V

Video Data Mining

JungHwan Oh
University of Texas at Arlington, USA

JeongKyu Lee
University of Texas at Arlington, USA

Sae Hwang
University of Texas at Arlington, USA

INTRODUCTION

Data mining, which is defined as the process of extracting previously unknown knowledge and detecting interesting patterns from a massive set of data, has been an active research area. As a result, several commercial products and research prototypes are available nowadays. However, most of these studies have focused on corporate data — typically in an alpha-numeric database, and relatively less work has been pursued for the mining of multimedia data (Zaïane, Han, & Zhu, 2000). Digital multimedia differs from previous forms of combined media in that the bits representing texts, images, audios, and videos can be treated as data by computer programs (Simoff, Djeraba, & Zaïane, 2002). One facet of these diverse data in terms of underlying models and formats is that they are synchronized and integrated hence, can be treated as integrated data records. The collection of such integral data records constitutes a multimedia data set. The challenge of extracting meaningful patterns from such data sets has lead to research and development in the area of multimedia data mining. This is a challenging field due to the non-structured nature of multimedia data. Such ubiquitous data is required in many applications such as financial, medical, advertising and Command, Control, Communications and Intelligence (C3I) (Thuraisingham, Clifton, Maurer, & Ceruti, 2001). Multimedia databases are widespread and multimedia data sets are extremely large. There are tools for managing and searching within such collections, but the need for tools to extract hidden and useful knowledge embedded within multimedia data is becoming critical for many decision-making applications.

BACKGROUND

Multimedia data mining has been performed for different types of multimedia data: image, audio and video. Let us first consider image processing before discussing image and video data mining since they are related. Image processing has been around for some time. Images include maps, geological structures, biological structures, and many other entities. We have image processing applications in various domains including medical imaging for cancer detection, and processing satellite images for space and intelligence applications. Image processing has dealt with areas such as detecting abnormal patterns that deviate from the norm, and retrieving images by content (Thuraisingham, Clifton, Maurer, & Ceruti, 2001). The questions here are: *what* is image data mining and *how* does it differ from image processing? We can say that while image processing focuses on manipulating and analyzing images, image data mining is about finding useful patterns. Therefore, image data mining deals with making associations between different images from large image databases. One area of researches for image data mining is to detect unusual features. Its approach is to develop templates that generate several rules about the images, and apply the data mining tools to see if unusual patterns can be obtained. Note that detecting unusual patterns is not the only outcome of image mining; that is just the beginning. Since image data mining is an immature technology, researchers are continuing to develop techniques to classify, cluster, and associate images (Goh, Chang, & Cheng, 2001; Li, Li, Zhu, & Ogihara, 2002; Hsu, Dai, & Lee, 2003; Yanai, 2003; Müller & Pun, 2004). Image data mining is an area with applications in numerous domains including space, medicine, intelligence, and geoscience.

Mining video data is even more complicated than mining still image data. One can regard a video as a collection of related still images, but a video is a lot more than just an image collection. Video data management has been the subject of many studies. The important areas include developing query and retrieval techniques for video databases (Aref, Hammad, Catlin, Ilyas, Ghanem, Elmagarmid, & Marzouk, 2003). The question we ask yet again is what is the difference between video information retrieval and video mining? There is no clear-cut answer for this question yet. To be consistent with our terminology, we can say that finding correlations and patterns previously unknown from large video databases is video data mining.

MAIN THRUST

Even though we define video data mining as finding correlations and patterns previously unknown, the current status of video data mining remains mainly at the pre-processing stage, in which the preliminary issues such as video clustering, and video classification are being examined and studied for the actual mining. Only a very limited number of papers about finding any patterns from videos can be found. We discuss video clustering, video classification and pattern finding as follows.

Video Clustering

Clustering is a useful technique for the discovery of some knowledge from a dataset. It maps a data item into one of several clusters which are natural groupings for data items based on similarity metrics or probability density models (Mitra & Acharya, 2003). Clustering pertains to unsupervised learning, when data with class labels are not available. Clustering consists of partitioning data into homogeneous granules or groups, based on some objective function that maximizes the inter-cluster distances, while simultaneously minimizing the intra-cluster distances. Video clustering has some differences with conventional clustering algorithms. As mentioned earlier, due to the unstructured nature of video data, preprocessing of video data by using image processing or computer vision techniques is required to get structured format features. Another difference in video clustering is that the time factor should be considered while the video data is processed. Since video is a synchronized data of audio and visual data in terms of time, it is very important to consider the time factor. Traditional clustering algorithms can be categorized into two main types: partitional and hierarchical clustering (2003). Partitional clustering algorithms (i.e., *K-means* and *EM*) divide the patterns into a set of spherical clusters, while minimizing the objective function. Here the number of clusters is predefined. Hierarchical algorithms, on the other hand, can again be grouped as agglomerative and divisive. Here no assumption is made about the shape or number of clusters, and validity index is used to determine termination.

Two of the most popular partitional clustering algorithms are *K-means* and *Expectation Miximization (EM)*. In K-means, the initial centroids are selected, and each data item is classified to a cluster with the smallest distance. Based on the previous results, the cluster centroids are updated, and all corresponding data items are re-clustered until there is no centroid change. It is easily implemented, and provides a firm foundation of variances through the clusters. We can find the papers using the K-means algorithm for video clustering in the literature (Ngo, Pong, & Zhang, 2001). EM is a popular iterative refinement algorithm that belongs to the model-based clustering. It differs from the conventional K-means clustering algorithm in that each data point belongs to a cluster according to some weight or probability of membership. In other words, there are no strict boundaries between clusters. New means are computed based on weighted measures. It provides a statistical model for the data and is capable of handling the associated uncertainties. We can find the papers using the EM algorithm for video clustering in the literature (Lu, & Tan, 2002; Frey, & Jojic, 2003).

Hierarchical clustering methods create hierarchical nested partitions of the dataset, using a tree-structured dendogram and some termination criterion. Every cluster node contains child clusters; sibling clusters partition the points covered by their common parent. Such an approach allows exploring data on different levels of granularity. Hierarchical clustering methods are categorized into agglomerative (bottom-up) and divisive (top-down). An agglomerative clustering starts with one-point (singleton) clusters and recursively merges two or more of the most appropriate clusters. Divisive clustering starts with one cluster of all data points and recursively splits the most appropriate cluster. The process continues until a stopping criterion is achieved. The advantages of hierarchical clustering include: embedded flexibility regarding the level of granularity, ease of handling of any forms of similarity or distance, and applicability to any attribute types. The disadvantages of hierarchical clustering are vagueness of termination criteria, and the fact that most hierarchical algorithms do not revisit constructed (intermediate) clusters for the purpose of their improvement. Hierarchical clustering is used in video clustering because it is easy to handle the similarity of extracted features from video, and it can represent the depth and granularity by the level of tree (Okamoto, Yasugi, Babaguchi, & Kitahashi, 2002).

Video Classification

While clustering is an unsupervised learning method, classification is a way to categorize or assign class labels to a pattern set under the supervision. Decision boundaries are generated to discriminate between patterns belonging to different classes. The data set is initially partitioned into training and test sets, and the classifier is trained on the former. A framework to enable semantic video classification and indexing in a specific video domain (medical video) was proposed (Fan,

Related Content