CLUSTERING OF MULTIMEDIA FILES CONCEPT

Sudha V[1], Mrs. Malathi A[2]

[1,] Student, Department of Computer Science & Engineering, Anand Institute of Higher Technology, Kazhipattur, Chennai, Tamil Nadu, India
[2] Assistant Professor, Department of Computer Science & Engineering, Anand Institute of Higher Technology, Kazhipattur, Chennai, Tamil Nadu, India

ABSTRACT

A file is capable of holding one or two elements (texts, videos, audios, videos, animations). It also consist of multimedia content.The file is not capable of storing more multimedia elements .The data has been collected from the data set then the data is about texts, images, videos and audios.There are so many similar data are mixed in the file system. The concept of Clustering is been used to find the same set of data in the file system and place it into the separate group. The process of organizing subjects into groups whose members are similar in some way. A clustering is therefore a collection of objects which are similar between them and or dissimilar to the objects belonging to other clusters. Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple word, the aim is to segregate groups with similar traits and assign them into clusters.

Keyword: - *clustering technique*

1. INTRODUCTION:

In computing, a file system controls how data is stored and retrieved. Without a file system, data placed in a storage medium would be one large body of data with no way to tell where one piece of data stops and the next begins. By separating the data into pieces and giving each piece a name, the data is easily isolated and identified. There are many different kinds of file systems. Each one has different structure and logic, properties of speed , flexibility, security, size and more. Multimedia applications and system are getting more and more involved in our everyday lives. Their main purpose is to deal with various media types like pictures, video data, audio data and text. Video and audio belong to continuous media data. When most people refer to multimedia, they generally mean the combination of two or more continuous media .In practice, the two media are normally audio and video, that is, sound plus moving pictures.

1.1 OBJECTIVE:

The purpose of the project is to separate multimedia file in the same set of file. (The file may have texts, audios, videos, and images). The file system mixed with the different set of the files, then the clustering concept is to separate the similar set of file into the group. The challenge of multimedia systems are media type that need to be played continuously. That means that the data that should be played has to arrive in time (or at least until a certain strict deadline).Continuous media data differs from discrete data but not only in its real time characteristics. A challenge for these systems is also the synchronization of pictures and the according sound. Hence these can be two different data streams.it is important to synchronize these before showing them on the monitor. Another difference to discrete data is the file size. Video and audio need much more storage space than next data and the multimedia file system has to organize the data on disk in a way that efficiently use the limited storage.

1.2 SCOPE OF THE PROJECT:

The main motivation of the project is to reduce unwanted space in the storage system. Easy to find the same set of data file in the system. The unwanted dataset has been cleanup from the file system.Their main purpose is to deal with the various media types like pictures,video data,audio data and text.Video and audio belong to continuous media data.

2. RELATED WORK:

I. Paper [1] A distributed algorithm is presented that partitions the nodes of the fully mobile network(multi-hop network) into clusters, thus giving the network a hierarchical organization. The algorithm is proven to be adaptive to changes in the network topology due to nodes mobility and to nodes addition/removal. A new weight – based mechanism is introduced for the efficient cluster formation/maintenance the allows the cluster organization to be configured for specific applications and adaptive to changes in the network status, not available in the previous solutions. Specifically, new and flexible criteria are defined the allow the choice of the node that coordinate the clustering process based on mobility parameters and/or their current status. Simulation results are provided that demonstrate up to an 85% reduction on the communication overhead associated with the cluster maintenance with respect to techniques used in clustering algorithms previously proposed.

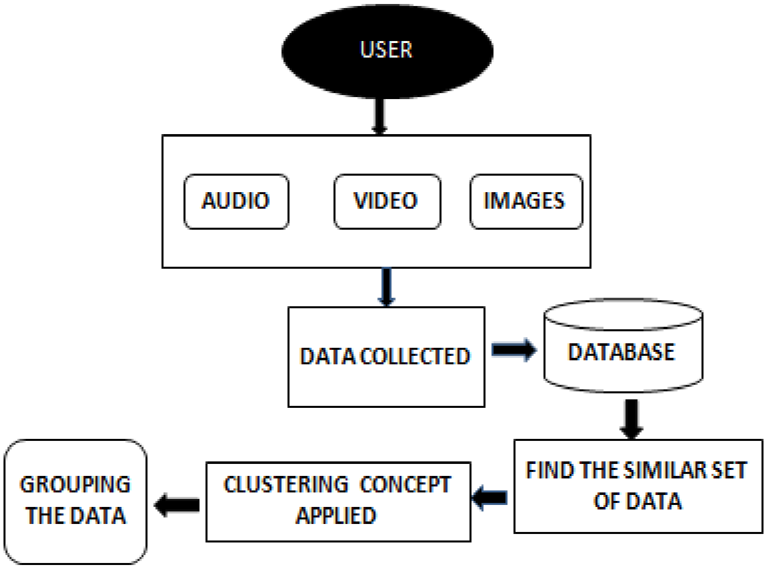
II. Paper [2] Several clustering algorithm can be applied to clustering in large multimedia databases. The effectiveness and efficiency of the existing algorithms, however, is somewhat limited, since clustering in multimedia databases requires clustering high-dimensional feature vectors and since multimedia database often contain large amount of noise. In this paper, we therefore introduce a new algorithm to clustering in large multimedia databases called DENCLUE. The basic idea of our new approach is to model the overall point density analytically as the sum of influence functions of the data points. Clusters can then be identified by determining density-attractors and clusters of arbitrary shape can be easily described by the simple equation based on the overall density function. The advantage of our new approach are (1) it has a firm mathematical basis, (2) is has good clustering properties in data set with large amounts of noise, (3) it allows the compact mathematical description of arbitral shaped cluster in high dimensional data set and (4) it is significantly faster than existing algorithms. To demonstrate the effectiveness and efficiency of DENCLUE, we perform a series of experiments on a number of different data set from CAD and molecular biology. A comparison with DBSCAN shows the superiority of our new approach.

III. Paper [3] Most of the online multimedia collections, such as picture galleries or video archives, are categorized in a fully manual process, which is very expensive and may soon be infeasible with the rapid growth of multimedia repositories. In this paper, we present an effective method for automating this process within the unsupervised learning framework. We exploit the truly multi-modal nature of multimedia collections - they have multiple views, or modalities, each of which contributes its own perspective to the collection's organization. For example, in picture galleries, image captions are often provided that form a separate view on the collection. Color histograms (or any other set of global features) form another view. Additional views are blobs, interest points and other sets of local features. Our model, called Comraf* (pronounced Comraf-Star), efficiently incorporates various views in multi-modal clustering, by which it allows great modeling flexibility. Comraf* is a light-weight version of the recently introduced combinatorial Markov random field (Comraf). We show how to translate an arbitrary Comraf into a series of Comraf* models, and give an empirical evidence for comparable effectiveness of the two. Comraf* demonstrates excellent results on two real-world image galleries: it obtains 2.5-3 times higher accuracy compared with a uni-modal k-means.

IV. Paper [4] after the generation of multimedia data turned digital, an explosion of interest in their data storage, retrieval, and processing has drastically increased. This includes videos, images, and audios, where we now have higher expectations in exploiting these data at hands. Typical manipulations are in some forms of video/image/audio processing, including automatic speech recognition, which require fairly large amount of storage and are computationally intensive. In our recent work, we have demonstrated the utility of time series representation in the task of clustering multimedia data using k-medoids method, which allows considerable amount of reduction in computational effort and storage space. However, k- means is a much more generic clustering method when Euclidean distance is used. In this work, we will demonstrate that unfortunately, k-means clustering will sometimes fail to give correct results, an unaware fact that may be overlooked by many researchers. This is especially the case when Dynamic Time Warping (DTW) is used as the distance measure in averaging the shape of time series. We also will demonstrate that the current averaging algorithm may not produce the real average of the time series, thus generates incorrect k-means clustering results, and then show potential causes why DTW averaging methods may not achieve meaningful clustering results. Lastly, we conclude with a suggestion of a method to potentially find the shape-based time series average that satisfies the required properties.

V. Paper [5] Multimedia information retrieval is the challenging problem because multimedia information is not inherently structured. Jabber is an experimental system that attempts to bring some structure to the task. Jabber allows users to retrieve records of video conferences based upon the concepts discussed. In this paper we introduce concept find, a sub system within jabber and show how it is able to process the spoken text of a meeting into meeting topics. Concept finder can make subtle distinctions among different sense of the same words, and is able to summarize a set of related words, giving a name to each topic. Users can then use this name to query or browse the stored multimedia, through jabber’s user interface. By presenting information that closely matches a user’s expectations, the challenge of multimedia retrieval is rendered more tractable.

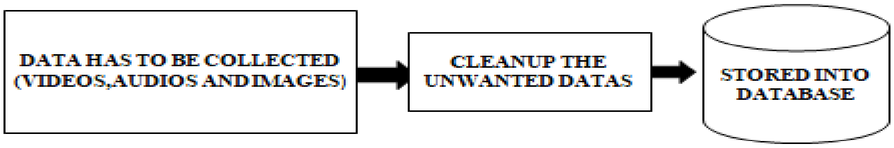
3. ARCHITECTURE DIAGRAM:



4. IMPLEMENTATION:

4.1 DATA SET COLLECTION:

Data set is the collection of data. In case of the tabular data set corresponds to one or more database tables. Every column of the table that represents the particular variable, and each row corresponds to the given record of the data set in the question. In the multimedia file system, the data may texts, videos, audios and images. All the data is been stored in the file system.


DATA HAS TO BE COLLECTED (VIDEOS,AUDIOS AND IMAGES)
CLEANUP THE UNWANTED DATAS
STORED INTO DATABASE

4.2 INPUT TO THE FILESYSTEM:

The data is been collected and input to the file system. Then the data is been stored in the different set of files. The files contain both similar and dissimilar data items. There are so many types of data are been stored in the file system.

4.3 IDENTIFY THE SIMILAR SET OF DATA:

In the multimedia file system, different types of the data are been stored. Using the tag name is used to find the similar set of data set in the file system. The same name of the tag name is been categorized into the separate group.

4.4 CLUSTERING ONCEPT APPLIED:

Clustering is the machine language technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each point into a specific group. In theory, data points that are in the same group should have similar property and are features. Clustering is the method of unsupervised learning and is a common technique for statistical data analysis used in many fields. In the multimedia file system, the same set of the data is been categorized into the different groups. Similar data is been identified and it is forming a group.

4.5 GROUPIN THE SAME SET OF FILE:

All the similar data is been identified and formed into the group. At the same time, the similar set of data in the data set is been separated to form a group using the tag name.

5. CONCLUSION:

In this project, we can approach a method to reduce the unwanted space in the file system. Unwanted and repeated data in the file system is been clean up the process. So many same set of files in the file system may leads to the confusion, In this approach separate the sane set of files then they can easily identify the file.

6. REFERENCES:

[1]. J Jang, WK cheung-2005 EEE international conference using learning the kernel matrix for XML document clustering.

[2]. J Yuvan, X li, L Ma -2008 fifth international conference using XML document clustering path features.

[3]. F De Francesca, G Gordano and reasoning o 2003 using a frame work for XML document clustering.

[4]. A Doucet, H ahonen-myka proceedings of the 1st INEX 2002

[5]. H leung, F Chung, SCF chan0 international workshop on 2007 using XML document clustering path.

[6]. T Tran, R nayak, PD Bruza -2008 using combining structure and content similarities for XML document.

[7]. An efficient approach to clustering in large multimedia database with noise, A Hinnebrug, DA keim-KDD 1998-aai.org.

[8]. Distributed and mobility adaptive clustering foe multimedia support in multi hop wireless network, S Basagni-Gateway to 21st century communication village 1998 ieeeexplore.iee.org .

[9]. Clustering and classification of multimedia data, C Acharya, K Purang, M plutowski US patent 2010.