MDM/KDD2002: Multimedia Data Mining between Promises and Problems

Simeon J. Simoff
Faculty of Information Technology, University of Technology, Sydney, Australia
simeon@it.uts.edu.au

Chabane Djeraba
IRIN, Nantes University, 2, Rue de la Houssiniere, 44322 Nantes Cedex, France
djeraba@irin.univ-nantes.fr

ABSTRACT

This report presents a brief overview of multimedia data mining and the corresponding workshop series at ACM SIGKDD conference series on data mining and knowledge discovery. It summarizes the presentations, conclusions and directions for future work that were discussed during the 3rd edition of the International Workshop on Multimedia Data Mining, conducted in conjunction with KDD-2002 in Edmonton, Alberta, Canada.

Keywords

Multimedia data mining, knowledge discovery, multimedia databases, image content mining, digital media, sound analysis, video analysis.

1. INTRODUCTION

Digital multimedia differs from previous forms of combined media in that the bits that represent text, images, animations, and audio, video and other signals can be treated as data by computer programs. One facet of this diverse data in terms of underlying models and formats is that it is synchronized and integrated, hence, can be treated as *integral data records*. Such records can be found in a number of areas of human endeavor. Modern medicine generates huge amounts of such digital data. For example, a medical data record may include SPECT images, DNA microarray data, ECG signals, clinical measurements, like blood pressure and cholesterol levels, and the description of the diagnosis given by the physician interpretation of all these data. Another example is architecture design and related Architecture, Engineering and Construction (AEC) industry. An architectural design data record may include CAD files with floor plans and data for generating 3D-models, links to building component data bases and the instances of this components, stored in corresponding formats, images of the environment, text descriptions and drawings from the initial and evolved design requirements, financial and personnel data, and other data related to the other disciplines involved in the AEC industry, e.g. civil and electrical engineering. Virtual communities (in the broad sense of this word, which includes any communities mediated by digital technologies) are another example where generated data constitutes an integral data record. Such data may include data about member profiles, the content generated by the virtual community, and communication data in different formats, including email, chat records, SMS messages, video conferencing records. Not all multimedia data is so diverse. An example of less diverse, but larger in terms of the collected amount, is the data generated by video surveillance systems, where each integral data record roughly consists of a set of time-stamped images – the video frames. In any case, 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 the 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, if not essential, in many applications. 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 useful knowledge embedded within multimedia data is becoming critical for many decision-making applications. The tools needed today are tools for discovering relationships between data items or segments within images, classifying images based on their content, extracting patterns from sound, categorizing speech and music, recognizing and tracking objects in video streams, relations between different multimedia components, and cross-media object relations.

2. MDM/KDD WORKSHOP SERIES

The workshop series was initiated in early 1998, however, the first successful event was staged at KDD2000. Since the beginning of the century, the MDM/KDD workshops at the ACM SIGKDD forums (MDM/KDD2000, MDM/KDD2001, MDM/KDD2002 in conjunction with KDD2000 (in Boston), KDD2001 (in San Francisco) and KDD2003 (in Edmonton), respectively) have been bringing together cross-disciplinary experts in analysis of digital multimedia content, multimedia databases, spatial data analysis, analysis of data in collaborative virtual environments, and knowledge engineers and domain experts from different applied disciplines related to multimedia data mining. MDD/KDD2000 revealed a variety of topics that come under the umbrella of multimedia data mining. Accepted papers were presented in three sessions: Mining spatial multimedia data (3 papers); Mining audio data and multimedia support (5 papers); and Mining image and video data (6 papers) [1]. The presentations at the workshop showed that many researchers and developers in the areas of multimedia information systems and digital media turn to data mining methods for techniques that can improve indexing and retrieval in digital media. The workshop ended with a discussion on the scope of multimedia data mining and the need to establish an annual forum. When text, maps, video, sound, and images typically fall into the realm of multimedia, research fields such as spatial data mining and text mining were already known active disciplines. There was a consensus that multimedia data mining is emerging as its own distinct area of research and development. The work in the area was expected to focus on algorithms and methods for mining from images, sound and video streams. Workshop participants identified that there is a need for (i) development and application of specific methods, techniques and tools for multimedia data mining; and (ii) frameworks that provide consistent methodology for multimedia data analysis and integration of discovered knowledge back in the system where it

can be utilized. These conclusions and to some extent decisions influenced the selection of the papers for presentation at MDM/KDD2001, grouped in the following streams: Frameworks for multimedia mining (2 papers); Multimedia mining for information retrieval (4 papers); and Applications of multimedia mining (6 papers) [2]. The workshop discussion revised the scope of multimedia data mining outlined during the previous workshop, clearly identifying the need to approach multimedia data as a "single unit" rather than ignoring some layers in favor of others. The participants acknowledged the high potential of multimedia data mining methods in medical domains, design and creative industries. There was an agreement that the research and development in multimedia mining should be extended in the area of collaborative virtual environments, 3D virtual reality systems, musical domain and e-business technologies.

3. THIS YEAR WORKSHOP

The 3rd International Workshop on Multimedia Data mining attempted to address the above-mentioned issues looking at specific issues in pattern extraction from image data, sound, and video; suitable multimedia representations and formats that can assist multimedia data mining; and advanced architectures of multimedia data mining systems. Papers, selected for presentation provided an interesting coverage of these issues and some technical solutions. They were grouped in the following three streams: *Frameworks for Multimedia Data Mining* (2 papers); *Multimedia Data Mining Methods and Algorithms* (4 papers); and *Applications of Multimedia Data Mining* including *applications in medical image analysis* (2 papers) and applications in content-based multimedia processing (4 papers).

3.1 Frameworks for Multimedia Data Mining

In *Multimedia Data Mining Framework for Raw Video Sequences* by JungHwan Oh and Babitha Bandi (Department of Computer Science and Engineering, University of Texas at Arlington, USA) presented a general framework for real time video data mining from "raw videos" (e.g. traffic videos, surveillance videos). In proposed technique the first step for mining raw video data was grouping the input frames into a set of basic units - segments, that became the organizational blocks of the video database for video data mining. Then, based on some features (motion, object, colors, etc.) extracted from each segment, the segments could be clustered into similar groups for detecting interesting patterns. The focus within the presented framework was on the motion as a feature, and how to compute and represent it for further processing. The multi-level hierarchical segment clustering procedure used category and motion. Preliminary experimental results were promising. Readers may see also the framework for multimedia data mining from traffic video sequences presented by a joint research team from Florida International University at an earlier workshop [3].

In *An Innovative Concept for Image Information Mining*, Mihai Datcu (Remote Sensing Technology Institute – IMF, German Aerospace Center – DLR, Germany) and Klaus Seidel (Computer Vision Lab, ETH, Switzerland) introduced the concept of 'image information mining' and discussed a system that implements this concept. The approach is based on modeling the causalities, which link the image-signal contents to the objects and structures within the interest of the users. The basic idea is to split the information representation into four steps: (1) image feature extraction using a library of algorithms to obtain a quasi-complete signal description; (2) unsupervised grouping in a large number of clusters to be suitable for a large set of tasks; (3) data reduction by parametric modeling of the clusters, and; (4) supervised learning of user semantics, that is the level where, instead of being programmed, the system is trained by a set of examples. Through these steps the links from image content to user semantics are created. Step 4 employs advanced visualization tools. At the time of the workshop the system has been prototyped for inclusion in a new generation of intelligent satellite ground segment systems, value-adding tools in the area of geo-information, and several applications in medicine and biometrics.

3.2 Multimedia Data Mining Methods and Algorithms

The first presentation in the session, on *Multimedia Data Mining Using P-trees*, focused on the specific research work of the DataSURG group at North Dakota State University, USA (co-authored by William Perrizo, William Jockheck, Amal Perera, Dongmei Ren, Weihua Wu and Yi Zhang). The group came to data mining from the context of evaluation of remotely sensed images for use in agricultural applications. During that research a spatial data structure was developed that provided an efficient, loss-less, data mining ready representation of the data. In time the similarity of a sequence of remotely sensed images and some categories of multimedia data became apparent. The Peano Count Tree (P-tree) technology provided an efficient way to store and mine sets of images and related data. For most multimedia data mining applications, feature extraction converts the pertinent data to a structured relational or tabular form, and then the tuples or rows are analyzed. Proposed P-tree data structure design addressed such data mining setting.

In *Scale Space Exploration for Mining Image Information Content* Mariana Ciucu, Patrick Heas, Mihai Datcu (Remote Sensing Technology Institute – IMF, German Aerospace Center – DLR, Germany) and James C. Tilton (NASA's Goddard Space Flight Center, USA) described an application of a scale space-clustering algorithm (melting) for exploration of image information content. Clustering by melting considers the feature space as a thermo-dynamical ensemble and groups the data by minimizing the free energy, having the temperature as a scale parameter. The authors developed clustering by melting for multidimensional data, and proposed and demonstrated a solution for the initialization of the algorithm. Due to computational reasons due to the curse of dimensionality, for initialization of clusters they had chosen the initial clusters centers with another algorithm, which performed fast cluster estimation with low computation cost. They further analyzed the information extracted by melting and proposed an information representation structure that enabled exploration of image content. This structure is a tree in the scale space showing how the clusters merge. The implementation of the algorithm has been done through a multi-tree structure. With this structure, they were able to explore the image content as an information mining function and to obtain a more compact data structure. They had maximum of information in scale space because they memorized the bifurcation points and the trajectories of the centers points in the scale space. The information encoded in the tree structure enabled the fast reconstruction and exploration of the data cluster structure and the investigation of hierarchical sequences of image classifications. The research team demonstrated examples using satellite

multispectral image (SPOT 4) and Synthetic Aperture Radar – SAR and Digital Elevation Models – DEM derived from SAR interferometry (SRTM). The paper was presented by Klaus Seidel.

In *Multimedia Knowledge Integration, Summarization and Evaluation* Ana B. Benitez and Shih-Fu Chang (Columbia University, New York, USA) presented new methods for automatically integrating, summarizing and evaluating multimedia knowledge. The work tackled issues, essential for multimedia applications to efficiently and coherently deal with multimedia knowledge at different abstraction levels such as perceptual and semantic knowledge (e.g., image clusters and word senses, respectively). The proposed methods included automatic techniques for: (1) interrelating the concepts in the multimedia knowledge using probabilistic Bayesian learning; (2) reducing the size of multimedia knowledge by clustering the concepts and collapsing the relationships among the clusters, and; (3) evaluating the quality of multimedia knowledge using notions from information and graph theory. Presented experimental results showed the potential of knowledge integration techniques in improving the knowledge quality, the importance of good concept distance measures for clustering and summarizing knowledge, and the usefulness of automatic measures in comparing the effects of different processing techniques on multimedia knowledge.

In the paper *Object Boundary Detection for Ontology-based Image Classification*, Lei Wang, Latifur Khan and Casey Breen (University of Texas at Dallas, USA) discussed a search mechanism guaranteeing delivery of minimal irrelevant information (high precision) while insuring that relevant information was not overlooked (high recall). The traditional solution works well in performing searches in which the user specifies images containing a sample object, or a sample textural pattern, in which the object or pattern is indexed. One can overcome this restriction by indexing images according to meanings rather than objects that appear in images. The authors proposed a solution to the problem of creating a meaning based index structure through the design and implementation of a concept-based model using domain-dependent ontologies. With regard to converting objects to meaning the key issue was to identify appropriate concepts that both described and identified images. Aiming at accurate identification of object boundaries, the authors proposed an automatic scalable object boundary detection algorithm based on edge detection and region growing techniques, and an efficient merging algorithm to join adjacent regions using an adjacency graph to avoid the over-segmentation of regions. They implemented a very basic system aimed at the classification of images in the sports domain, in order to illustrate the effectiveness of the algorithm. The results showed that the approach worked well when objects in images had less complex organization. Similar approach for retrieval of audio information was presented at MDM/KDD2000 [4]. Unfortunately, this year none of the authors could make it to the workshop to discuss the issues and results of this research.

3.3 Applications of Multimedia Data Mining

3.3.1 Applications in Medical Image Analysis

Osmar R. Zaïane, Maria-Luiza Antonie and Alexandru Coman (University of Alberta, Edmonton, Canada) presented in *Mammography Classification by an Association Rule-based Classifier* the continuation of their work in classification of medical (mammography) images, discussed at MDM/KDD2001 [5]. They proposed an association rule-based classifier, that was tested on a real dataset of medical images. The proposed approach included a significant (and important) pre-processing phase, a phase for mining the resulted transactional database, and a final phase to organize derived association rules in a classification model. Presented experimental results showed that the method performed well reaching over 80% in accuracy. Authors emphasized the importance of the data-cleaning phase in building an accurate data mining architecture for image classification.

In *An Application of Data Mining in Detection of Myocardial Ischemia Utilizing Pre- and Post-Stress Echo Images* Pramod K. Singh, Simeon J. Simoff (University of Technology Sydney, Australia) and David Feng (University of Sydney, Australia) presented the initial work in the development of a data mining approach for computer-assisted detection of myocardial ischemia. Proposed approach is based on automatic identification of endocardial and epicardial boundaries of left ventricle (LV) from images generated from echocardiogram data. These images are of poor quality and similar to the mammography images, discussed in the previous presentation, require significant preprocessing. The overall algorithm includes LV-wall boundary identification, segmentation and further comparative analysis of wall segments in pre- and post stress echocardiograms.

3.3.2 Applications in Content-Based Multimedia Processing

In *From Data to Insight: The Community of Multimedia Agents* Gang Wei, Valery A. Petrushin and Anatole V. Gershman (Accenture Technology Labs, Chicago, USA) presented a work devoted to creating an open environment for developing, testing, learning and prototyping multimedia content analysis and annotation methods. The work is part of the multimedia agents project COMMA. The environment served as a medium for researchers to contribute and share their achievements while protecting their proprietary techniques. Each method was represented as an agent that could communicate with the other agents registered in the environment using templates that were based on the Descriptors and Description Schemes in the emerging MPEG-7 standard. Such approach allowed agents developed by different organizations to operate and communicate with each other seamlessly regardless of their programming languages and internal architecture. To facilitate the construction of media analysis methods, the research team provided a development environment. This environment enabled researchers to compare the performance of different agents and combine them in more powerful and robust system prototypes. The COMMA could also serve as a learning environment for researchers and students to acquire and test cutting edge multimedia analysis algorithms, and sharing media agents. Some background to this work can be found in Valery's previous work on the PERSEUS project [6].

In *A Content Based Video Description Scheme and Video Database Navigator* Sadiye Guler and Ian Pushee (Northrop Grumman Information Technology, USA), introduced a unified framework for a comprehensive video description scheme and presented the "database navigator" – a browsing and manipulation tool for video data mining. The proposed description scheme was based on the structure and the semantics of the video, incorporating scene, camera, and object and behavior information pertaining to a large class of video data. The Database Navigator

was designed to exploit both the hierarchical structure of video data, the clips, shots and objects, as well as the semantic structure, such as scene geometry the object behaviors. The navigator provided means for visual data mining of multimedia data: intuitive presentation, interactive manipulation, ability to visualize the information and data from a number of perspectives, and to annotate and correlate the data in the video database. The authors demonstrated a working prototype.

In *Subjective Interpretation of Complex Data: Requirements for Supporting Kansei Mining Process* Nadia Bianchi-Berthouze and Tomofumi Hayashi (University of Aizu, Aizu Wakamatsu, Japan) presented the continuation of Nadia's work on modeling of visual impression from point of view of multimedia data mining, presented at MDM/KDD2001 [7]. The new work described a data warehouse for the mining process of multimedia information, where a unique characteristic of the data warehouse was its ability to store multiple hierarchical descriptions of the multimedia data. Such characteristic is necessary to allow mining multimedia data, not only at different levels of abstraction, but also according to multiple interpretation of the content. Proposed framework could be generalized to support the analysis of any type of complex data that relate to subjective cognitive processes, whose interpretation would be greatly variable.

In *User Concept Pattern Discovery Using Relevance Feedback and Multiple Instance Learning for Content-Based Image Retrieval* Xin Huang, Shu-Ching Chen (Florida International University, USA), Mei-Ling Shyu (University of Miami, USA) and Chengcui Zhang (Florida International University, USA) proposed a multimedia data mining framework that incorporated multiple instance learning into the user relevance feedback to discover users concept patterns, especially to find where the user's most interested region was allocated and how to map the local feature vector of that region to the high-level concept pattern of users. This underlying mapping could be progressively discovered through proposed feedback and learning procedure. The role of the user in the retrieval system was in guiding the mining process according to user's focus of attention.

4. CONCLUSION

Workshop participants agreed that in spite of the increasing efforts in multimedia data mining during the last three last years, extracting relevant and accurate knowledge from multimedia data sets remains a very difficult task. The contributions over the three years show that multimedia data mining is still an explorative domain of research and development. Figure 1 summarizes the MDM/KDD events so far in terms of submission/acceptance statistics and no more than three words 'summary'.

- MDM/KDD-99 – 'initial attempt'
  - **5** papers submitted, only **2** *related*
- MDM/KDD-2000 – 'second "initial attempt"'
  - **27** papers submitted, **14** accepted
- MDM/KDD-2001 – 'we are sailing …'
  - **20** papers submitted, **12** accepted
- MDM/KDD-2002 – 'it's a long way …'
  - **22** papers submitted, **12** accepted

Figure 1. MDM/KDD workshop series

The few real world applications showed that without accessing efficiently the content of media collections, it would not be realistic to consider knowledge extraction technologically. This year workshop identified that the research in multimedia data mining should be extended in the area of image and video information content mining.

Overall, MDM/KDD2002 was a successful workshop, demonstrating live interest in theoretical research and a growing area of application of multimedia data mining methods. The workshop remained a unique opportunity for exchanging ideas and extending the research community in this data mining area. All participants emphasized the importance to ensure that this event will continue on an annual basis.

The workshop URL:
http://www-staff.it.uts.edu.au/~simeon/mdm_kdd2002/

5. ACKNOWLEDGMENTS

We are thankful to the members of the KDD workshop committee for providing the opportunity to stage this annual event. Special thanks go to the workshop program committee for promptly providing the organizers with high quality paper reviews, and to Osmar Zaiane for his assistance in the organization of the workshop and local arrangements this year.

6. REFERENCES

[1] Simoff, S. and Zaïane, O. (2000). Report on MDM/KDD2000: The 1st International workshop on multimedia data mining, *SIGKDD Explorations*, **2** (2), pp. 103-105.
[http://www.acm.org/sigkdd/explorations/issue2-2.htm]

[2] Zaïane, O. and Simoff, S. (2002). Report on MDM/KDD2001: The 2nd International workshop on multimedia data mining, *SIGKDD Explorations*, **3** (2), pp. 65-67.
[http://www.acm.org/sigkdd/explorations/issue3-2.htm]

[3] Chen, S-C., Shyu, M-L., Zhang, C. and Strickrott, J. (2001). Multimedia data mining for traffic video sequences. In O. R. Zaïane and S. J. Simoff (eds), *Proceedings 2nd Int. Workshop on Multimedia Data Mining MDM/KDD2001*, pp 78-86.

[4] Khan, L. and McLeod, D. (2001). Effective retrieval of audio information from annotated text using ontologies. In S. J. Simoff and O. R. Zaïane (eds), *Proceedings 1st Int. Workshop on Multimedia Data Mining MDM/KDD2000*, pp 37-45.

[5] Antonie, M-L., Zaïane, O. R. and Coman, A. (2001). Application of data mining techniques for medical image classification. In O. R. Zaïane and S. J. Simoff (eds), *Proceedings 2nd Int. Workshop on Multimedia Data Mining MDM/KDD2001*, pp 94-101.

[6] Kulesh, V., Petrushin, V. A. and Sethi, I. K. (2001). The PERSEUS project: Creating personalized multimedia news portal. In O. R. Zaïane and S. J. Simoff (eds), *Proceedings 2nd Int. Workshop on Multimedia Data Mining MDM/KDD2001*, pp 31-37.

[7] Bianchi-Berthouze, N. (2001). An interactive environment for Kansei data mining. In O. R. Zaïane and S. J. Simoff (eds), *Proceedings 2nd Int. Workshop on Multimedia Data Mining MDM/KDD2001*, pp 58-67