Imperial Journal of Interdisciplinary Research (IJIR)
Vol.2, Issue-1 , 2016
ISSN : 2454-1362 , www.onlinejournal.in

Social web clustering on mails using an improved k-means algorithm

Manikyam Nageswara Raju

M.Tech Student,
Department of Computer science and Engineering,
Madanapalle Institute of Technology and Science
Madanapalle, Ap,India.

Abstract-- *Now a day's internet is the very important to gathering the information are improved over knowledge, lot of e-documents such, as html pages, digital libraries etc. Email is one of the most popular forms of communication between each ether nowadays, mainly due to its very low cost and compatibility of diverse types of information. Clustering is a division of data into groups of similar objects, each group called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups, Application of document clustering techniques to cluster e-mails is an interesting application Techniques, like k-means. Clustering is an important technique that organizes large number of objects into smaller coherent groups, Email is one of the most frequently used e-document by individual or organization, Email categorization is one of the major tasks of email mining. The E mail are sending into one (or) more receivers Categorizing emails into different groups help easy retrieval and maintenance, Like other e-documents, emails can also be classified using clustering algorithms. In this paper a similarity measure called Similarity Measure for Text Processing is suggested for email clustering in to dynamically because of number of text message is allocating on the email based on the text emails is classified.*

KEYWORDS

"Similarity Measure, Clustering Algorithm, Dynamic data Clustering,"

1. INTRODUCTION

Clustering is a division of data into groups of similar objects each group called cluster, A Web Application is a type of software that is hosted on server and can be accessed remotely by a human through an Internet Browser. Other types of software are also widely used and distributed. But web-based application faces some additional challenges like maintaining a consumer base and ensuring acceptability of the application. Most of the work in text processing can easily be broadly categorized into two areas, clustering and classification,[1]

We are living in cyber world dumped with whole lot of information Efficient and accurate retrieval of the information is important for survival of many web portals, Text processing is an important aspect of information retrieval data mining and web search.[2] It is equally important to cluster the emails into different groups, so as to retrieve the similar emails i.e. email containing search string or key words, We can further group the mails based on time stamp for more easy retrieval we point out that the difficulty of extending in dynamic data clustering is that, the pre existing objects have established certain relationships

Compared with the previous works another remarkable feature work, the k mean clustering algorithms are proposed based on a message-passing framework before using an the K mean's algorithm it is not getting required information previously Message can be displayed static way but now dynamically displayed, each object is a node in a graph and weighted edges between nodes correspond to pair wise similarity between objects.[3] When a new object is observed it will be added on the message passing is implemented to find a new exemplar set, Because that only one or a few of nodes entering will not

change the structure of the whole message, a local adjustment of availabilities and responsibilities is enough Therefore messages passing on graphs will reconverted quickly. Based on these features, the K mean clustering algorithms proposed in this paper don't need to re-implemented. The Emails was classified into dynamic order it was easy to identify the user required data.

2. RELATED WORK

In this paper related to the social web sites i.e. email the paper was passing a mail from sender to receive the receiver was finding of the message not easy because of the data (or) message is static way, in that way getting the message classified such way to checking the all the mails in such case checking of the time is more and risk accruing, but now the message classified in to dynamically by using an the K mean algorithm in such case to identified required information, I have taken to the Database in that database related words was inserted (or) created on the database in to the server side , baside on the Database the message was find out the server side easy, in this two modules are taken one is email (or) social web site i.e. Email and server, in this first we can active the server side and the go through the mail login and then compose the mail to enter the text (or) message in that entering of the message and then send to the receiver after sending of the message the server side easy to identified, which mode the message was getting and size of the message also, when the mail was passing to the sender to receive the receiver was check the received mail box but the before checking of the mail box, the server was displayed on where the message was clustering easy to finding.

Clustering is one of the data mining methods which is used for the purpose of email mining, Clustering is used for grouping emails for the purpose easy management .Commonly used clustering algorithms for email grouping are k-means clustering algorithm, As we nowadays are so much used to our emails, they have become an important part of our corporate and social life, it is very important to the social web[4] sites to easy to find out the required information.

In this we have create data from Database which is related data can be inserted in to the DataBase.

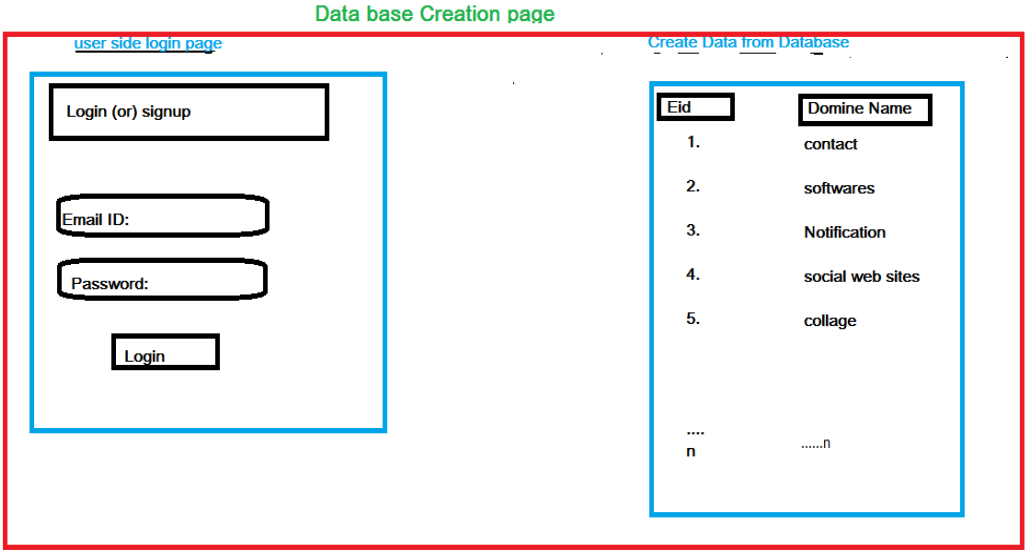
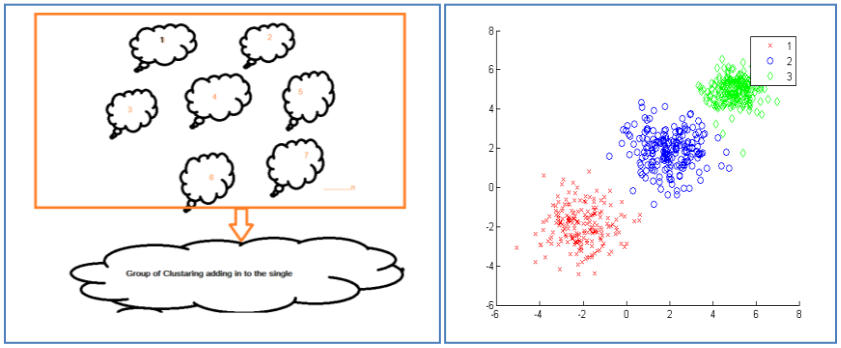


Fig : Demine name creation from DB



Number of Cluster grouped in to single mode and Data Clustering

3.K-Means Algorithm

BEGIN
initialize c; //first initialize cluster
c' = n; //her num of cluster
$M_i = \{x_i\}$; i = 1,...,n //min num of cluster at start 1 –n
do
c' = c' - 1 / /Decrement the cluster
Find nearest clusters M //find the nearest cluster
until 0 = c‘ //until cluster is 0
Merge M_i ,C[maxCw];

//merge the min num cluster until cluster wait is 0 //
return c clusters
END

By using K mean algorithm data is clustering in same manner finding the data from database Suppose a set of observations are given as (x1, x2, …, x*n*), where each observation is a *d*-dimensional real vector, *k*-means clustering aims to partition the *n* observations into *k* sets

($c \leq n$)S = {*S*1, *S*2, …, *Sc*} so as to minimize the within-cluster sum of squares

In the K-means algorithm was we have to chose the any data in mining, in that data is cluster that name is “C” in taking of the data stating to ending to apply the clustering by using an the K means, in the algorithm[5] “Mi” is the minimum number of the cluster the cluster can start with the 0 or n-1, here C’ is the finding of the time mails was decrement because of the clustering can be start with the ‘n-1’ to 0 until the clustering count is 0 it is the final result of the clustering.

4. PROPOSED SYSTEM

Now a day’s the social web sites are most important to communicate each one, there are number of social web sites are available, for Exp Gmail, Face book, whatApp, Skype, twitter etc, now i am taken one of the web site is E-mail it is most important to communicate others, we are taken to the Gmail it is very important for education peoples because of each educated peoples are to use mails that is Gmail, yahoo mail ect, for Exp we are taken to the Gmail user Required information can getting from sender to receiver, that is communicate between the sender and receiver now we send mail from sender to receiver that message was reached from the sender after the message was clustering from the receiver side, this process is data clustering technique the data was clustering static way clustering in this disadvantage is more time was taken [6]. In the static clustering very risk user required mail not easy to identified in such way we are taken to other technique dynamic data clustering this technique will be to overcome the that problem it will easy to find out the required message easy the message will be sending to one person are multiple persons, the dynamic data technique is user required mails easy

to finding because of the data will be clustering in the mail side the mails was which area clustered it can be displayed on the server side clustering map, We use Pearson-Correlation-Distance for Initialization of centurions, which helps in pruning and improving the performance of the proposed algorithm.

The main purpose of the e-mail clustering technique to identifying the clustering message in the received mailbox and server side, user cannot check the all the mails because it is very long time and risk also, in the database mode to check the server side it can be easy to identified, in the same manner the receiver mail box which area the data was getting baside on the related mode the server side and mailbox was finding the message easily in this mode we are taken to the necessary data only.

In the Pearson-correlation-distance method uses 0 to represent to items are equal, 1 represents the first item is higher and -1 represents the first item is smaller. Finally the dissimilarity value is rounded to 0-2.

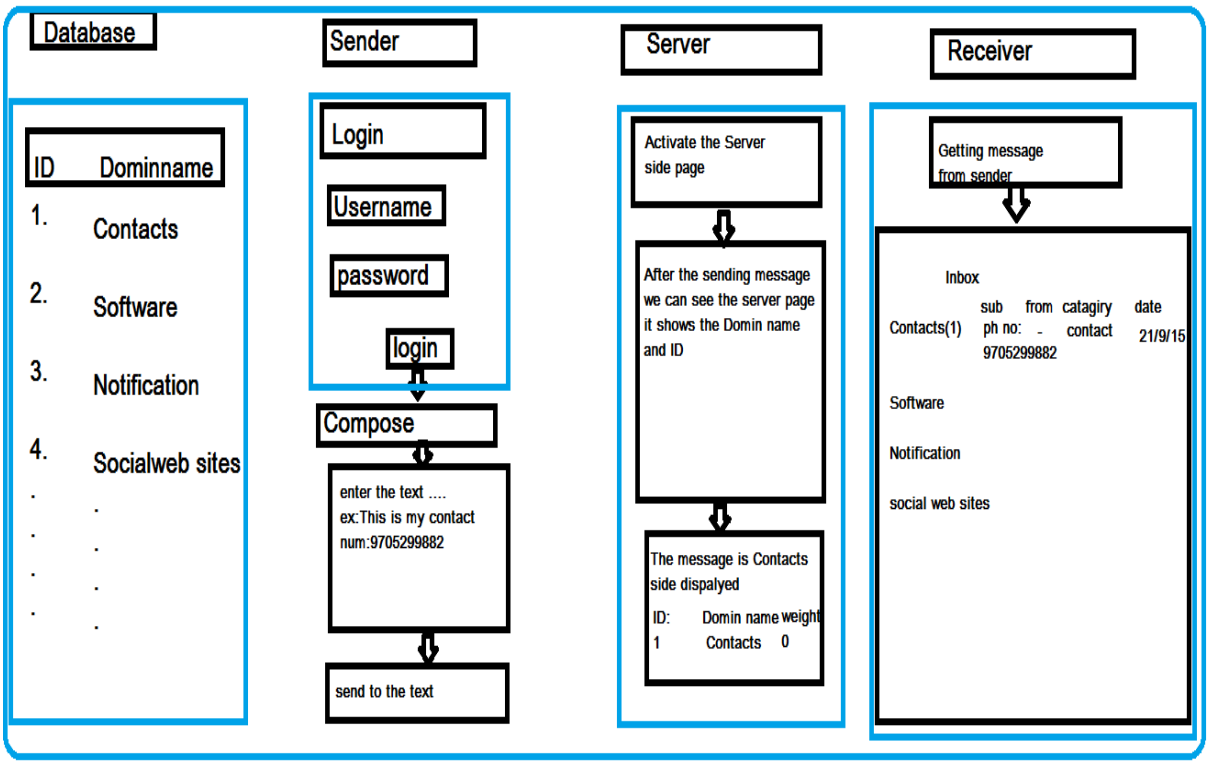


Fig: K mean Technique for data clustering in Gmail

5.DATA COLLECTION AND PARSING:

In this paper data can be written on the database system, the database well be created on the related words which is stored on the database, what sender sending message in that message related to it was stored in the database server side, After the data can be clustering to easy to identified in the server side as well as mail box, in order to compose mail this mail have, the sender was send to data received got it, after sending of the data reviewer can go to inbox it is not easy to got the information because of the receiver was searching all the mails but now using an K-means algorithm[5] the user was easy to get the required information mail by using on Dynamic data clustering technique.

i) S: (xi,xj) ∈ S if xi,xj are similar
ii) D: (xi,xj) ∈ D if xi,xj are dissimilar

Advantage

<i>Application of database is related words were created on the DB, the time maintenance it is very less time to find out the data.

<ii> Avoiding unnecessary data from the database

<iii> It cab be working on the off-mode and on-mode, because of the clustering is technique was working

<iV> Good result can get the less time and also higher speed.

Conclusion :

In this paper we have addressed the problem of eliminating unnecessary calculations associated with the K-Means clustering algorithm and applied on high dimensional e-mail messages to classify the e-mails by semantics of their content. The computational time and the computational resources are minimized, but the performance is improved.Additionally, some other problems such as how to determine the unnecessary data and without database to connect with internet (or) without to identify the data, how to implement the mails to easy to getting required information.

REFERANCE:

[1] W. W. Cohen: Learning Rules that classify E-mail. In Proc. of the 1996 AAAI Spring Symposium in Information Access, 1996.

[2] LUIA FILIPE DA CRUZ NASSIF AND EDUARDO RAUL HRUSCHKA **—Document Clustering for Forensic Analysis: An Approach for Improving Computer Inspection.‖** ieee transactions on information forensics and security,1, january 2013**,** vol. 8, no **.**

[3] B.J. Frey and D. Dueck, "Clustering by Passing Messages between Data Points," Science, vol. 315, no. 5814, pp. 972-976, Feb. 2007.

[4] Oren Zamir and Oren Etzioni. Grouper: **—a dynamic clustering interface to Websearch results‖ .** Computer Networks (Amsterdam, Netherlands: 1999), 31(11–16):1361–1374, 1999.

[5] A.K. Jain, "Data Clustering: 50 Years Beyond K-Means," PatternRecognition Letters, vol. 31, no. 8, pp. 651-666, June 2009.

[6] Dawid Weiss and Jerzy Stefanowski.**—Web search result dynamic clustering in Polish:Experimental evaluation of Carrot —.** In Proceedings of the New Trends in IntelligentInformation Processing and Web Mining Conference, Zakopane, Poland,2003.

Authors

Ms. NAGESWARA RAJU is a post graduate student in Computer science & Engineering from madanapalle institute of technology and science JNTUA ap State, India