The Eleventh Workshop on Multimedia Data Mining

Aaron Baughman
IBM
6710 Rockledge Drive
Bethesda, MD 20817, USA
baaron@us.ibm.com

Jia-Yu Pan
Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043, USA
jypan@google.com

Jiang (John) Gao
Nokia Inc.
200 S. Mathilda Ave
Sunnyvale, CA 94086, USA
jiang.gao@nokia.com

ABSTRACT

In this report we provide a summary of the eleventh Multimedia Data Mining Workshop (MDMKDD 2011) that was held in conjunction with the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011), August 21-24 in San Diego, CA.

Keywords

Multimedia data mining, multimedia information retrieval, knowledge discovery, semantic visual content indexing, video analysis, audio analysis, song extraction, social community evolution, user-generated content analysis, topic detection, mobile devices, multi-objectives optimization

1. INTRODUCTION

Big data and large scale analytics is a fundamental problem within computer science. The accelerating data avalanche is gaining unimpeded momentum that is increasing the volume and variety of information. Specifically, a growing and increasingly relevant component of today's corpora of information is the multimedia data. Within the knowledge discovery and data mining community and as evidenced by the success of the previous decade of this Workshop series, there is an increasing interest in new techniques and tools that can detect and discover patterns in multimedia data, including those that can lead to new knowledge. Multimedia information is ubiquitous, and is a digital capsule that is deliverable, artful, and empirical. An enabler for multimedia data and a new focus for the Workshop examined the content acquisition and delivery within mobile devices. Consumer-grade tablets, cell phones, and cameras provide affordable multimedia devices for sound, video, and images. For example, in tablet computing, forward facing cameras capture video and images that can be edited for individual use and published to social media sites. Business and events, such as the 2011 Wimbledon, provide the science of tennis, scores, and statistics to mobile devices so that patrons can enjoy streaming updates of simultaneous matches. Theme parks, such as Walt Disney, publish park information in the form of text, sound, video, and images that is accessible by mobile devices.

The entire MDM Workshop series has formed a trend of common work that lends itself towards a multimedia analytic pipeline for the accumulation of evidence within multimedia. Data fusion, machine learning, and multi-modal multimedia mining are core interests for the workshop. Within the medical community, combining CAT scans, x-rays, and MRI's provides evidential reasoning for diagnosis and prognosis. Topic and event discovery through photos, video, text, and sound produces emergent computing systems that tend towards individualized consumer experiences, novel business applications, and innovative solutions for an accessible world. Human computer interfaces and tools for multimedia data mining maintain importance within the workshop. Interfaces on websites to explore video clips, and facial recognition algorithms on social media sites are two facets of usable multimedia data mining. Additional important tools and functions include the classification of images, recognition of patterns in sound, object tracking, managing multimedia data on the WWW, feature extraction, and automatic annotation.

Multimedia is designed to stimulate the human senses beyond text or to capture a rich event. As such, data mining within media rich platforms and immersive environments such as online communities, blogs, social networks, and virtual worlds integrate the digital and physical. This Workshop explores the role multimedia data mining takes for enhanced technological experiences.

The Multimedia Data Mining (MDM) Workshop is one of the longest supported workshop series that accompany the ACM SIGKDD Conference (KDD). In 2011, the eleventh workshop of the series (MDMKDD2011) was held successfully in San Diego. The previous ten workshops on Multimedia Data Mining have been held in conjunction with KDD 2000 (Boston, MA), KDD 2001 (San Francisco, CA), KDD 2002 (Edmonton, Canada), KDD 2003 (Washington, DC), KDD 2004 (Seattle, WA), KDD 2005 (Chicago, IL), KDD 2006 (Philadelphia, PA), KDD 2007 (San Jose, CA), KDD 2008 (Las Vegas, NV), and KDD 2010 (Washington, DC), respectively. The MDMKD 2011 Workshop was held in a half-day format.

The major topics of the MDMKDD 2011 workshop include the following:

- Data mining on media rich platforms and immersive environments (on-line communities, blogs, social networks, virtual communities, virtual worlds).
- Mining large datasets of user generated content (e.g., YouTube, Flickr, etc.).
- Multimedia data mining across platforms, including web and mobile devices.
- Emerging technology of data mining for mobile applications.
- Predictive and prescriptive multimedia data modeling.
- Privacy preserving data mining.
- Mining multimedia time series.
- Multi-objective multimedia data mining.
- Anomaly and outlier detection in multimedia databases.

- Merging and integration of mining results from different sources (using ensembles, fusion techniques, etc.).
- Scalable data mining techniques for large-scale multimedia databases.
- Human-computer interfaces for multimedia data mining.
- Topic and event discovery in large multimedia repositories.

The submissions included papers by authors from countries around the world, including Japan, Thailand, Tunisia, United Kingdom, and United States. Each submission was reviewed by at least three program committee members. Five papers were selected for publication and presentation at the workshop. The workshop was honored to have Dr. John R. Smith deliver a keynote talk.

2. Keynote Speech

Dr. John R. Smith from IBM T.J. Watson Research Center provided a superb keynote, "Mining Images and Video for Meaning", for the workshop. The speaker's impressive background and experience provided a lively discussion. He is the Senior Manager of Intelligent Information Management that includes research within multimedia, database systems, computer vision, and cross-domain analytics. Currently he leads the development of IBM Multimedia Analysis and Retrieval System (IMARS) and was the Principal Investigator for IARPA Video Analysis and Content Extraction (VACE) project. Dr. Smith served as the chair of the MPEG Multimedia Description Schemes Group from 2001 to 2004, was the editor of the MPEG-7 Standard, Editor-in-Chief of IEEE Multimedia, and is an IEEE Fellow.

The keynote focused on large data sets, where multimedia data is big data. In particular, recognizing temporal and semantic activities (e.g. protest, violence, explosions, burning) is very complex. The IMARS system provides tens of thousands of classifiers to facilitate complex pattern recognition in broadcast video, web, social media, user content, and mobile computing. A few specific techniques of retrieving training samples include Random Subspace Bagging (RSBag) and Model-Shared Subspace Boosting (MSSBoost). The system provides manual cataloging by professionals, automated tagging by machines, and social tagging by crowd sourcing. A hierarchical faceted browsing interface enables users to view training data and annotations. During the keynote, a demo of IMARS was provided. The IMARS system is available at http://www.alphaworks.ibm.com/tech/imars.

An important ontology for multimedia data, Large Scale Concept Ontology for Multimedia (LSCOM), is fundamental for Dr. Smith's work. LSCOM defines a formal standard for the retrieval and annotation of video. The definition of the ontology began through a series of workshops held from April 2004 to September 2006. Within the IMARS system, labeled data for 449 visual concepts, out of the 1,000 LSCOM concepts, are available.

Another area that Dr. Smith discussed was the problem of turning visual information into actionable insights. Image and video analytics need to support Data-in-Motion and Data-at-Rest. To progress the field further, Dr. Smith's work includes the IBM Smart Vision Suite (SVS). The system categorizes activities, scenes, people, and objects. Specific activities that have been tracked by SVS include shoplifting and abandoning a bag, in diverse background scenes such as traffic, buildings, waterfront, and cityscapes. A brief discussion also revealed that his work included machine learning scale-up. The framework within SVS provides the ability to rapidly train new classifiers. The SVS is an outgrowth from years of research by the Exploratory Computing Vision group that started the smart surveillance research.

Towards the end of the keynote, a question and answer session provided additional insights into "Mining Images and Video for Meaning". The questions focused on feature extraction techniques and machine learning frameworks. A common trend of the audience was to learn and further understand how video analytics could be both scaled up and scaled out. Large-scale data processing techniques such as Unstructured Information Management Architecture (UIMA), InfoSphere Streams, and Hadoop were briefly discussed. Additional details of acquiring training data for thousands of classifiers, which includes crowd sourcing, manual and automated labeling, were provided. Finally, Dr. Smith showed that the MARVel (Multimedia Analysis and Retrieval) system contains links to functionalities such as classification, annotation, and video index searching. After the talk, an offline discussion of feature and score fusion from thousands of classifiers provided the capstone for the session. Methods such as average weighting, logistic regression, and decaying sum were talking points.

3. CONTRIBUTED PAPERS

The first accepted paper "A Fuzzy Ontology-Based Framework for Reasoning in Visual Video Content Analysis and Indexing" (*Elleuch, Zarka, Ammar, Alimi*) presents research on semantic visual content analysis and indexing. During the last decade, matching semantic concepts and visual data has attracted the attention of the research community in order to facilitate semantic indexing and concept-based retrieval of multimedia contents. Semantic concept detection is generally carried out by using supervised learning from manually annotated image samples, and the majority of research is almost exclusively focused on the development of independent concept detectors, where a main focus is on the extraction of low-level visual features in order to model the high level concepts. However, the same semantic concept may appear in various contexts and its appearance may be very different with respective to the different contexts. Therefore, indexing visual contents based on concept detectors is not optimal. The core contribution of this paper is the implementation of a fuzzy contextual ontology for representing semantic knowledge of concepts extracted by a contextual annotation framework. The extracted knowledge is represented as fuzzy rules, which are generated by an abduction engine. The paper shows how to apply the contextual ontology to enhance concept detection over the TRECVID 2010 corpus, and presents its effectiveness in terms of precision and recall on diverse concepts.

The second paper “Mining Movies to Extract Song Sequences” (*Doudpota, Guha*) proposes a method for extracting song segments from movies. The potential applications for the proposed method include music/song search and media collection and production. In particular, this paper studies the problem of extracting songs from Bollywood movies. The proposed method contains two major steps: an audio classifier which classifies segments that contain music from those that do not, and a song extractor that generates potential song sequences ("PSSs") and analyzes the vocal/non-vocal structure of each PSS to determine whether it is really a song sequence or not. The proposed audio classifier is a SVM classifier that uses features such as zero crossing rate, spectrum flux, and short-time energy to classify whether a frame is one with music or not. Consecutive music frames are grouped into PSSs, using heuristics on the song lengths observed from typical Bollywood songs. To estimate the likelihood of a PSS being a song, rather than a segment of conversation or action with background music, the proposed method first analyzes the vocal structure of the PSS using a vocal detector that identifies the vocal and non-vocal segments in the PSS. Similar to the audio classifier, the vocal detector is also a SVM classifier, which uses features as those used in the audio classifier, as well as additional ones such as the mel-frequency cepstral coefficients and the spectral centroid. With the information of the vocal structure, the likelihood of a PSS being a song or not is estimated using probabilistic timed automata, one automaton for songs and one for the non-songs. Each automaton is trained on a set of Bollywood songs and includes some heuristics derived from the structure of Bollywood songs via human analysis. Experiments on a set of 10 Bollywood movies show that the proposed method is able to extract the 74 songs in these movies with high precision and recall.

The third paper "Social Bookmark Data Mining Using Extended Graph Kernel" (*Niimi, Konishi*) studies the problem of extracting the transition patterns of communities from a sequence of graphs. In this study, the authors propose a procedure to convert a time-stamped transactional data set into a sequence of graphs, a method to detect communities from each single graph, and an algorithm to identify the evolution of a community over time. In particular, the formation of a graph sequence is done by first detecting "burst" events in the timeline of the transactions and then representing the events between two consecutive "burst" events as one graph in the graph sequence. For detecting the communities from each graph, the paper proposes to use the Clauset-Newman-Moore algorithm to find communities that maximizes the "modularity" metric. The evolution of each community is then found by grouping the detected communities at different time points into "community clusters". Furthermore, the paper defines various primitives (the "Community Changes") to describe the community transition over time. Using the primitives, a "community transition rule" is generated to describe the evolution of a community, where the generation is done by analyzing the relationship between the community instances in a community cluster at different time points. Experiments on the synthetic data sets and a data set from an online social bookmarking website show that the proposed methods are able to identify interesting rules from the data sets.

The fourth paper "What's Trending? Mining Topical Trends in UGC Systems with YouTube as a Case Study" (*Reed, Elvers, Srinivasan*) proposes a general procedure for mining emerging topics in user-generated content (UGC) systems, such as Twitter, Facebook, and YouTube. In this work, an emerging topic is represented as a set of correlated (co-occurred) terms that experience a sudden increase of interest during a specific time period. The proposed procedure is a generalization of an earlier work designed for detecting emerging topics in Twitter posts. The main challenge for this work is to make the procedure general for various UGC systems. To be general for a wide range of UGC systems, the paper (1) proposes a generic, query-based method for sampling/collecting user posts, and (2) develops measures that are adaptive and robust to the characteristics of the data from different UGC systems. In particular, the proposed improvements on the measures include (1) a term weighting scheme that is unbiased among different users, (2) a timestamp-weighted linear regression scheme on the "ranks" of terms for detecting emerging terms, and (3) an adaptive method to determine the threshold for extracting emerging terms. In addition, the implementation of the proposed method provides an interactive interface for users to visualize and explore the relationship among terms of an emerging topic. A case study on a data set of 2.2 million YouTube posts over a period of 15 days shows that the proposed procedure is able to detect major emerging topics and extract meaningful terms that describe those topics.

The fifth paper "Intelligent Call Routing: Optimizing Contact Center Throughput" (*Ali*) presents an architectural framework for applying real-time analytics to intelligent call centers. Industry research consistently shows that the central driver of customer satisfaction is the degree to which customer service representatives (CSR) and clients establish a rapport over the course of their interaction. The proposed framework applies machine-learning algorithms that recognize data patterns through repeated learning techniques and develops an effective matching strategy between customers and CSRs. Based on features representing the underlying customer and CSR demographics, the psychographics, and the historical performance information, the paper proposes predictive analytics that consider three scores: sales, satisfaction, and costs (call handle time), and these scores are combined in finding the best match between a customer and a CSR. The analytics engine has two main components: offline training and prediction. Experimental results based on three supervised learning algorithms: Neural Network, Random Forest, and Support Vector Machines, are benchmarked. The proposed predictor is capable of automatically updating and refining itself by cycling through all the machine learning algorithms and choosing the most accurate one based on the training data at a specific time. An average improvement of ten to fifteen percent on call outcomes is achieved based on the proposed system.

The list of the contributed papers is also available at the Workshop's website at https://sites.google.com/site/mdmkdd2011/home, along with supplementary materials such as presentation slides.

4. ACKNOWLEDGMENTS

We thank the KDD Workshop Co-chairs, Ian Davidson and Jennifer Dy, for their efforts in staging this annual event. We are also very grateful to the MDMKDD 2011 Workshop Program Committee members who promptly provided high quality paper reviews during a compressed schedule.

5. WORKSHOP CO-CHAIRS

Aaron Baughman, *IBM*

Jia-Yu Pan, *Google Inc.*

Jiang (John) Gao, *Nokia Research*

6. PROGRAM COMMITTEE

Cees Snoek, *University of Amsterdam*

Chaabane Djeraba, *LIFL - UMR CNRS*

Chong-Wah Ngo, *City University of Hong Kong*

Christian Eggenberger, *IBM*

Fatma Bouali, *University of Lille 2*

Florent Masseglia, *INRIA Sophia Antipolis*

Hanghang Tong, *IBM Research*

Hari Sundaram, *Arizona State University*

Henning Muller, *University of Applied Sciences Western Switzerland, Sierre*

Jessie Hsu, *Industrial Technology Research Institute (ITRI)*

Jingrui He, *IBM Research*

John Smith, *IBM Research*

Junfeng Pan, *Facebook*

K. Selçuk Candan, *Arizona State University*

Latifur Khan, *University of Texas at Dallas*

Lei Li, *Carnegie Mellon University*

Maria Luisa Sapino, *University of Torino*

Mark Zhang, *Binghamton University*

Max Lin, *Google Inc.*

Michael Perlitz, *IBM*

Qi Tian, *University of Texas at San Antonio*

Russell Vane III, *IBM*

Shann-Ching Chen, *St. Jude Children's Research Hospital*

Shu-Ching Chen, *Florida International University*

Stefan Stockt, *IBM South Africa*

Trista Chen, *Gracenote*

Valery Petrushin, *Neilson Company*

Vasileios Mezaris, *CERTH*

Vincent Tseng, *National Cheng Kung University*

William Grosky, *University of Michigan*

Xin Chen, *Navteq*

Yu-Gang Jiang, *Columbia University*