. In fact, morphological similar words are clubbed together under the assumption that they are also semantically similar. For example, „map", „maps", and „mapping" are forms of the equivalent lexeme, with „map" as the morphological root form. To transform variant word forms to their common root called stem, various kinds of stemming algorithms, such as Lovins stemmer, Dawson Stemmer, Paice/Husk Stemmer, and Porter Stemmer, have been proposed. Among them, Porter Stemmer (http://tartarus.org/martin/PorterStemmer/) is one of the most widely used stemming algorithms. It applies cascaded rewrite rules that can be run very quickly and do not require the use of a lexicon. In ClubCF approach, the words in $Dt$ are gotten from service Bigtable where row key = "$st$" and column family = "$Description$". The words in $Dj$ are gotten from service Bigtable where row key = "$sj$" and column family = "$Description$". Then these words are stemmed by Porter Stemmer and put into $Dt'$ and $Dj'$, respectively.

***Step 1.2: Compute Description Similarity and Functionality Similarity***

Description similarity and functionality similarity are both computed by Jaccard similarity coefficient (JSC) which is a statistical measure of similarity between samples sets [15]. For two sets, JSC is defined as the cardinality of their intersection divided by the cardinality of their union. Concretely, description similarity between $st$ and $sj$ is computed by formula (1):

$$D\_sim(s_t, s_j) = \frac{|D_t' \cap D_j'|}{|D_t' \cup D_j'|} \quad (1)$$

It can be inferred from this formula that the larger $|Dt' \cap Dj'|$ is, the more similar the two services are. Dividing by $|Dt' \cup D||'$ is the scaling factor which ensures that description similarity is between 0 and 1. The functionalities in $Ft$ are gotten from service Bigtable where row key = "$st$" and column family = "$Functionality$". The functionalities in $Fj$ are gotten from service Bigtable where row key = "$sj$" and column family = "$Functionality$". Then, functionality similarity between $st$ and $sj$ is computed using JSC as follow:

***Step 1.3: Compute Characteristic Similarity***

Characteristic similarity between $st$ and $sj$ is computed by weighted sum of description similarity and functionality similarity, which is computed as follow:

$$F\_sim(s_t, s_j) = \frac{|F_t \cap F_j|}{|F_t \cup F_j|} \quad (2)$$

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016

In this formula, $\alpha \in 0, 1$ is the weight of description similarity, $\beta \in 0, 1$ is the weight of functionality similarity and $\alpha+\beta=1$. The weights express relative importance between these two.

$$C\_sim(s_t, s_j) = \alpha \times D\_sim(s_t, s_j) + \beta \times F\_sim(s_t, s_j) \quad (3)$$

Provided the number of services in the recommender systemic $n$, characteristic similarities of every pair of services are calculated and form a $n \times n$ characteristic similarity matrix $D$. An entry $dt,j$ in $D$ represents the characteristic similarity between $st$ and $sj$.

***Step 1.4: Cluster Services***

Clustering is a critical step in our approach. Clustering methods partition a set of objects into clusters such that objects in the same cluster are more similar to each other than objects in different clusters according to some defined criteria.

Generally, cluster analysis algorithms have been utilized where the huge data are stored. Clustering algorithms can be either hierarchical or partitioned. Some standard partitioned approaches (e.g., K-means) suffer from several limitations: 1) results depend strongly on the choice of number of clusters K, and the correct value of K is initially unknown; 2) cluster size is not monitored during execution of the K-means algorithm, some clusters may become empty ("collapse"), and this will cause premature termination of the algorithm; 3) algorithms converge to a local minimum. Hierarchical clustering methods can be further classified into agglomerative or divisive, depending on whether the clustering hierarchy is formed in a bottom-up or top-down fashion. Many current state-of-the-art clustering systems exploit agglomerative hierarchical clustering (AHC) as their clustering strategy, due to its simple processing structure and acceptable level of performance. Furthermore, it does not require the number of clusters as input. Therefore, we use an AHC algorithm for service clustering as follow. Assume there are $n$ services. Each service is initialized to be a cluster of its own. At each reduction step, the two most similar clusters are merged until only $K$ $K<n$ clusters remains.

***B. Deployment of Collaborative Filtering Stage***

Step 2.1: Compute Rating Similarity Rating similarity computation between items is a time-consuming but critical step in item-based CF algorithms. Common rating similarity measures include the Pearson correlation coefficient (PCC) and the cosine similarity between ratings vectors. The basic intuition behind PCC measure is to give a high similarity score for two items that tend to be rated the same by many users. PCC which is the preferred choice in most major systems was found to perform better than cosine vector similarity. Therefore, PCC is applied to compute rating similarity between each pair of services in ClubCF. Provided that service $st$ and $sj$ are both belong to the same cluster, PCC-based rating similarity between $st$ and $sj$ is computed by formula (4):

$$R\_sim(s_t, s_j) = \frac{\sum_{u_i \in U_t \cap U_j} \left(r_{u_i,s_t} - \bar{r}_{s_t}\right)\left(r_{u_i,s_j} - \bar{r}_{s_j}\right)}{\sqrt{\sum_{u_i \in U_t \cap U_j} \left(r_{u_i,s_t} - \bar{r}_{s_t}\right)^2} \sqrt{\sum_{u_i \in U_t \cap U_j} \left(r_{u_i,s_j} - \bar{r}_{s_j}\right)^2}} \quad (4)$$

Here, $Ut$ is a set of users who rated $st$ while $Uj$ is a set of users who rated $sj$, $ui$ is a user who both rated $st$ and $sj$, $rui,st$ is the rating of $st$ given by $ui$ which is gotten from service big table where row key = "$st$" and column key = "$Rating$:$ui$", $rui,sj$ is the rating of $sj$ given by $ui$ which is gotten from service big table where row key = "$sj$" and column key = "$Rating$:$ui$", $rst$ is the average rating of $st$, and $rsj$ is the average rating of $sj$. It should be noted that if the denominator of formula (4) is zero, we make 0, in order to avoid division by 0. Although PCC can provide accurate similarity computation, it may overestimate the rating similarities when

$$R\_sim'(s_t, s_j) = \frac{2 \times |U_t \cap U_j|}{|U_t| + |U_j|} \times R\_sim(s_t, s_j) \quad (5)$$

There are a small amount of co-rated services. To address this problem, the enhanced rating similarity [23] between $st$ and $sj$ is computed by formula (5):

In this formula, $|Ut \cap Uj|$ is the number of users who rated both service $st$ and $sj$, $Ut$ and $Uj$ are the number of users who rated service $st$ and $sj$, respectively. When the number of co-rated services is small, for example, the weight

$\frac{2 \times |U_t \cap U_j|}{|U_t| + |U_j|}$

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016

Will decrease the rating similarity estimation

Between these two users. Since the value of $\frac{2\times|U_t\cap U_j|}{|U_t|+|U_j|}$ is between the interval of [0,1] and the value of $R\_sim(st,sj)$ is in the interval of [-1,1], the value of $R\_sim'(st,sj)$ is also in the interval of [-1,1].

***Step 2.2:*** Select Neighbors Based on the enhanced rating similarities between services, the neighbors of a target service $st$ are determined according to constraint formula (6):

$$N(s_t) = \{ s_j \,|R\_sim'(s_t, s_j) > \gamma , s_t \neq s_j \} \quad (6)$$

Here, $R\_sim'\,(st,\ sj)$ is the enhanced rating similarity between service $st$ and $sj$ computed by formula (5), $\gamma$ is a rating similarity threshold. The bigger value of $\gamma$ is, the chosen number of neighbors will relatively less but they may be more similar to the target service, thus the coverage of collaborative filtering will decrease but the accuracy may increase. On the contrary, the smaller value of $\gamma$ is, the more neighbors are chosen but some of them may be only

Slightly similar to the target service, thus the coverage of collaborative filtering will increase but the accuracy would decrease. Therefore, a suitable $\gamma$ should be set for the tradeoff between accuracy and coverage. While $\gamma$ is assigned, $sj$ will be selected as a neighbor of $st$ and put into the neighbor set $N(\mathrm{S}t)$ if $R\_sim'(st,sj) > \gamma$.

***Step 2.3:*** Compute Predicted Rating for an active user $ua$ for whom predictions are being made, whether a target service $st$ is worth recommending depends on its predicted rating. If $N(st)\neq\Phi$, similar to the computation formula proposed by Wu et al. [24], the predicted rating $P(ua,st)$ in an item-based CF is computed as follow:

$$P_{u_a,s_t} = \bar{r}_{s_t} + \frac{\sum_{s_j\in N(s_t)} \left(r_{u_a,s_j} - \bar{r}_{s_j}\right) \times R\_sim'(s_t, s_j)}{\sum_{s_j\in N(s_t)} R\_sim'(s_t, s_j)} \quad (7)$$

Here, $rst$is the average rating of $st$, $N(st)$ is the neighbor set of $st$, $sj\in N(st)$ denotes $sj$ is a neighbor of the target service $st$, $rua,sj$ is the rating that an active user $ua$ gave to $sj$, $rsj$ is the average rating of $sj$, and $R\_sim'(st,sj)$ is the enhanced rating similarity between service $st$ and $sj$ computed using formula (5). If the predicted rating of a service exceeds a recommending threshold, it will be a recommendable service for the active user. A service is generally rated on a five-point scale from 1 (very dissatisfied) to 5 (very satisfied). Therefore, we set the recommending threshold to 2.5 which is the median value of the max rating. All recommendable services are ranked in non-ascending order according to their predicted ratings so that users may discover valuable services quickly.

## II. LITERATURE SURVEY

### *1. ClubCF: A Clustering-based Collaborative Filtering Approach for Big Data Application*

In this paper, we present a ClubCF approach for big data applications relevant to service recommendation. Before applying CF technique, services are merged into some clusters via an AHC algorithm. Then the rating similarities between services within the same cluster are computed. As the number of services in a cluster is much less than that of in the whole system, ClubCF costs less online computation time. Moreover, as the ratings of services in the same cluster are more relevant with each other than with the ones in other clusters, prediction based on the ratings of the services in the same cluster will be more accurate than based on the ratings of all similar or dissimilar services in all clusters. These two advantageous of ClubCF have been verified by experiments on real-world data set. Future research can be done in two areas. First, in the respect of service similarity, semantic analysis may be performed on the description text of service. In this way, more semantic-similar services may be clustered together, which will increase the coverage of recommendations. Second, with respect to users, mining their implicit interests from usage records or reviews may be a complement to the explicit interests (ratings). By this means, recommendations can be generated even if there are only few ratings. This will solve the sparsely problem to some extent.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016

### *2. Collaborative Filtering Approach For Big Data Applications in Social Networks*

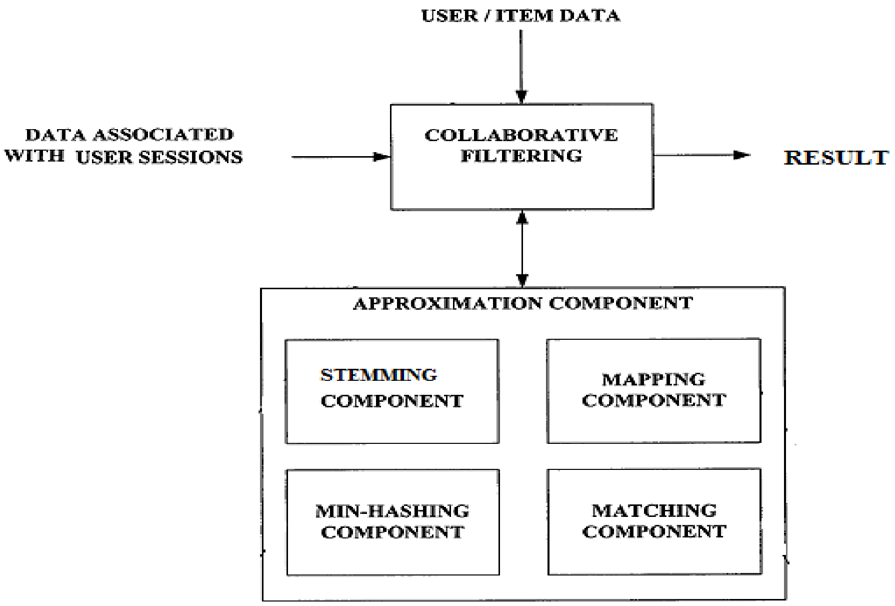
In attendance a Club CF approach for big data applications relevant to service suggestion. Proceeding to applying CF technique, services are merged into some clusters via an AHC algorithm. Then the rating similari-ties between services within the same cluster are computed. As the number of services in a cluster is much less than that of in the whole system, ClubCF costs less online computation time. Moreover, as the ratings of services in the same cluster are more relevant with each other than with the ones in other clusters, prediction based on the ratings of the services in the same cluster will be more accurate than based on the ratings of all similar or dis-similar services in all clusters.

### *3. Strength and Accuracy Analysis of Affix Removal Stemming Algorithms*

Thus, it has been conclude that all the stemming algorithms discussed in this paper are comparatively strong and aggressive, but are less accurate. All tends to produce both over-stemming and under stemming errors. However, the occurrence of under-stemming errors in Paice/Husk stemmer is comparatively low. The ACWF obtained by Lovins and Porter1 stemmer shows negative percentage. This is because the number of words that stems to incorrect words is more than the correctly stemmed words. Thus in both the cases over-stemming and under-stemming errors occurred more than the others. Further the AWCF of Paice/husk stemmer is comparatively Positive still it has the problem of occurrence of over-stemming errors, as the ICF and WSF is comparatively high. The CSF and AWCF is obtained by Porter2 stemmer is quite good, but it produces the over-stemming errors as compare to under-stemming errors.

## III. SYSTEM DESIGN

*System Architecture*




## IV. CONCLUSION

We present a ClubCF approach for big data applications relevant to service recommendation. Before applying CF technique, services are merged into some clusters via an AHC algorithm. Then the rating similarities between services within the same cluster are computed. As the number of services in a cluster is much less than that of in the whole system, ClubCF costs less online computation time. Moreover, as the ratings of services in the same cluster are more relevant with each other than with the ones in other clusters, prediction based on the ratings of the services in the same cluster will be more accurate than based on the ratings of all similar or dissimilar services in all clusters. These two advantageous of ClubCF have been verified by experiments on real-world data set. Future research can be done in two areas. First, in the respect of service similarity, semantic analysis may be performed on the description text of service. In this way, more semantic-similar services may be clustered together, which will

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 1, January 2016
increase the coverage of recommendations. Second, with respect to users, mining their implicit interests from usage records or reviews may be a complement to the explicit interests (ratings). By this means, recommendations can be generated even if there are only few ratings. This will solve the sparsity problem to some extent.

## V. FUTURE SCOPE

- We proposed a Agglomerative Hierarchal Clustering or Hierarchal Agglomerative Clustering
- Clustering are such techniques that can reduce the data size by a large factor by grouping similar services together.
- A cluster contains some similar services just like a club contains some like-minded users. This is another reason besides abbreviation that we call this approach ClubCF.
- This approach is enacted around two stages. In the first stage, the available services are divided into small-scale clusters, in logic, for further processing. At the second stage, a collaborative filtering algorithm is imposed on one of the clusters.
- This similarity metric computes the Euclidean distance d between two such user points this value alone doesn’t constitute a valid similarity metric, because larger values would mean more-distant, and therefore less similar, users. The value should be smaller when users are more similar.

## REFERENCES

[1] X. Wu, X. Zhu, G. Q. Wu, et al., “Data mining with big data,” IEEE Trans. on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107, January 2014.

[2] Z. Zheng, J. Zhu, M. R. Lyu. “Service-generated Big Data and Big Data-as-a-Service: An Overview,” in Proc. IEEE BigData, pp. 403-410, October 2013.

[3] A. Bellogín, I. Cantador, F. Díez, et al., “An empirical comparison of social, collaborative filtering, and hybrid recommenders,” ACM Trans. on Intelligent Systems and Technology, vol. 4, no. 1, pp. 1-37, January 2013.

[4] R. S. Sandeep, C. Vinay, S. M. Hemant, “Strength and Accuracy Analysis of Affix Removal Stemming Algorithms,” International Journal of Computer Science and Information Technologies, vol. 4, no. 2, pp. 265-269, April 2013.

[5] V. Gupta, G. S. Lehal, “A Survey of Common Stemming Techniques and Existing Stemmers for Indian Languages,” Journal of Emerging Technologies in Web Intelligence, vol. 5, no. 2, pp. 157-161, May 2013.

## AUTHOR BIOGRAPHY

**Prof. S. V. Phulari**

Lecturer at College of engg Manjari BK. Pune, Maharashtra, India.

**Prasad P. Shah**

He is pursuing BE (COPMUTER) degree from College of Engg Manjari BK. Pune, Maharashtra, India.

**Atul D. Kalpande**

He is pursuing BE (COPMUTER) degree from College of Engg Manjari BK. Pune, Maharashtra, India.

**Vikas A. Pawar**

He is pursuing BE (COPMUTER) degree from College of Engg Manjari BK. Pune, Maharashtra, India.