



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Object Predictive Analysis of Video Web Mining Using J48 and Naive Bayesian Algorithms

P. Kavitha, S. Sasikala

M.Phil Research Scholar, Dept. of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi, India

Head of the Department, Dept. of Computer Science [UG], Sree Saraswathi Thyagaraja College, Pollachi, India

ABSTRACT: Web mining is a combination of data collected by historic data mining procedure and techniques with information gathered by over the web. Web mining is used to understand customer needs, calculate the effectiveness of a specific Web site, and help quantify the success of a marketing operation. Video Mining can be characterized as the unsupervised discovery of design in video-visual content. Furthermore, surveillance video often considers of events that are not known earlier, and is hence an obvious target for unsupervised discovery of patterns. For example, a video sequence captured by a camera trained at a full marketing place would be analyzed through the simple motion detection. In such a case, we do not need to know what is usual and unusual, let alone a finer classification. At present days, the Data Engineering is beginning to rise to discover knowledge from web video visual data such as YouTube videos, Yahoo Screen, Face Book videos etc. Various categories of web video are being shared on the social websites and are being used by the lakhs of users in the all over the world. The imported web videos will have the various kinds of meta data as attribute information available of the video data. The metadata attributes define the contents and features and characteristics of the web videos conceptually. So, accomplishing web video mining by extracting features of web videos in terms of metadata is a challenging task. This effective work attempt is made to classify and predict the metadata features of web videos in various category such as length of the web videos, number of comments of the web videos, number of ratings information and view counts of the web videos using data mining algorithms such as J48 and naive Bayesian algorithms as a part of web video mining. The results of J48 and naive Bayesian classification models are analyzed and compared as a step in the process of knowledge discovery from web videos.

KEYWORDS : Web videos, classification, web video classification , j48 classification Naive Bayesian classification.

I.INTRODUCTION

The Data Engineering is beginning to rise to discover knowledge from web video visual data social Medias such as- YouTube videos, Face Book videos etc. various types of web video are present shared on such social websites and are present used by the plenty of users all over the world. The uploaded web videos will have the different kinds of metadata as attribute information of the web video data[6]. The metadata attributes describe the contents and features and characteristics of the web videos conceptually. Hence, accomplishing web video mining by extracting features of web videos in terms of metadata is a challenging task. In this work, effective attempts are made to classify and predict the metadata features of web videos such as length of the web videos, number of comments of the web videos, ratings information and also view counts of the web videos using data mining algorithms such as Decision J48 and naive Bayesian algorithms as a part of web video mining[7]. The results of Decision tree J48 and naive Bayesian classification models are analyzed and compared as a step in the process of knowledge discovery from web videos.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 4, Issue 12, December 2016

II. RELEATED WORK

Amjad Mahmood reviewed "Semi-supervised evolutionary At once for Web video categorization"[1], worked on various categorization of web videos based on textual metadata. The scheduled method to various categorize of web videos are based on Semi supervised developmental At once framework. For (e.g), The web video categorization using minimizes cost textual appearance such as title, definition, tag etc., built in relations and extrinsic web support. In sequence to implement the scheduled technique, the common Vector Space Model extend to the Semantic technique for VSM. Lastly, analysis on real world social Web data has been perform to over all the framework.

Aggarwal Reviewed "Mining YouTube Metadata for Detecting privacy Invading Harassment Videos"[2], The common web videos shared will have different kind of web video metadata such as category, comment information, rating information, and view counts etc . Different kind of video web metadata is assign to the characteristics of the web videos. The web video classification prediction and analysis of web videos in conclusion of such different kind of video web metadata is a difficult and challenging work and then many kind of classification models and algorithms and data mining and machine learning tools are developed in few years.

Anil Kale Reiewed "An Automated Video Classification and Annotate Using Embedded Audio for Content Based Retrieval"[3], this work an computerized web video classification technique. This technique presents a model that issues the automation of web video classification and video observation. The videos are classified and observed on the keywords.

Bin Cui Ce Zhange Reviewed "Content Enriched Classifier for Web Video Classification"[4], This work implement a different video classification system which is able to apply both content and text characteristics for video classification while ignore the price estimation of extracting content characteristics at classification time. This develop approach apply the content features extracted from instruction data to improve and enrich the text based correct substance, yielding content enriched acceptable substance. Content improved and enriched semantic substance utilize both content and the text features for classifying web videos without using extracting their content and characteristics. The experimental results show that the proposed technique extreamly performs well on the state and art video classification methods.

Chunneng Huang, Reviewed "Text-based video content classification for web video sharing sites"[5], This work a textual base methodology for video content classification and categorization of web videos sharing on the Web sites. The various kinds of user performed data and for example, tag, titles, definition , and comments, and used as intermediately for web videos, and three types of text features were extracted. The two characteristics based classification techniques (C4.5 Naïve Bayes, and Support Vector Machine) were used to classify web site videos.

Chueh-Wei Chang, Reviewed "Abnormal Spatial Event Detection and Video Content Searching in a Multi-Camera Surveillance System" [9]. This works schema for structural communication development, unexpected action disclosure and web video content analysis with respect to observed control applications. This proposed system can computerize find the abnormal events from guiding places, and select the model key frame from the web video clips as an basis, and store the color feature of the thick objects into the control database[8]. A chart model has been defined to correlative the chase the objects between the multiple views. This work was helpful to the control system to audit the schedule of objects whether go into difficult path or not. To attain the contented base web video object analysing, a substance based access has been selected as a comparison calculate between the color provide of the collect object and target applicant in the control database.

III. PROPOSED SYSTEM

The proposed work describes the mining process for web videos using metadata .A large number of videos is available in different video websites such as YouTube, and Google Videos. The web metadata of online videos are extracted using the info extracting tool. This metadata includes uploaded information, category, comments, ratings, length of the video, descriptions about content of the video information available etc. Here propose a problem and effective methodology to classify and predict the web video metadata features such as length, view counts, number of comments, status and rating information by applying data mining techniques. The novel work to classification model built using J48 and navie Bayesian classification methods. The classification and prediction results of each considered metadata features are analyzed and the efficiency of the proposed method has been demonstrated.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 4, Issue 12, December 2016

A. ADVANTAGE

- Website metadata included over all the information of the website video.
- Descriptive metadata includes sufficient description of the video.
- This raw metadata has to be preprocessed for the refinement.
- Main advantage comparable file format communication and to identify detect the unimportant metadata

B. STEPS INVOLVED IN PROPOSED SYSTEM

A .Web video metadata extraction

B .Metadata clarification

C . Classification models

A. Web Video Metadata Extraction

The various categories of web videos available like as Entertainment sports, news are randomly selected and given to the information extract tool, to extract different types of web metadata such as length, rating information, status, category, etc. The extracted raw metadata will be in the form of text and these metadata are then stored in such file formats.

B .Metadata Clarification

The input component is basic metadata extracted from the web videos. This basic metadata has to be preprocessed for the clarification such as file format communication and to detect the unimportant metadata The extracted basic metadata are converted to ARFF or CSV format from the text format for effective classification. Some web videos might have less metadata information, where as some web videos might have more metadata information. Through observations, it is found that, all web videos contains minimum metadata information such as ratings, ratings and average ratings, count of views and number of comments, author information and URL.

C .Classification Models

This component has two sub components:

1. J48 classification model
2. The naive Bayesian probabilistic classification model.

V.ALGORITHMS IMPLEMENTATIONS

A. J48 CLASSIFICATION MODEL

J48 classification model has three types available such as:

A. Attribute selection measures

B. J48 algorithm

C. Classification rules.

A. Attribute Selection Measures:

Attribute selection measures implement a splitting criteria for each attribute describing the given tuples. The web video metadata for attribute selection measures are: Information needed to identify the category of an element of a metadata tuple, Information gain of each attributes and Splitting criteria.

B. J48 classification algorithm:

J48 is version of C4.5 classification algorithm. The algorithm consist all the possible tests that can split the metadata data set and selects a test that gives the best information. For each one metadata features of the web videos such as Views Length, and Ratings, Comments etc, binary tests involving every distinct values of the attribute are considered.

C. Classification Rules

A part or segment of the Tree structure of J48 classification model for the dataset chosen is represented.

Ratings Very High(1.0)High(14.0/0.4)Low(135.0/10.0) Medium(10.0/4.0)

Veryhigh=High=Low=Medium

The above tree can be converted to classification rules by traversing the path from root node to each leaf node in the tree.

J48 PRUNED

Ratings=very heigh:very heigh



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 4, Issue 12, December 2016

Ratings=High:High(15.0/4.0)
Ratings=Low:Low(156.010.0)
Ratings=Medium:Medium(10.04.0)
Number of leaves:4
Size of tree:5

This assumption is called class conditional independence.

- $P(c|x)$ is the posterior expectation of class (target) given predictor (attribute).
- P naïve is the prior probability of class.
- $P(x|c)$ is the likelihood which is the of expectation predictor given class.
- $P(x)$ is the prior probability of predictor.

Hybrid classification using J48 and Naive Bayesian

The following issues are faced by most j48 algorithms:

- To choose splitting attributes
- Order of splitting attributes
- Number of splits to be taken
- Balance of data structure and pruning
- The stopping criteria

The following is the working of j48 Algorithm:

Step1: importing video data set

Step2: selecting the separate category of whole video

Step3: Assign the label of having a data object

Step4: split the objects like length, command and status finally shows the result on over all objects.

Step5: End

B. HYBRID CLASSIFICATION USING J48 AND NAIVE BAYESIAN

Naive Bayesian Algorithm steps:

Step1: Convert the data set into a frequency table.

Step 2: Create Likelihood table by finding the probabilities like probability cost.

Step 3: Use Naive Bayesian comparison to evaluate the posterior probability for each class. The class with the highest posterior probability is the outcome of predictive.

Step 4: The posterior probability can be calculated by first, constructing a object table for each attribute against the target.

Step5: Then, transforming the frequency tables to likelihood tables and finally use the Naive Bayesian equation to calculate the posterior probability for each class.

Step6: The class with the highest posterior probability is the outcome of predictive.

Step 7: Take little time classify the video files.

Step 8: Classify is completed after display the result and End.

VI. SIMULATION AND RESULT

The following Table 1 explains the structure of web video metadata dataset.

TABLE 1

Category	Length	Views	Rate	Ratings	Comments
People Blogs	300	1900	4	5	6
Comedy	439	557	4	4	4
Entertainment	250	1154	4.4	33	12
Sports	298	265	1	1	2

The following Table 2 explains the Attribute selection details and the overall attribute details and class lables shows in this table.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 4, Issue 12, December 2016

S.no	Attributes	Description	Class labels
1	Length	Length of video	Low Medium High Very High
2	Rate	Rating given by users	Low Medium High Very High
3	View	View counts of web videos	Low Medium High Very High
4	Rating	Rating of the web videos	Low Medium High Very High
5	Comments	No of comments given by users	Low Medium High Very High

TABLE 2

The Naive Bayesian Probabilistic Classification Model labels are following:

NODE

1.High

2.Low

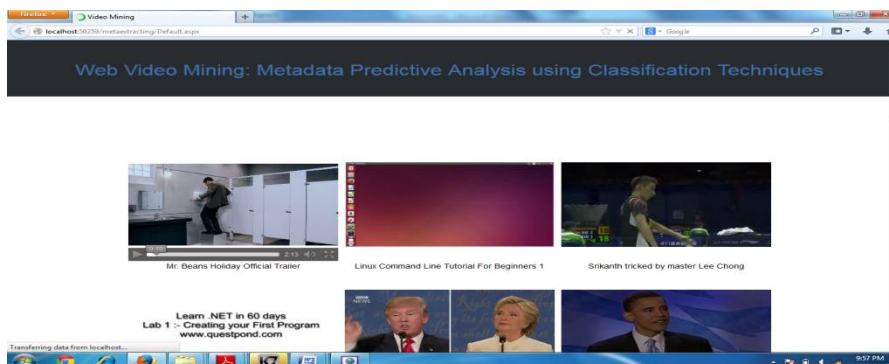
3.Medium

4.High

5.VeryHigh

The empirical results shown the proposed technique extremely performs well on the state video classification methods. Automatic different kind of videos in a Web-scale unconstrained collection such as YouTube is a challenging task.

A. CATEGORY OF VIDEOS



B. METADATA REFINEMENT

The input to this fundamental is raw metadata concentrate from the web videos. This basic metadata has to be preprocessed for the refinement such as file format communication and to identify the unimportant video metadata.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

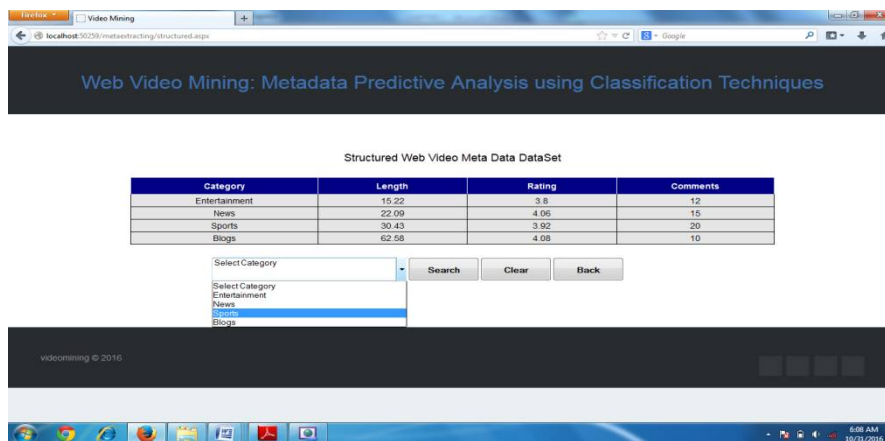
Website: www.ijirce.com

Vol. 4, Issue 12, December 2016



C. WEB VIDEO METADATA EXTRACTION

In this basic, the quality and efficiency of the mining result is depends on richness of the applicable metadata of web videos. Effective and efficient method is desired to extract metadata from web videos. A common procedure to extract metadata from web pages. To concentrate metadata from web videos, different open source tools available such as Media Info and Video Inspector, Info Extractor. Using these tools are all three types of metadata internal metadata, web metadata and descriptive metadata can be concentrate effectively.



Category	Length	Rating	Comments
Entertainment	15.22	3.8	12
News	22.09	4.06	15
Sports	30.43	3.92	20
Blogs	62.58	4.08	10

A. J48 Algorithm Classification



Filename	Category	Length	Rating	Comments
Mis dhoni interview with mayank tanger dhoni childhood PART- 1	Sports	7.34	4.2	4
Sunil Chhetri All Goals & Dribbling Skills HD	Sports	10.36	4	3
Portland Trail Blazers vs LA Lakers	Sports	9.28	3.9	6
Star Sports Pro Kabaddi Season 2 All Stars: Rahul Chauhan	Sports	3.10	3.8	5
Srikanth tricked by master Lee Chong	Sports	0.35	3.7	2



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 4, Issue 12, December 2016

B. HybridNavie Bayesian Classification

The screenshot shows a web browser window with a URL starting with 'localhost'. The page title is 'Chaudhari View'. The main content area is titled 'J48 Classification' and contains a table with the following data:

Filename	Category	Length	Rating	Comments
Ms dhoni interview with mayank...	Sports	7.34	4.2	4
langer dhoni's childhood PART-1				
Sunil Chhetri All Goals & Dribbling Skills HD	Sports	10.35	4	3
Portland Trail Blazers vs LA Lakers	Sports	9.28	3.9	6
Star Sports Pro Kabaddi Season 2 All-Stars: Rahul Chaudhari	Sports	3.10	3.8	5
Srinanth tripped by master Lee Chong	Sports	0.35	3.7	2

Below this table is a section titled 'Result J48 Classification & Navie Bayesian' containing a smaller table:

Sno	Attributes	Description	Class Labels
1	Length	30.43	Low
2	Rate	3.92	High
3	Comments	30	Medium

V.CONCLUSION AND FUTURE ENHANCEMENT

CONCLUSION

The dissertation proposed as “objective predictive analysis of video web mining using j48 algorithm and Naive Bayesian” This is effective methods to classify predict the video web objects. This work the uploaded video will have different kind of video metadata as attributes information of the video data. The web video metadata define the attributes contents and feature of the web video conceptually. This work experiment are made to classify and predict the metadata characteristics of the web videos such as length of the videos, comments, status rating information and view count of the web videos using data mining algorithm j48 and Navie Bayesian classification models analyzed.

FUTURE ENHANCEMENT

In, this work, we classified web videos based on their metadata attributes/features such as- length, view counts, rate, ratings, and number of comments as a part of knowledge discovery from web videos. The web video metadata are extracted from standard website and stored in a database for classification.

The J48 and naive Bayesian (NB) classification algorithms are chosen to classify and predict the class labels of different attributes chosen. In future work can be extended by using the recent Algorithms used in web video mining and also whole category of video analyzed and the objects can be predictable and classified using the data mining tools.

REFERENCES

1. Aggarwal N, Agrawal, S. and Sureka, A., “Mining YouTube Metadata for Detecting privacy Invading Harassment and Misdemeanor Videos”, IEEE Twelfth Annual International Conference on , vol., no., pp.84,93, 23-24 July 2014.
2. Amjad Mahmood, Tianrui Li, Yan Yang, Hongjun Wang and Mehtab Afzal, “Semi-supervised evolutionary ensembles for Web video categorization”, Elsevier- Knowledge-Based Systems 76 (2015) 53–66.
3. Anil Kale and D.G. Wakde, “An Automated Video Classification and Annotation Using Embedded Audio for Content Based Retrieval”, Journal of Industrial and Intelligent Information, December 2013.
4. .Bin Cui Ce Zhang and Gao Cong, “Content Enriched Classifier for Web Video Classification”, 2010 ACM 978-1-60558-896-4/10/07.
5. .Bin Zhao et al., “Online Detection of Unusual Events in Videos via Dynamic Sparse Coding”, 2011.
6. Chunneng Huang, Tianjun Fu and Hsinchun Chen, “Text-based video content classification for online video-sharing sites”, Journal of the American Society for Information Science and Technology, May 2010.
7. Chirag Shah, Charles File, “Info extractor – A Tool for Social Media Data Mining”, JITP 2011.
8. .Chueh-Wei Chang, Ti-Hua Yang and Yu-Yu Tsao, “ Abnormal Spatial Event Detection and Video Content Searching in a Multi-Camera Surveillance System”, MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009.
9. Cewu Lu et al., “Abnormal Event video Detection at 150 FPS in MATLAB”, The Chinese University of Hong Kong.