Middle-East Journal of Scientific Research 13 (12): 1556-1558, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.13.12.1510

A New Method for Data Mining in Multimedia Environment

S. Pothumani

Department of Computer Science and Engineering, Bharath University, India

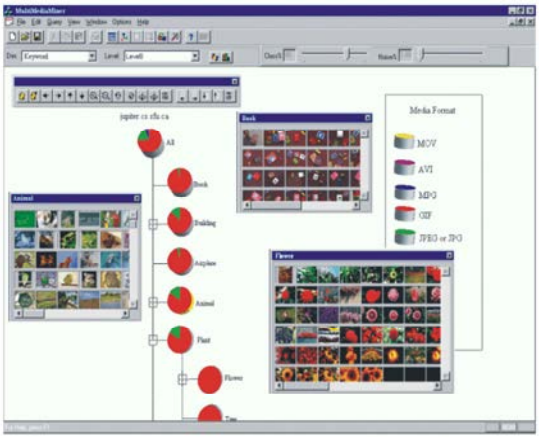
Abstract: Most of the existing frequent item sets mining techniques are based up on Multimedia data mining. In this paper we propose a novel approach for frequent item sets mining using color, texture and shape. Frequent item set is an item set that satisfies minimum support. The data bases tested in the Multimedia Miner System is constructed. Each Image contains two descriptors: a feature descriptor and a Layout descriptor. The original Image is not stored directly in the database only its descriptors are stored. The description information encompasses fields like Image file name, Image URL, Image type like gif, tiff, jpeg, mpeg, bmp, avi. A list of all known web pages referring to the image i.e parent URLs. A list of keywords and a thumbnail used by the user interface for image and video Browsing. The feature descriptor is set of vectors for each visual characteristic. The main vectors are a color vector containing the color histogram quantized to 512 colors (8*8 for R*G*B, an Most Frequent Color vector and Most Frequent Orientation Vector. The image Excavator component of Multimedia Minor uses image contextual information.

Key words: Data Mining · Multimedia · Frequent item set

INTRODUCTION

Multimedia data cube seems to be an interesting model for multidimensional analysis of multi media data we should note that it is difficult to implement a data cube efficiently given a large number of dimensions. This curse of dimensionality is especially serious in case of multimedia data cubes. It may like to model color, Orientation, texture, keywords, soon. As multiple dimensions in multimedia data cube many of these attributes are set oriented instead of single, valued for eg: one Image may correspond to a set of keywords. It may contain a set of objects, each associated with a set of colors. If use each keyword as a dimension or each detailed color as a dimension in the design of the data cube. It will create huge number of dimensions. on the other hand not doing so may lead to the modeling of an image at rather rough, limited and imprecise scale. More research is needed on how to design a multimedia data cube like a balance between efficiency and the power of representation.

Implementation: Classification and predictive modeling have been used for mining multimedia data especially in scientific research, such as astronomy seismology and



geoscientific research. We construct models for the recognition of galaxies, stars and other stellar objects based on properties like magnitudes, areas, intensity, image moments and orientation. A large number of sky images taken by telescopes or space probes can be tested against the constructed models in order to identify volcanoes on Venus. Data processing is important when mining image data and can include data cleaning, data transformation and feature extraction and though transformations, techniques can be explored such as decomposition of images to eigenvectors or the adoption

Corresponding Author: S. Pothumani, Department of Computer Science and Engineering, Bharath University, India.

of probabilistic models to deal with uncertainty since the Image data are often in huge volumes and may require substantial processing power, parallel and distributed processing are useful. Image data mining classification and clustering are closely inked to image analysis and data mining and thus many image analysis techniques and scientific data analysis methods can be applied to image data mining the popular use of world wide web has made the web a rich and gigantic repository of multimedia data [1]. The web not only collects a tremendous number of photos, pictures, albums and video images in the form of online multimedia libraries.

The web pages are linked with another web page in a complicated way. such text, image location and web linkage information, if used properly may help understand the context of the text or assist classification and linkages among images, text, blocks with in a page and page links on the web becomes an important direction in web analysis. Association rules involving multimedia objects can be mined in image and video databases. At least three categories can be observed:

- Association between image content and non image content features the upper part of the picture is blue then it is likely to represent sky.
- Association among image contents that are not related to spatial relationships like if a picture contains two blue squares, then it is likely to contain on red circle.
- Association among image contents related to spatial relationships like if a red triangle is between two yellow squares then it is likely a big oval-shaped object is underneath

An may contain multiple objects each with many features such as color, shape, texture, keyword and spatial location, so there could be many possible associations in many cases a feature may be considered as the same in two images at certain level of resolution, but different at a finer resolution level therefore it is essential to promote a progressive resolution refinement approach. That is we can first mine frequently occurring patterns at relatively rough resolution level and then focus only on those that have passed the minimum support threshold when mining at a finer resolution this is because the patterns that are not frequent at a rough level cannot be frequent at finer resolution levels. Such a multi resolution mining strategy substantially reduces the overall data mining cost without loss of the quality and completeness of data mining results. This leads to an efficient methodology for mining frequent item sets and associations in large multimedia data bases.

A picture containing multiple recurrent objects is an important feature in image analysis, recurrence of the same objects should not be ignored in association analysis. A picture containing two golden circles is treated quite differently from the containing only one this is quite different from that in transaction database, where the fact that a person buys a gallon of milk are two or may often be treated the same as “buys milk’. The definition of multimedia associations and its measurements such as support and confidence should be adjusted accordingly [2].

Important spatial relationships among multimedia objects such as above, beneath, between, nearby, left-off and soon. These features are very useful for exploring object associations and correlations. Spatial relationships with other content based multimedia features, such as color, shape, texture and keywords form interesting associations. Thus spatial data mining methods and properties of topological spatial relationships become important for multimedia mining. Besides still images, audio visual information is becoming available digital form, in digital archives, on world wide web, in broadcast data streams and in personal and professional data bases [3].

This amount is rapidly growing. There are great demands for effective content base retrieval and data mining methods for audio and video data. Typical examples include searching for and multimedia editing of particular video clips in a TV studio, detecting suspicious persons or sheens in surveillance videos, searching for particular events in a personal multimedia repository such as MyLifeBits, a discovering patterns and outliers in weather radar recordings and finding a particular melody in our mp3 audio album [4]. To facilitate the recording, search and analysis of audio and video information from multimedia data, industry and standardization committees have made great strides toward developing a set of standards for multimedia information description and compression [5-8].

For example MPEG-k Moving Picture Experts Group and JPEG are Typical video compressions schemes. The most recently MPEG-7, formally named Multimedia Content Description Interface is a standard for describing the multimedia content data. It supports some degree of interpretation of the information meaning, which can be passed on to, or accessed by, a device or a computer.

The audio visual data description in MPEG-7 includes still pictures, video, graphics, audio, speech, three-dimensional models and information about how these data elements are combine in the multimedia presentation. MPEG committee standardizes the following elements in MPEG-7 [9, 10].

- A set of descriptors where each descriptor defines the syntax and semantics of a feature, such as Color, shape, texture, image, topology, motion, title.
- A set of descriptor schemes where each scheme specifies the structure and semantics of relationships between its components.
- A set of coding schemes for the descriptors.
- A description definition language (DDL) to specify the schemes and descriptors. such standardization greatly facilitates content based video retrieval and video data mining. mining is a very popular data mining technique and it finds relationships among the different entities of records. The introduction of frequent itemset has received a great deal of attention in the field of knowledge discovery and data mining.

CONCLUSION

In this paper, we have presented a Multimedia Datamining association between the image, color, shape, Audio and video visualization predictive analysis multimedia datamining [11-13].

Future Scope: Video data mining is till in its invancy. There are still a lot of research issues to be solved before it becomes general practice similarity-based preprocessing, compressing, indexing and retrieval, information extraction redundancy removal, frequent pattern discovery, classification, clustering and trend and outlier detection are important data mining tasks in this domain

REFERENCES

1. Kumaravel, A., 2013. Introducing an Efficient Programming Paradigm for Object-oriented Distributed Systems, Indian Journal of Science and Technology, 6(5S): 4597-4603.
2. Kumaravel, A., 2013. Application of Non-Uniform Cellular Efficient Cryptography Automata, Indian Journal of Science and Technology, 6(5S): 4561-4566.
3. Kumaravel, A., 2013. Routing Algorithm over Semi-regular Tessellations, Xplore, pp: 1180-1184.
4. Kumaravel, A., 2013. AN Application of Non-uniform Cellular Automata for Efficient Cryptography, Xplore, pp: 1200-1205.
5. Han, J., J. Pei and Y. Yin, 2000. Mining Frequent of Data, pp: 1-12, May. Patterns without Candidate Generation, Proc. ACM SIGMOD Int'l Conf. Management.
6. Zaki, M.J., S. Parthasarathy, M. Ogihara and W. Li, 1997. NewAlgorithms for Fast Discovery of Association Rules, Proc. Third nt'l Conf. Knowledge Discovery and Data Mining, pp: 283-286.
7. Agrawal, R., C. Aggarwal and V. Prasad, 2000. A Tree Projection Algorithm for Generation of Frequent Item Sets, Parallel and Distributed Computing, pp: 350-371.
8. Zaki, M.J., S. Parthasarathy, M. Ogihara and W. Li, 1997. New Algorithms for Fast Discovery of Association Rules, Proc. Third Int'l Conf. Knowledge Discovery and Data Mining, pp: 283-286.
9. Goethals, B., 2003. Survey on Frequent Pattern Mining, manuscript.
10. Kumaravel, A., 2013. Algorithm for Automaton Specification for Exploring Dynamic Labyrinths, Indian Journal of Science and Technology, 6(5).
11. Pattanayak, Monalisa and P.L. Nayak, 2013. Green Synthesis of Gold Nanoparticles Using Elettaria cardamomum (ELAICHI) Aqueous Extract World Journal of Nano Science and Technology, 2(1): 01-05.
12. Chahataray, Rajashree and P.L. Nayak, 2013. Synthesis and Characterization of Conducting Polymers Multi Walled Carbon Nanotube-Chitosan Composites Coupled with Poly (P-Aminophenol) World Journal of Nano Science and Technology, 2(1): 18-25.
13. Parida, Umesh Kumar, S.K. Biswal, P.L. Nayak and B.K. Bindhani, 2013. Gold Nano Particles for Biomedical Applications World Journal of Nano Science and Technology, 2(1): 47-57.