ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

THE ROLE OF METADATA IN WEB VIDEO MINING: ISSUES AND PERSPECTIVES

Siddu P. Algur, Prashant Bhat*, Suraj Jain
School of Mathematics and Computing Sciences, Department of Computer Science
Rani Channamma University, Belagavi-591156, Karnataka-India

ABSTRACT

Due to the complexity in rapid growth of audiovisual information over the web, it is becoming difficult to extract useful information from the web audiovisual data such as YouTube, Face Book, and Yahoo Screen etc. Web video mining is the process of extracting useful information from the web videos by applying data mining techniques. There are two approaches for web video mining- using traditional image processing/signal processing approach and metadata based approach. A number of techniques and algorithms are developed in image/signal processing approach to mine the video contents. But nowadays, mining of web videos without using image processing techniques is a challenging task. This paper represents a new approach for mining web videos using metadata as leading contribution for knowledge discovery.

Keywords: Web Video Mining, Metadata, Classification, Clustering, Outlier Detection, Object/Event Mining.

INTRODUCTIO

Video The World Wide Web (WWW) has huge amount of information and raw data, and continues to increase in size and complexity as well. It is very staggering task to search relevant information from huge amount of data. In recent years, web videos are widely deployed with the advance of digital devices/computer networks and hardware. Especially with the popularity of Web 2.0 and online video service, there are more and more online video sharing service providers, such as YouTube, Google Video, Yahoo Screen, etc. The amount of video clips also becomes huge accordingly. Due to the explosion of video clips over the web, it is important to provide an effective way to retrieve/extract useful information from the web videos (web video mining). To mine web videos, we classify mining methodology as- image processing based approach and metadata based approach.

In recent years many works have been implemented to discover knowledge from videos using traditional image processing techniques such as, video image retrieval and indexing technique, object recognition/detection in videos such as face recognition, vehicle detection etc, video object tracking from one frame to another and many more. A number of algorithms have been developed to mine the videos using image processing approach. However to discover knowledge from web videos, it has been observed that, in recent years a less number of research works have been implemented using metadata based approach which is a challenging task nowadays. The reason for this fact is, as metadata is different for different videos, sometimes there exist insufficient metadata for many web videos. This will lead to ineffective and poor result in knowledge discovery. In our previous work [1], we proposed metadata construction model for web videos to mine the web videos effectively and efficiently. The metadata for web videos can be of two types- implicit web metadata and explicit web metadata. Implicit web metadata (such as- file name of the video, description of the video, author name etc) is created by uploader of the video, whereas explicit metadata (such as information of- likes, dislikes, comments, tags etc) created by users of the web video. Further, metadata of a web video can be classified as- web metadata, descriptive metadata and technical metadata [1]. The web metadata includes all the information available in the website of the video, descriptive metadata includes sufficient description/summary of the video content and technical metadata includes basic/internal parameters of the video information such as- bit rate, resolution, aspect ratio, frame counts and duration etc. In our previous work [1] we showed that, if there is strong and sufficient metadata for web videos will lead to

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

better result in the knowledge discovery. Few works have been implemented using metadata to classify and clustering web videos. The section 2 represents some related works which are implemented using metadata to discover knowledge from web videos.

PRIOR WORK

The web video mining task using metadata based approach includes- classification of web videos, clustering of web videos, outlier detection among large number of web videos etc. And also few research works on web videos are implemented based on graph methodology. The authors Ullas Gargi, Wenjun Lu and others [2], proposed a multi stage community detection algorithm for large–scale YouTube video graphs. Local partitioning algorithms are implemented in a parallel fashion which is used to efficiently generate clusters which covers big portions of the graph. Pre-processing and post-processing steps are used to optimize coherence metrics and multiple graph–connectivity, such as coverage, conductance, and a new text coherence measure. The proposed algorithm performs clustering over tens of millions of YouTube videos, scalable to larger graphs, and produce very coherent clusters with good coverage. In this technique, clusters are labeled using entities extracted from the titles of constituent videos. These named clusters can be used to improve content discovery on YouTube effectively.

The authors John R. Zhang, Yang Song and Thomas Leung [3], proposed an approach which exploits YouTube video co-watch data to improve the performance of a video taxonomic classification method. A graph is built whereby edges are created based on video co-watch relationships and weakly-labeled videos are selected for classifier training through local graph clustering. Evaluation is performed by comparing against classifiers trained using manually labeled web documents and videos. The data collected through the proposed approach can be used to train competitive classifiers versus the state of the art, particularly in the absence of expensive manually-labeled data.

The authors Bingbing Ni, Yang Song and Ming Zhao [4], investigated the general event classification problem from uncontrolled YouTube videos. It is a challenging task due to the number of possible classes and large intra-class variations. The authors [4] worked on the problem- how to define proper event category labels and how to obtain training samples for these categories and also it was non-trivial to achieve satisfactory classification performance. To address these problems, a text mining pipeline is proposed to automatically discover a collection of video event categories. Part-of-Speech (POS) analysis technique applied to YouTube video titles and descriptions, and Word Net hierarchy was employed to refine the category selection, results in 29,163 video event categories. A POS-based query method is then applied to video titles, and 6,538,319 video samples results are obtained from YouTube to represent different categories. In this method, to improve classification performance, video content-based high-level features are complemented with scores from a set of classifiers. Extensive evaluations demonstrated the effectiveness of the proposed automatic event label mining and classification results.

The authors Subhabrata Mukherjee and Pushpak Bhattacharyya [5], proposed a weakly supervised system called YouCat for categorizing/classifying YouTube videos into different classes like Comedy, Horror, Romance, Sports and Technology. The proposed system takes a YouTube video URL as input and gives it a belongingness score for each type. The key features of this research work can be summarized as:

- Unlike other types of identification works, in which most cases are supervised, the YouCat system is mostly unsupervised, as it doesn't require labeled data for training.
- The system can easily incorporate new types/classes without requiring labeled data for the genres.
- YouCat extracts information from the video title, Meta description and user comments.
- The YouCat uses Wikipedia and Word Net for concept expansion.
- The prior algorithm with a time complexity of O (|W|) (where (|W|) is the number of words in the video descriptor) is efficient to be deployed in web for real-time video categorization.

Experimentations have been performed on real world YouTube videos where YouCat achieves an F-score of 80.9%, without using any labeled training set, compared to the supervised, multiclass SVM F-score of 84.36% for single genre prediction. YouCat performs better for multi-genre prediction with an F-Score of 90.48%. Weak supervision in the system arises out of the usage of manually constructed Word Net and genre description by a few root words. The authors Katja Filippova and Keith B. Hall [6],

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

presented a text-based approach to the task of assigning relevant categories to videos. In this approach the authors showed that a competitive classifier can be trained on high-scoring predictions made by a weakly supervised classifier learned from video features, and also the two provide complementary views on the data and that a simple product model which combines two sets of predictions outperforms each of them taken on their own. The prediction rate of 41% is quite high given the unprecedentedly large size of our category set (75) and the authors did not use any human-labeled data. Furthermore, for popular videos with at least ten comments, the authors achieved an even higher accuracy of 46%. It has been found that that, all of the text sources – description, title, tags and comments – are helpful for category predictions. A more significant result is that accurate predictions can be made from the users' comments, provided that there are enough metadata. The authors [6], analyzed a set of video-category pairs rated by humans and suggested three reasons for why a model which also looks at the viewers' comments outperforms the one which lacks this information source.

Tags are the new valuable source information in the web. The multi-media content has been heavily tagged by the owners and the viewers. Thus, Tags represent a social classification of the content and at the same time it also adds to its semantics. The author Ankur Satyendrakumar Sharma [7], intend to base our classification on the information carried in tags. This relieves us from the pain of explicitly performing the analysis and summarization of multimedia, which potentially could be a costly affair. However, the basic shortcoming in the usage of tags lies in the looseness of it's the representation. As tags are created by users (humans) they represent the human interpretation of the multimedia content, the personal bias of the human being comes into the context, which in turn acts as noise in this situation. Thus, it becomes complicated to identify tags that are relevant to the multimedia content. Tags not only act as a method of describing a multi-media object, but as means of communication to the other users. But due to the unspoken nature of the communication, the concept of focal point (from Thomas Schelling) comes into play. This project has been implemented to classify the videos which are located on YouTube on the basis of Tags and Focal Points.

The authors Zheshen Wang, Ming Zhao, Yang Song and others [8], proposed a technique to mine web videos manually creating a small labeled set and then extending it using additional sources such as related videos, searched videos, and text-based web pages. The data from such different sources has different properties and labeling quality, and thus fusing them in a coherent fashion is a practical challenging task. The authors [8] proposed a fusion framework in which each data source is first combined with the manually-labeled set independently. Then, using the hierarchical taxonomy of the categories, a Conditional Random Field (CRF) based fusion strategy has been designed. Based on the final fused classifier, category labels are predicted for the new web videos. Extensive experiments on about 80K videos from 29 most frequent categories in YouTube show the effectiveness of the proposed method for categorizing large-scale wild Web videos. It is observed from the reviews of existing web video mining techniques that, there is no ideal framework for knowledge discovery. Hence we propose an effective web video mining framework for knowledge discovery in the next section.

PROPOSED VIDEO OBJECT MINING FRAMEWORK

The proposed framework describes the mining process for web videos using metadata is shown in the Figure 1. A huge number of videos are available in different video websites such as YouTube, Google Videos and Yahoo Screen. These websites are open to the users and users can access any videos which are located on such social websites. This framework uses web video metadata as preliminary aspect for mining purpose. Hence, the entire mining result and efficiency depends on metadata of web videos. The metadata of web videos can be extracted using different available tools such as InfoExtractor (for web and descriptive metadata), MooO, MediaInfo (for technical metadata) etc. All three types of metadata, i.e. web metadata, descriptive metadata and technical metadata of large number of web videos are extracted and stored in a metadata database for mining purpose. The web video mining process includes various data mining strategies-preprocessing, data transformation, attribute selection, feature extraction/selection, filtering etc. In this stage, the metadata of web videos are reformed/ transformed to another format if it is necessary for mining process. Care must be taken while transforming metadata values from one form to another which will lead to a good result in knowledge discovery. The mining of web videos task includes classification, clustering, objects/events prediction and outlier videos, objects or events detection processes of large number of web videos. The

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

proposed framework consists of the following components, such as- Web Video Dataset, Metadata Extraction from Web Video Dataset, Metadata Repository of Web Videos, Web Video Mining Process, Mined Result Evaluation and Knowledge Discovery.

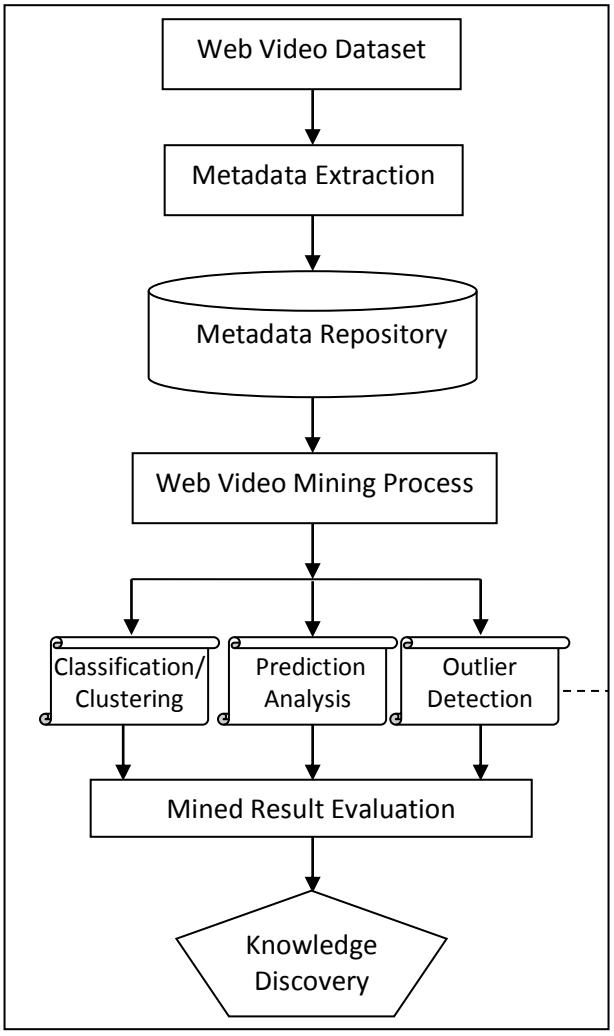

Figure 1: Proposed Framework for Web Video Mining

A) Web Video Dataset

With the exponential growth of social media and network in Web 2.0, the huge volume of videos being transmitted and searched on the Internet has increased massively. Users can capture videos by smart phones, video camcorders, or directly obtain videos from the web, and then distribute them again with some variations. The social web such as YouTube, Face Book, Yahoo Screen etc, provides different platforms for users to interact and exchange information. This has resulted in the rich sets of contextual information related with web videos. These contextual resources provide complementary information to the video content itself. In this dataset, in addition to the video itself, the metadata information- internal metadata, web metadata and descriptive metadata are also provided. These web video datasets can be encoded in the format flv, wmv, avi, mpg, mp4, ram etc with frame rate 15fps, 25fps, 29.97fps etc, and bit rate 529kbps, 819kbps etc and frame resolution 174x144, 320x240, 240x320 and so on.

B) Metadata Extraction Process

In this component, the quality and efficiency of the mining result is directly depends on richness of the available metadata of web videos. An effective and efficient method is needed to extract metadata from web videos. A traditional procedure to extract metadata from web pages is using XML platform. To extract metadata from web videos, different open source tools available such as Media Info, Video Inspector, Info Extractor, MooO etc. Using these tools all 3 types of metadata- internal metadata, web metadata and descriptive metadata can be extracted effectively.

C) Metadata Repository

The different types of metadata extracted from each web video and are stored in metadata repository with respect to appropriate Video ID of web videos. The proposed repository has 4 attributes- Video ID, Internal Metadata, Web Metadata and Descriptive Metadata. A sample schematic representation of metadata repository is shown in Table 1.

D) Web Video Mining Process

The next component in the proposed framework is web video mining, where web videos are mined using the metadata by applying any data mining techniques such as, classification, clustering, association rule mining, predictive analysis, outlier detection etc.

Web Video Classification and Clustering

All web videos are belongs to a particular category/class. These categories may be- Entertainment, Sports, News and Discussion, Education, Adventure etc. The entertainment category includes all web videos such as movie, serials, reality shows, fun videos etc. The sports category includes all web videos such as cricket, hockey, football, tennis etc. The web videos such as- live news, discussions, or interviews belongs to news

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

and discussion category. The lecturing videos, demonstration and laboratory practical oriented videos are comes under education category and so on. A common research problem in classification/categorization aspect is to classify web videos without watching it.

Video ID	Internal Metadata	Web Metadata	Descriptive Metadata
V0001	File Name: Lakshmi Baramma - 20th January 2015 - Full Episode-1QS5VjUk5ls.mp4 File Size (Bytes): 92,389,234 [Format] Format (Short Name): mov,mp4,m4a,3gp,3g2,mj2 Format (Long Name): QuickTime / MOV Duration: 00:20:18.72 Duration (Microseconds): 1,218,721,667 Bit Rate (bits/sec): 606,466 Number of Streams: 2 [Stream #0] Type: Video Real Base Frame Rate: 25/1 Average Frame Rate: 25/1 Duration: 00:00:00.72 Duration (Microseconds): 30,468 Bit Rate (bits/sec): 507,998 Number of Frames: 30,468	Author: ETV Kannada Published: 2015-01-20T16:22:30.000Z Category: Shows Keywords: Duration: 1219 Views: 3440 Ratings: 28 Avg. rating: 4.142857 Comments: 0 Favorited: 0 Thumbnail: http://i1.ytimg.com/vi/1QS5VjUk5ls/0.jpg	Shruthi was upset hearing the statements about Chandan. An unknown couple gives lift to Chandan and Laxmi when Chandan's car breaks down. Chandan and Laxmi get worried that they had taken lift from Shruthi's childhood friend. There was heated argument between Ramu and Ranjith, about Laxmi. For more stories please watch full episode.
V002	File Name: Comedy nights with Kapil, Full Episode -142, 10 January, 2015-e2osCTGoWfg.mp4 File Size (Bytes): 35,715,687 [Format] Format (Short Name): mov,mp4,m4a,3gp,3g2,mj2 Format (Long Name): QuickTime / MOV Duration: 00:47:56.88 Duration (Microseconds): 2,876,881,667 Bit Rate (bits/sec): 99,317 Number of Streams: 2 [Stream #0] Type: Video Codec (Short Name): h264 Codec (Long Name): H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 Codec Tag: 0x31637661 Codec Tag (String): avc1 Codec Time Base: 1/16 Time Base: 1/8 Real Base Frame Rate: 8/1 Average Frame Rate: 8/1 Duration: 00:00:00.88 Duration (Microseconds): 23,015 Bit Rate (bits/sec): 1,501	Video data from Comedy nights with Kapil, Full Episode -142, 10 January, 2015 (http://www.youtube.com/watch?v=e2osCTGoWfg&feature=youtube_gdata_player) Author: K19 Productions Published: 2014-12-28T11:57:58.000Z Category: People Keywords: Duration: 2878 Views: 16094 Ratings: 35 Avg. rating: 2.3714285 Comments: 1 Favorited: 0 Thumbnail: http://i1.ytimg.com/vi/e2osCTGoWfg/0.jpg	Video data from Comedy nights with Kapil, Full Episode -142, 10 January, 2015 (http://www.youtube.com/watch?v=e2osCTGoWfg&feature=youtube_gdata_player) URL: http://www.youtube.com/watch?v=e2osCTGoWfg&feature=youtube_gdata_player

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

	Number of Frames: 23,015		
V003 ⋮ Vn			

Table 1.: Metadata repository with available metadata of respective web video

The metadata such as, descriptions about the video contents, comments, keywords etc are useful to classification and clustering of web videos. The web metadata of a web video can be classified as two types- implicit metadata and explicit metadata. The implicit metadata such as title of the video, description of the video etc are given by the uploader of the web video. The explicit metadata such as comments, likes, dislikes, ratings etc are the independent of uploader the web video. A web video can be classified using implicit metadata or explicit metadata or both. A novel video clustering strategy for web video clustering is proposed by the authors Shouqun Liu Ming and Zhu Quan Zheng [9], which facilitates the users to effectively browsing web videos. The clustering method is evaluated by experiments with real web video clips. Furthermore, clustering with difference features and multimodal information may be investigated. Also another research aspect is to classification and clustering of objects and events of the web videos. Objects and events of similar types and dissimilar types are present in videos. Object classification based on video category, event classification, object clustering based on certain condition and event clustering are the challenging task during knowledge discovery.

Prediction Analysis

The video data contains various objects as well as different events associated in it. The objects and events may vary from one frame to another frame. It is a challenging task to predict the category of the video, to predict possible objects and events that a web video may contain. This prediction can either supervised or unsupervised. Various data mining prediction techniques such as linear regression can be used for knowledge discovery. A standard effective and efficient model is required to predict objects and events of the the web video.

Outlier Detection

Another complex task in web video mining is to detect and analysis of outliers. Outlier video detection among several web videos, outlier object detection and outlier event detection are the different mechanisms under this process. For example, song/dance events in a set of news/discussion videos may be considered as outlier event. Also the outlier detection mechanism is useful in the case of fraud detection in security aspects.

Mining Result Evaluation and Knowledge Discovery

At the end of web video mining strategy, the results of classification, clustering, prediction analysis and outlier analysis are collected and evaluated effectively and efficiently using statistical and mathematical models (if necessary) to discover the knowledge from web video datasets.

CONCLUSION AND FUTURE WORK

In this paper, we reviewed some existing techniques to mine web videos using metadata and observed that there is no ideal framework to discover knowledge from web videos. This paper describes a new effective framework for web video mining using metadata. Even though there exists many traditional image/signal processing technique to discover knowledge from videos, metadata based web video mining is complex and challenging task due to insufficient amount of available metadata of web videos. To overcome from this problem, we strongly recommend to give sufficient metadata for videos while uploading to websites

REFERENCES

[1] Siddu P. Algur, Prashant Bhat and Suraj Jain, "Metadata Construction Model for Web Videos: A Domain Specific Approach", IJECS, December 2014.
[2] Ullas Gargi, Wenjun Lu and others, "Large-Scale Community Detection on YouTube for Topic Discovery and Exploration", AAAI, 2011.
[3] John R. Zhang, Yang Song and Thomas Leung, "Improving Video Classification via YouTube Video Co-Watch Data", SBNMA,

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

2011.
[4] http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/37392.pdf
[5] http://www.aclweb.org/anthology/C12-1114.
[6] Katja Filippova and Keith B. Hall, "Improved Video Categorization from Text Metadata and User Comments", SIGIR, 2011.
[7] http://wwwusers.cs.umn.edu/~ankur/FinalReport_PR-1.pdf
[8] Zheshen Wang, Ming Zhao, Yang Song, Sanjiv Kumar, and Baoxin Li, "YouTubeCat: Learning to CategorizeWildWeb Videos", Google Research.
[9] Shouqun Liu Ming Zhu Quan Zheng, "Mining Similarities for Clustering Web Video Clips", International Conference on Computer Science and Software Engineering, 2008.

Author Biblography

Dr. Siddu P. Algur is working as Professor, Dept. of Computer Science, Rani Channamma University, Belagavi, Karnataka, India. He received B.E. degree in Electrical and Electronics from Mysore University, Karnataka, India, in 1986. He received his M.E. degree in from NIT, Allahabad, India, in 1991.He obtained Ph.D. degree from the Department of P.G. Studies and Research in Computer Science at Gulbarga University, Gulbarga.

He worked as Lecturer at KLE Society's College of Engineering and Technology and worked as Assistant Professor in the Department of Computer Science and Engineering at SDM College of Engineering and Technology, Dharwad. He was Professor, Dept. of Information Science and Engineering, BVBCET, Hubli, before holding the present position. He was also Director, School of Mathematics and Computing Sciences, RCU, Belagavi. His research interest includes Data Mining, Web Mining, Big Data and Information Retrieval from the web and Knowledge discovery techniques. He published more than 42 research papers in peer reviewed International Journals and chaired the sessions in many International conferences.

Mr. Prashant Bhat is pursuing Ph.D programme in Computer Science at Rani Channamma University Belagavi, Karnataka, India. He received B.Sc and M.Sc (Computer Science) degrees from Karnatak University, Dharwad, Karnataka, India, in 2010 and 2012 respectively. He has PG teaching experience and his research interest includes Data Mining, Web Mining, web multimedia mining, and Information Retrieval from the web and Knowledge discovery techniques and published research papers in various International Journals.

He has attended and participated in International and National Conferences and Workshops in his research field.

Mr. Suraj Jain is pursuing Ph.D programme in Computer Science at Rani Channamma University Belagavi, Karnataka, India. He received BCA and MCA degrees from Indira Gandhi National Open respectively. His research interest includes Data Mining, Web Mining, and web multimedia mining. He is also working as lecturer and PG coordinator, JSS College Dharwad. He has 13 years of professional experience in training and academic background.