TOOLS, TECHNIQUES AND MODELS FOR MULTIMEDIA DATABASE MINING

Chitra Wasnik
Department of Computer Engineering, Lokmanya Tilak College of Engineering University of Mumbai, India
chitrawasnik@yahoo.com

Abstract

Advances in multimedia acquisition and storage technology have led to tremendous growth in very large and detailed multimedia databases. Analyzing this huge amount of multimedia data to discover useful knowledge is a challenging problem. This challenge has opened the opportunity for research in Multimedia Data Mining (MDM). Multimedia data mining can be defined as the process of finding interesting patterns from media data such as audio, video, image and text that are not ordinarily accessible by basic queries and associated results. The motivation for doing MDM is to use the discovered patterns to improve decision making. MDM has therefore attracted significant research efforts in developing methods and tools to organize, manage, search and perform domain specific tasks for data from domains such as surveillance, meetings, broadcast news, sports, archives, movies, medical data, as well as personal and online media collections. If these multimedia files are analyzed, useful information to users can be revealed. Multimedia mining deals with the extraction of implicit knowledge, multimedia data relationships, or other patterns not explicitly stored in multimedia files. Multimedia mining is more than just an extension of data mining, as it is an interdisciplinary endeavor that draws upon expertise in computer vision, multimedia processing, multimedia retrieval, data mining, machine learning, database and artificial intelligence. This paper describes the tools, well known techniques and models for multimedia database mining.

Keywords: text mining; image mining; audio mining; video mining.

1. Introduction

At present, data mining is the most active branch of database research, development and applications, and it also has been studying the hot. In digital data acquisition and storage technology, the rapid progress has led to the fast growing tremendous and amount of data stored in databases. Although valuable information may be hiding behind the data, the overwhelming data volume makes it difficult (if not impossible) for human beings to extract them without powerful tools. Multimedia mining systems that can automatically extract semantically meaningful information (knowledge) from multimedia files are increasingly in demand. Data mining technology is gradually integrated into the multimedia databases. For this reason, a large number of techniques have been proposed ranging from simple measures (e.g. color histogram for image, energy estimates for audio signal) to more sophisticated systems like speaker emotion recognition in audio, automatic summarization of TV programs. Generally, multimedia database systems store and manage a large collection of multimedia objects, such as image, video, audio and hypertext data. Thus, in multimedia documents, knowledge discovery deals with non-structured information. For this reason, we need tools for discovering relationships between objects or segments within multimedia document components, such as classifying images based on their content, extracting patterns in sound, categorizing speech and music, and recognizing and tracking objects in video streams. In general, the multimedia files from a database must be first preprocessed to improve their quality. Subsequently, these multimedia files undergo various transformations and features extraction to generate the important features from the multimedia files. A way to apply advanced data mining techniques is to have a flexible and interactive data mining tool that is fully integrated with a database or data warehouse. Using a tool that operates outside of the database or data warehouse is not as efficient. Using such a tool will involve extra steps to extract, import, and analyze the data. When a data mining tool is integrated with the data warehouse, it simplifies the application and implementation of mining results. Furthermore, as the warehouse grows with new decisions and results, the organization can mine best practices continually and apply them to future decisions. Regardless of the technique used, the real value behind data mining is modeling — the process of building a model based on user-specified criteria from already captured data.

2. Feature Extraction

There are two kinds of features: description-based and content-based. The former uses metadata, such as keywords, caption, size and time of creation. The later is based on the content of the object itself.

2.1 Feature extraction from text

Text categorization is a conventional classification problem applied to the textual domain. It solves the problem of assigning text content to predefined categories. In the learning stage, the labeled training data are first pre-processed to remove unwanted details and...
to “normalize” the data. For example, in text documents punctuation symbols and non-alphanumeric characters are usually discarded, because they do not help in classification. Moreover, all characters are usually converted to lower case to simplify matters. The next step is to compute the features that are useful to distinguish one class from another. For a text document, this usually means identifying the keywords that summarize the contents of the document. How are these keywords learned? One way is to look for words that occur frequently in the document. These words tend to be what the document is about. Of course, words that occur too frequently, such as “the”, “is”, “in”, “of” are no help at all, since they are prevalent in every document. These common English words may be removed using a “stop-list” of words during the preprocessing stage. From the remaining words, a good heuristic is to look for words that occur frequently in documents of the same class, but rarely in documents of other classes. In order to cope with documents of different lengths, relative frequency is preferred over absolute frequency. Some authors used phrases, rather than individual words, as indexing terms, but the experimental results found to date have not been uniformly encouraging results. Another problem of text is the variant. Variant refers to the different forms of the same word, e.g. “go”, “goes”, “went”, “gone”, “going”. This may be solved by stemming, which means replacing all variants of a word by a standard one.

2.2 Feature extraction from images

Image categorization classifies images into semantic databases that are manually precategorized. In the same semantic databases, images may have large variations with dissimilar visual descriptions (e.g. images of persons, images of industries etc.). In addition, images from different semantic databases might share a common background (some flowers and sunset have similar colors). There are three types of feature vectors for image description are distinguished: 1) pixel level features, 2) region level features, and 3) tile level features. Pixel level features store spectral and textural information about each pixel of the image. For example, the fraction of the end members, such as concrete or water, can describe the content of the pixels. Region level features describe groups of pixels. Following the segmentation process, each region is described by its boundary and a number of attributes, which present information about the content of the region in terms of the end members and texture, shape, size, fractal scale etc. Tile level for image features present information about whole images using texture, percentages of end members, fractal scale and others. Moreover, other researchers proposed an information-driven framework that aims to highlight the role of information at various levels of representation. This framework adds one more level of information: the Pattern and Knowledge Level that integrates domain, related alphanumeric data and the semantic relationships discovered from the image data.

2.3 Feature extraction from Audio

Audio data play an important role in multimedia applications. Music information has two main branches: symbolic and audio information. Attack, duration, volume, velocity and instrument type of every single note are available information. Therefore, it is possible to easily access statistical measures such as tempo and mean key for each music item. Moreover, it is possible to attach to each item high-level descriptors, such as instrument kind and number. On the other hand, audio information deals with real world signals and any features need to be extracted through signal analysis. only perceptual features such as loudness, brightness, pitch etc.

2.4 Feature extraction from Video

In video mining, there are three types of videos: a) the produced (e.g. movies, news videos, and dramas), b) the raw (e.g. traffic videos, surveillance videos etc.), and c) the medical video (e.g. ultra sound videos including echocardiogram). Higher-level information from video includes: detecting trigger events (e.g. any vehicles entering a particular area, people exiting or entering a particular building) determining typical and anomalous patterns of activity, generating person-centric or object-centric views of an activity classifying activities into named categories (e.g. walking, riding a bicycle), clustering and determining interactions between entities. The first stage for mining raw video data is grouping input frames to a set of basic units, which are relevant to the structure of the video. In produced videos, the most widely used basic unit is a shot, which is defined as a collection of frames recorded from a single camera operation. Shot detection methods can be classified into many categories: pixel based, statistics based, transform based, feature based and histogram based. Color or grayscale histograms (such as in image mining) can also be used. To segment video, color histograms, as well as motion and texture features can be used. Generally, if the difference between the two consecutive frames is larger than a certain threshold value, then a shot boundary is considered between two corresponding frames. The difference can be determined by comparing the corresponding pixels of two images.

3. DATA MINING TOOLS

Organizations that wish to use data mining tools can purchase mining programs designed for existing software and hardware platforms, which can be integrated into new products and systems as they are brought online, or they can build their own custom mining solution. For instance, feeding the output of a data mining exercise into another computer system, such as a neural network, is quite common and can give the mined data more value. This is because the data mining tool gathers the data, while the second program (e.g., the neural network) makes decisions based on the data collected. Different types of data mining tools are available in the marketplace, each with their own strengths and weaknesses. Internal auditors need to be aware of the different kinds of data mining tools available and recommend the purchase of a tool that matches the organization’s current detective needs. This should be considered as early as possible in the project’s lifecycle, perhaps even in the feasibility study. Most data mining tools can be classified into one of three categories: traditional data mining tools, dashboards, and text-mining tools. Below is a description of each.
3.1 Traditional Data Mining Tools Traditional data mining programs help companies establish data patterns and trends by using a number of complex algorithms and techniques. Some of these tools are installed on the desktop to monitor the data and highlight trends and others capture information residing outside a database. The majority are available in both Windows and UNIX versions, although some specialize in one operating system only. In addition, while some may concentrate on one database type, most will be able to handle any data using online analytical processing or a similar technology.

3.2 Dashboards Installed in computers to monitor information in a database, dashboards reflect data changes and updates onscreen — often in the form of a chart or table — enabling the user to see how the business is performing. Historical data also can be referenced, enabling the user to see where things have changed (e.g., increase in sales from the same period last year). This functionality makes dashboards easy to use and particularly appealing to managers who wish to have an overview of the company's performance.

3.3 Text-mining Tools The third type of data mining tool sometimes is called a text-mining tool because of its ability to mine data from different kinds of text — from Microsoft Word and Acrobat PDF documents to simple text files, for example. These tools scan content and convert the selected data into a format that is compatible with the tool's database, thus providing users with an easy and convenient way of accessing data without the need to open different applications. Scanned content can be unstructured (i.e., information is scattered almost randomly across the document, including e-mails, Internet pages, audio and video data) or structured (i.e., the data's form and purpose is known, such as content found in a database). Capturing these inputs can provide organizations with a wealth of information that can be mined to discover trends, concepts, and attitudes.

4. DATA MINING TECHNIQUES
In addition to using a particular data mining tool, internal auditors can choose from a variety of data mining techniques. The most commonly used techniques include artificial neural networks, decision trees, and the nearest-neighbor method. Each of these techniques analyzes data in different ways:

4.1 Artificial neural networks are non-linear, predictive models that learn through training. Although they are powerful predictive modeling techniques, some of the power comes at the expense of ease of use and deployment. One area where auditors can easily use them is when reviewing records to identify fraud and fraud-like actions. Because of their complexity, they are better employed in situations where they can be used and reused, such as reviewing credit card transactions every month to check for anomalies.

4.2 Decision trees are tree-shaped structures that represent decision sets. These decisions generate rules, which then are used to classify data. Decision trees are the favored technique for building understandable models. Auditors can use them to assess, for example, whether the organization is using an appropriate cost-effective marketing strategy that is based on the assigned value of the customer, such as profit.

4.3 The nearest-neighbor method classifies dataset records based on similar data in a historical dataset. Auditors can use this approach to define a document that is interesting to them and ask the system to search for similar items.

5. MODELS FOR MULTIMEDIA MINING
Multimedia classification and clustering are the supervised and unsupervised classification of multimedia files into groups.

5.1 Classification models
Machine learning (ML) and meaningful information extraction can only be realized, when some objects have been identified and recognized by the machine. The object recognition problem can be referred as a supervised labeling problem. Starting with the supervised models, we mention the decision trees. Decision trees can be translated into a set of rules by creating a separate rule for each path from the root to a leaf in the tree. However, rules can also be directly induced from training data using a variety of rule-based algorithms. Artificial Neural Networks (ANNs) are another method of inductive learning, based on computational models of biological neurons and networks. Instance-based learning algorithms are lazy-learning algorithms as they delay the induction or generalization process until classification is performed. During the training phase, the lazy-learning algorithms require less computation time than eager-learning algorithms (e.g. decision trees, neural and Bayes nets). However, during the classification process, they require more computation time. The Support Vector Machines (SVMs) is the newest technique that considers the notion of a “margin”. Maximising the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it, is proven to reduce an upper bound on the expected generalisation error.

5.2 Clustering Models
In unsupervised classification, the problem is to group a given collection of unlabeled multimedia files into meaningful clusters according to the multimedia content without a priori knowledge. Clustering algorithms can be categorized into partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods. Partitioning methods are divided into two major subcategories, the centroid and the medoids algorithms. The centroid algorithms represent each cluster by using the gravity centre of the instances. The medoid algorithms represent each cluster by means of the instances closest to the gravity centre. The hierarchical methods group data instances into a tree of clusters. Density-based clustering algorithms try to find clusters based on
density of data points in a region. The key idea of density-based clustering is that, for each instance of a cluster, the neighborhood of a given radius has to contain at least a minimum number of instances. Grid-based clustering algorithms first quantize the clustering space into a finite number of cells (hyper-rectangles) and then perform the required operations on the quantized space. Cells that contain more than a certain number of points are treated as dense and the dense cells are connected to form the clusters.

5.3 Association rules

The most association rules studies have been focusing on the corporate data typically in alphanumeric databases. There are three measures of the association: support, confidence and interest. The support factor indicates the relative occurrence of both X and Y within the overall data set of transactions. It is the total number of instances. The confidence factor is the probability of Y given X and is defined as the ratio of the number of instances satisfying both X and Y over the number of instances satisfying X. The support factor indicates the frequencies of the occurring patterns in the rule, and the confidence factor denotes the strength of implication of the rule. The interest factor is a measure of human interest in the rule. For example, a high interest means that if a transaction contains X, then it is much more likely to have Y than the other items. Relatively little research has been conducted on mining multimedia data. There are different types of associations: association between image content and non-image content features. For example, if the upper part of the picture is at least 50% blue, it is likely to represent sky. Association mining in multimedia data can be transformed into problems of association mining in traditional transactional databases. The image is can be modeled as a transaction, assigned with an ImageID, and the features of the images are the items contained in the transaction. Therefore, mining the frequently occurring patterns among different images becomes mining the frequent patterns in a set of transactions.

6. CONCLUSIONS

This paper describes tools, techniques and methods for multimedia mining. In text mining there are two open problems: polysemy, synonymy. Polysemy refers to the fact that a word can have multiple meanings. Distinguishing between different meanings of a word (called word sense disambiguation) is not easy, often requiring the context in which the word appears. Synonymy means that different words can have the same or similar meaning. In audio and video mining, a fundamental open problem also remains: The combination of information across multiple media (combining video and audio information into one comprehensive score). In image mining an open problem remains: the combination of different types of image data. Documents from an OCR library and a video library need to be presented in a single ranked list. It also describes the tool, techniques and models for multimedia database mining.

References


**Author**

Prof Chitra T Wasnik, B.Eng in Computer Science and Engineering and M.Tech in Computer Science and Engineering, is an assistant professor in the Department of Computer Engineering at Lokmanya Tilak College of Engineering, University of Mumbai, India. With over 15 years of teaching experience, Prof Wasnik has presented a number of papers on various topics in national and international conferences. She is a life member of ISTE.