IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-6 E-ISSN: 2347-2693

Abnormal Web Video Prediction Using RT and J48 Classification Techniques

Siddu P. Algur[1], Prashant Bhat[2*]

[1]Department of Computer Science, Rani Channamma University, Belagavi-591156, Karnataka, India
[2] Department of Computer Science, Rani Channamma University, Belagavi-591156, Karnataka, India

www.ijcseonline.org

Received: May/16/2016 Revised: May/30/2016 Accepted: Jun/18/2016 Published: Jun/30/ 2016

Abstract— Now a days, the 'Data Science Engineering' becoming emerging trend to discover knowledge from web videos such as- YouTube videos, Yahoo Screen, Face Book videos etc. Petabytes of web video are being shared on social websites and are being used by the trillions of users all over the world. Recently, discovering outliers among large scale web videos have attracted attention of many web multimedia mining researchers. There are plenty of outliers abnormal video exists in different category of web videos. The task of classifying and prediction of web video as- normal and abnormal have gained vital research aspect in the area of Web Mining Research. Hence, we propose novel techniques to predict outliers from the web video dataset based on their metadata objects using data mining algorithms such as Random Tree (RT) and J48 Tree algorithms. The results of Decision Tree and J48 Tree classification models are analyzed and compared as a strategy in the process of knowledge discovery from web videos**.**

Keywords—Outliers, Decision Tree, J48 Tree, Web Video Outliers, Prediction, Knowledge Discovery

I. INTRODUCTION

YouTube is recognized as one of the most successful user-generated video sharing sites nowadays. YouTube has over a billion users — almost one-third of all people on the Internet — and everyday people watch hundreds of millions of hours on YouTube and generate billions of views [1] [2]. In order to facilitate users to find interesting videos from a large number of videos, YouTube provides different features/metadata objects such as – view counts, rate, ratings, number of comments, favorites, key words, information regarding likes and dislikes etc.

The objective of this study is to classify and predict outlier (abnormal) videos among large scale web videos using their metadata objects. To thrive in the proposed objective of the work, large scale web video metadata objects are extracted from the standard YouTube dataset website [3]. This metadata objects includes various attributes such as- 'Category', 'Length', 'Views', 'Rate', 'Avg Ratings' 'Number of Comments' and 'Outlier (Nominal)'of each web videos.

The schematic structure of the dataset is represented in Table 1.

The main contributions of our work are as follows:

- Web Video Metadata Object dataset extraction and effective preprocessing for the experiment.
- The analysis and knowledge discovery process from the results of 10 cross validation classification and predictions of outliers by the built Random Tree (RT) and J48 Tree classification models.

Many outlier models/algorithms and data mining machine learning tools are developed in recent years. Using different data mining algorithms and machine learning tools such as R programming and WEKA, it is possible to classify and predict outliers from the web videos based on their features/metadata objects.

The rest of the paper is organized as follows: The section 2 represents related works on the clustering of web videos, section 3 represents proposed abnormal web video prediction methodology, section 4 represents performance evaluation analysis of outlier prediction models and comparison of efficiency of outlier models, and finally section 5 represents conclusion and future enhancements.

Table 1: Schematic Structure of Web Video Metadata Object Dataset.

No.	1: Category Nominal	2: Length Numeric	3: Views Numeric	4: Rate Numeric	5: Ratings Numeric	6: Comments Numeric	7: Outlier Nominal
30163	Music	111.0	4953.0	4.2	5.0	6.0	no
30164	Music	637.0	21804.0	4.75	24.0	15.0	no
30165	Music	316.0	31029.0	4.43	37.0	9.0	no
30166	Music	322.0	28641.0	4.89	35.0	17.0	no
30167	Music	293.0	7346.0	4.88	8.0	5.0	no
30168	Entertainm...	297.0	41697.0	4.83	53.0	36.0	yes
30169	Music	254.0	2117.0	4.5	6.0	4.0	no
30170	Music	220.0	13900.0	4.27	11.0	10.0	no
30171	Music	272.0	4753.0	4.5	4.0	2.0	no
30172	Music	291.0	8677.0	4.85	13.0	13.0	no
30173	Music	244.0	34339.0	4.77	35.0	22.0	yes
30174	Comedy	227.0	3888.0	3.94	18.0	17.0	no
30175	Music	249.0	37299.0	4.47	76.0	35.0	yes
30176	Music	200.0	1459.0	5.0	5.0	3.0	no
30177	Music	274.0	8755.0	4.86	7.0	4.0	no
30178	Entertainm...	267.0	45385.0	4.97	67.0	52.0	yes
30179	Music	228.0	108805.0	4.79	141.0	81.0	no
30180	Music	85.0	6542.0	5.0	4.0	6.0	no
30181	Music	261.0	96760.0	4.91	153.0	170.0	no
30182	Entertainm...	343.0	2776.0	4.78	9.0	6.0	no
30183	Music	228.0	13715.0	5.0	17.0	11.0	no
30184	Music	206.0	1826.0	5.0	1.0	0.0	no

*Corresponding Author: ***Prashant Bhat***, *prashantrcu@gmail.com*
Department of Computer Science, Rani Channamma University, Belagavi-591156, Karnataka, India

II. PRIOR WORKS

This section represents some related previous works which are implemented to find abnormal web videos/ abnormal web video events using metadata objects.

The authors Chueh-Wei Chang, et al. [2], proposed a framework for spatial relationship construction, abnormal event detection and video content searching with respect to visual surveillance applications. The proposed system [2] can automatically detect the abnormal events from monitoring areas, and select the representative key frame(s) from the video clips as an index, then store the color features of the suspect objects into the surveillance database. A graph model has been defined to coordinate the tracking of objects between multiple views. This was helpful to the surveillance system to check the route of objects whether go into a critical path or not. A variety of spatio-temporal query functions can be provided by using this spatial graph model. To achieve the content-based video object searching, a kernel- based approach has been employed as a similarity appraise between the color distribution of the deduce object and target candidates in the surveillance database.

In the work of [4], the authors Fan Jiang, Ying Wu, and Aggelos K. Katsaggelos have proposed a multi-sample-based similarity measure, where HMM training and distance measuring were based on multiple samples. Such multiple training data were acquired by a novel dynamic hierarchical clustering (DHC) method. By iteratively reclassifying and retraining the data groups at different clustering levels, the initial training and clustering errors due to over fitting was consecutively corrected in soon after steps. The proposed experimental results on real surveillance video showed an enhancement of the presented method over a baseline method that uses single sample- based similarity measure and spectral clustering approach.

The authors [5] Tushar Sandhan et al. have proposed the unsupervised learning algorithm - Proximity (Prx) clustering for abnormality detection in the video sequence. The proposed Prx clustering method tried to select only the dominant class sample points from the dataset. For each data sample, the algorithm assigned the degree of belongingness to the dominant cluster. The proposed motion features such as - circulation, motion homogeneity, motion orientation and stationary attempt has been made to extract vital information which was essential for abnormality discovery. After performing Prx clustering, each sample belongs to dominant cluster with the membership value. When Prx clustering is being performed in the space constructed from the proposed motion features, it helps to improve the abnormality detection performance. Experimental results for clustering performance evaluation on artificial dataset show that the Prx clustering outperforms the other clustering methods, for clustering the single dominant class from the dataset. Abnormality detection experiments show the comparable performance with other methods; in addition it has an advantage of incremental learning that it learns about the new normal events in an unsupervised manner.

In the work of [6], the authors Thi-Lan Le and Thanh-Hai Tran proposed a technique which, we can apply only HOG-SVM detector on extended regions detected by background subtraction. This method takes advantages of the background subtraction method (fast computation) and the HOG-SVM detector (reliable detection). Moreover, the authors [4] have done multiple objects tracking based on HOG descriptor. The HOG descriptor, computed in the detection phase, was used in the phase of observation and track association. This descriptor was more robust than standard grayscale (color) histogram based descriptor. As a conclusion, the paper [6] discussed a hybrid method for abnormal event detection which allows to remove several false detection cases.

The authors Yang Cong et al [7] proposed the Sparse Reconstruction Cost (SRC) over the normal dictionary to measure the normalness of the testing sample. By introducing the prior weight of each basis during sparse reconstruction, the proposed SRC is more robust compared to other outlier detection criteria. To condense the over-completed normal bases into a compact dictionary, a novel dictionary selection method with group sparsity constraint is designed, which can be solved by standard convex optimization. Observing that the group sparsity also implies a low rank structure, the authors [7] reformulated the problem using matrix decomposition, which can handle large scale training samples by reducing the memory requirement at each iteration from O (K2) to O (k) where k is the number of samples. Then the proposed technique of [7] used the column wise coordinate descent to solve the matrix decomposition represented formulation, which empirically leads to a similar solution to the group sparsity formulation.

Based on inherent redundancy of video structures, Cewu Lu et al [8] proposed an efficient sparse combination learning framework. It was accomplished decent performance in the uncovering phase without compromising result value. The short running time was guaranteed because the new method effectively turns the original complicated problem to one in which only a few costless small-scale least square optimization steps are involved. The proposed method of [8] arrived at high detection rates on benchmark datasets at a rate of 140∼150 frames per second.

The authors Yang Cong et al [9] have made experimental attempt to identify abnormal events via a sparse reconstruction over the normal bases. Given an over-complete normal basis set (e.g., an image sequence or a collection of local spatio-temporal patches), the authors [9] commenced the sparse reconstruction cost (SRC) over the normal dictionary to measure the normalness of the testing model. To condense the size of the dictionary, a novel dictionary selection method is designed with sparsity consistency constraint. By introducing the prior weight of each basis during sparse reconstruction, the proposed SRC is more robust compared to other outlier detection criteria. The method of [9] provides a unified solution to detect both local abnormal events (LAE) and global abnormal events (GAE). The experiment of [8] further extended to maintain online abnormal event recognition by updating the dictionary incrementally. Also researches on three benchmark datasets and the comparison to the state-of-the-art methods authenticated the compensation of proposed algorithm.

The experimental results of Bin Zhao et al [10] revealed a fully unsupervised dynamic sparse coding method for discovering abnormal events in videos based on online sparse re-constructability of query signals from an atomically learned event dictionary, which generates sparse coding bases. Using an intuition that normal events in a video are more likely to be re-constructible from an event dictionary, whereas abnormal events are not. The proposed algorithm [10] employed a principled convex optimization formulation that permits both a sparse reconstruction code, and a web dictionary to be together inferred and updated. The techniques were fully unsupervised, making no prior hypothesis of what unusual events may look like and the settings of the cameras. The fact that the bases dictionary is updated in a web fashion as the algorithm examined new data, avoids any matters with concept drift. Investigational results on hours of real world surveillance video and numerous YouTube videos showed that, the proposed algorithm might reliably locate the abnormal events in the video frames/sequences, outperforming the present state-of-the-art methods.

The authors Du Tran et al [11] depicted a method to discover abnormal motion in videos. The interior of the approach was to detect portion of video that corresponds to sudden changes of motion variations of a set of defined points of curiosity. The proposed optical flow technique tracked those points of curiosity. There were plenty variations in the optical flow patterns in a mob scene when there are cases those showing abnormalities. The geometric clustering algorithm, k-means, clusters the obtained optical flow information to get the distance between two successive frames. In general, relatively high distance indicates abnormal motion. To demonstrate the interest of the approach, the authors [11] presented the results based on the discovery of abnormal motions in video, which consists of both normal and abnormal motions.

The authors Du Tran et al [12] proposed to discover spatiotemporal paths for video event detection. This new formulation was accurately found and locate video events in cluttered and crowded scenes, and was vigorous to camera motions. It was also well handled the scale, shape, and intra-class disparities of the event. Compared to event detection using spatiotemporal sliding windows, the spatiotemporal paths correspond to the event trajectories in the video space, thus can better handle events composed by moving objects. The authors [12] proved that, the proposed search algorithm can achieve the global optimal solution with the lowest complexity. Experiments were made on realistic video data sets with various event detection tasks, such as anomaly event discovery, walking person identification, and running recognition. The proposed method [12] was compatible with different types of video features or object detectors and robust to false and missed local recognitions. It was significantly developed the overall detection and localization accurateness over the state-of-the-art techniques.

III. PROPOSED METHODOLOGY

In this section we present novel methodology for classification of web videos. The web video metadata objects are extracted from standard web video database website [13], preprocessed and stored in a database [14]. Then the refined data will be given to proposed classification/prediction models as input and resultant classified/predicted web videos will be extracted and performance evaluation metrics of Random Tree and J48 classification models will be analyzed for knowledge discovery. The system model of the proposed technique is represented in Fig. 1, and it consists of the following components:

A) Web Video Meta-Objects Collection Process
B) Data Refinement Process
C) Classification/Prediction Process
D) Result Analysis and KDD Process

A) Web Video Metadata Objects Collection Process

The different kind of web video metadata objects are extracted using InfoExtractor tool [15] and web video metadata objects are then preprocessed stored in a disk [14] with CSV or ARFF file format for experimental purpose. The summary of the dataset is represented in Table 2.

Table 2: Summary of the dataset

Summary	***Length***	***Views***	***Avg Rate***	***Ratings***	***Comments***
Min.	0.0	1	0.0	0.0	0.0
1st Qu.	83.0	579	3.67	2.0	1
Median	194.0	2220	4.69	6.0	4
Mean	223.5	11342	3.87	20.93	18.23
3rd Qu.	296.0	8176	5.0	17.0	13
Max.	5412.0	3281256	5.0	4629	5772

B) Data Refinement Process

The raw web video metadata objects are preprocessed to get stable result in the experiment. Missing values are replaced by median value of each attribute. The noise and redundancy in the database are removed for the better accuracy in the results using R programming. A typical structure of refined web video metadata object dataset is presented in Table 1. In the Table 1, the attribute 'Category' is nominal and contains 16 different classes (ex- 'Comedy', 'Music', 'UNA', 'Sports' etc) of web videos [13]. The remaining attributes are numeric and represents meta-objects of each web videos.

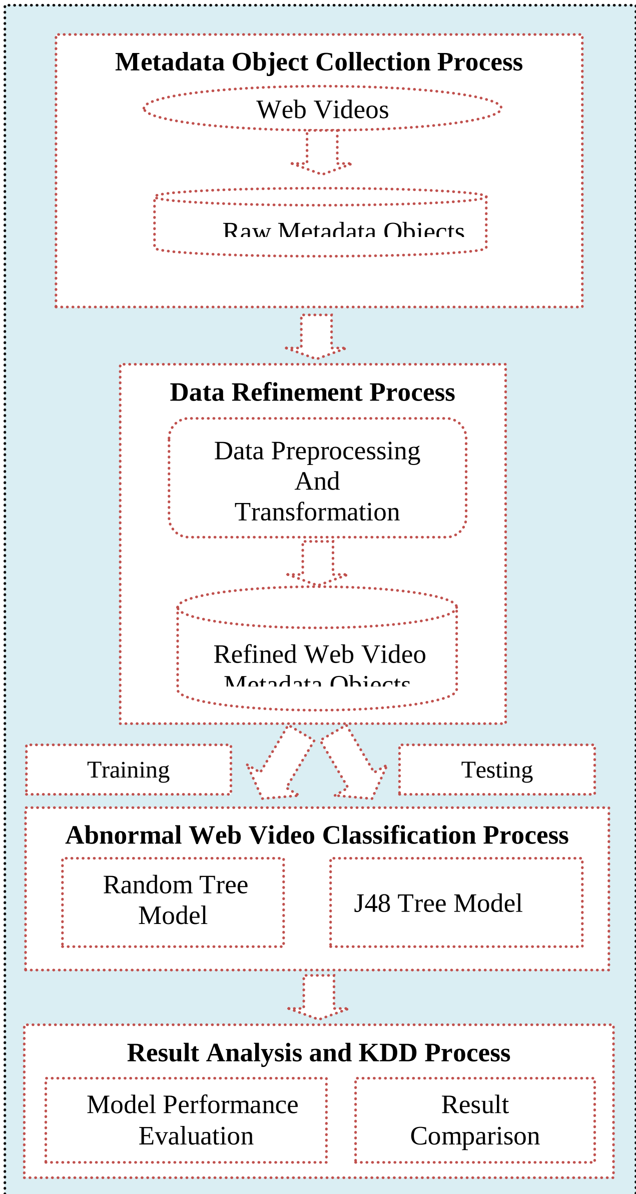

Fig. 1: System model of the proposed methodology

C) Abnormal Web Video Classification Process

The proposed work uses Random Tree and J48 Tree models to classify/predict abnormal videos in the web video metadata object dataset. The procedure to classify/predict abnormal videos using metadata objects based on Random Tree and J48 Tree methods are discussed as follows:

- ***Random Tree (RT) Classification Model***

The procedure to construct the Random Tree Classification Model has been discussed in our previous work [13]. The rules part of constructed Random Tree outlier model is represented in the Table 3.

Table 3: RandomTree Classification Model Rules Part

```
Views < 30778
|  Ratings < 62.5
|  |  Length < 931 : no (42418.91/0)
|  |  Length >= 931
|  |  |  Category = People & Blogs
|  |  |  |  Length < 1751.5 : yes (1/0)
|  |  |  |  Length >= 1751.5 : no (11.01/0)
|  |  |  Category = Comedy : no (0/0)
|  |  |  Category = Entertainment
|  |  |  |  Length < 1641 : yes (24.02/0.02)
|  |  |  |  Length >= 1641 : no (29.02/0)
|  |  |  Category = Howto & Style : no (2/0)
|  |  |  Category = Music
|  |  |  |  Length < 1416 : yes (3/0)
|  |  |  |  Length >= 1416 : no (5/0)
|  |  |  Category = Sports : no (2/0)
|  |  |  Category = News & Politics
|  |  |  |  Ratings < 26.5
|  |  |  |  |  Length < 1573 : yes (3/0)
|  |  |  |  |  Length >= 1573 : no (11.01/0)
|  |  |  |  Ratings >= 26.5 : yes (5/0)
|  |  |  Category = Film & Animation
|  |  |  |  Length < 1550.5 : yes (2/0)
|  |  |  |  Length >= 1550.5 : no (1/0)
|  |  |  Category = Nonprofits & Activism : no (0/0)
|  |  |  Category = UNA : no (0/0)
|  |  |  Category = Travel & Events : no (0/0)
|  |  |  Category = Autos & Vehicles
|  |  |  |  Length < 1650.5 : yes (1/0)
|  |  |  |  Length >= 1650.5 : no (1/0)
|  |  |  Category = Education : no (7.01/0)
|  |  |  Category = Pets & Animals : no (0/0)
|  |  |  Category = Gaming : no (1/0)
|  |  |  Category = Science & Technology
|  |  |  |  Length < 1488 : yes (2/0)
|  |  |  |  Length >= 1488 : no (37.03/0)
|  Ratings >= 62.5
|  |  Ratings < 107.5 : yes (831.65/0.65)
|  |  Ratings >= 107.5
|  |  |  Length < 1080.5 : no (405.31/0)
|  |  |  Length >= 1080.5
|  |  |  |  Category = People & Blogs : no (1/0)
|  |  |  |  Category = Comedy : no (0/0)
|  |  |  |  Category = Entertainment : no (0/0)
|  |  |  |  Category = Howto & Style : no (0/0)
|  |  |  |  Category = Music : yes (1/0)
|  |  |  |  Category = Sports : no (0/0)
|  |  |  |  Category = News & Politics : yes (1/0)
|  |  |  |  Category = Film & Animation : yes (1/0)
|  |  |  |  Category = Nonprofits & Activism : no (0/0)
|  |  |  |  Category = UNA : no (0/0)
|  |  |  |  Category = Travel & Events : no (0/0)
.........................................................................
.........................................................................
.........................................................................
```

The attribute 'Views' has highest information gain and labeled at the root node. The trees were formed randomly in accordance with the attribute selection methods.

- ***J48 Tree Classification Model***

The procedure to construct the J48 Tree Classification Model has been discussed in our previous work [13]. The pruned rules part of constructed J48 Tree outlier model is represented in the Table 4.

Table 4: J48 pruned tree Outlier Model

```
Views <= 30776
|  Ratings <= 62
|  |  Length <= 921: no (42418.91)
|  |  Length > 921
|  |  |  Length <= 1563: yes (41.03/0.03)
|  |  |  Length > 1563: no (107.08)
|  Ratings > 62
|  |  Ratings <= 107: yes (831.65/0.65)
|  |  Ratings > 107: no (409.32/3.0)
Views > 30776
|  Views <= 54013
|  |  Views <= 31126
|  |  |  Ratings <= 61: no (26.02)
|  |  |  Ratings > 61
|  |  |  |  Ratings <= 108: yes (14.01/0.01)
|  |  |  |  Ratings > 108: no (6.0)
|  |  Views > 31126: yes (1822.41/1.41)
|  Views > 54013
|  |  Ratings <= 107
|  |  |  Ratings <= 62: no (701.54)
|  |  |  Ratings > 62: yes (382.3/0.3)
|  |  Ratings > 107
|  |  |  Length <= 733: no (911.71)
|  |  |  Length > 733
|  |  |  |  Rate <= 4.77: no (2.0)
|  |  |  |  Rate > 4.77: yes (23.02/0.02)
```

IV. Results and Discussions

The meta-objects of different categories of 47697 web videos of same age (1400 days) are extracted, preprocessed and stored in database [13-15] for abnormal video prediction/classification using RT and J48 classification models. The 10 cross validation experimental attempts with RT and J48 classification models have been employed on the web video meta-object dataset to classify/predict outliers/abnormal videos using. The Fig. 2 and Fig.4 represents resultant classification tree generated by the RT and J48 classification model respectively. The results obtained from the RT and J48 classification models are shown in Table 5.

The Table 5 reveals, out of 47697 web videos, 47628 web videos were correctly classified, and 69 web videos were incorrectly classified by the RT classification model, whereas, 47681 web videos were correctly classified and, 16 web videos were wrongly classified by the J48 classification model.

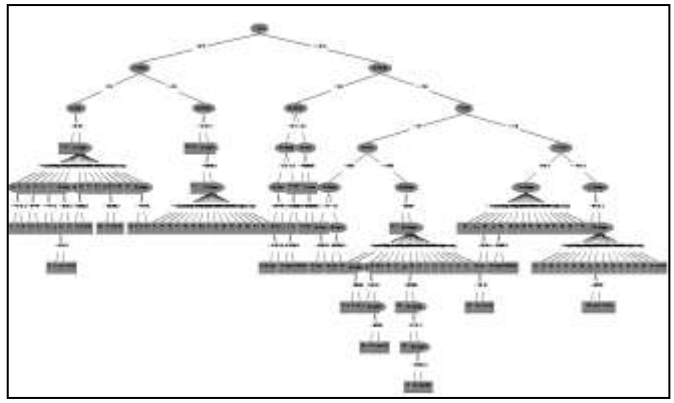

Fig.2: Tree structure of Random Tree Classification Model

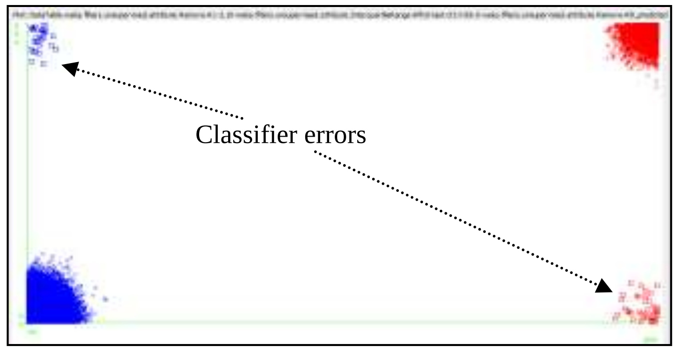

Fig.3: Classifier errors of RT Classification Model

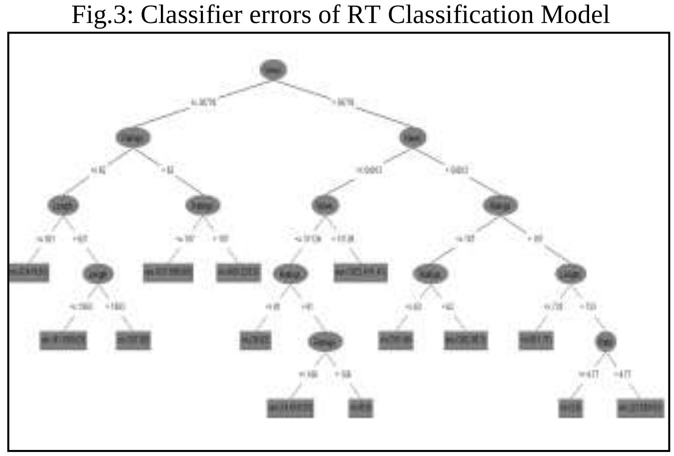

Fig.4: Tree structure of J48 Tree Classification Model

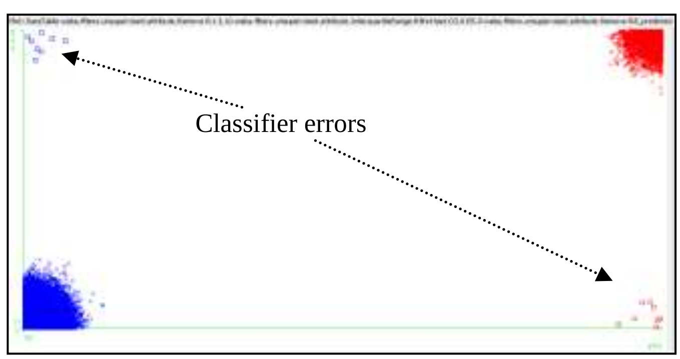

Fig.5: Classifier errors of J48 Tree Classification Model

Table 5: Classification Results of RT and J48 Classification Models

Classification Model	***Total Instances: 47697***		***TP***	***FP***	***Precision***	***Recall***	***F-Score***	***ROC Area***
	Correctly Classified	***Incorrectly Classified***						
Random Tree	47628	69	0.99	0.01	0.99	0.99	0.993	0.993
J48 Tree	47681	16	1	0.002	1	1	1	0.999

Table 6: Confusion Matrix for RT and J48 Classification Models

Confusion Matrix

RT Classification Model			J48 Classification Model		
a	**b**	**←Classified as**	**a**	**b**	**←Classified as**
44552	30	a= Normal	44574	8	a= Normal
39	3076	b= Abnormal	8	3107	b=Abnormal

Also, the TP rate, FP rate, precision, recall, F-Score , ROC area were more accurately simulated by the J48 classification model as compared to RT classification model. However, the results of performance evaluation metrics as described in the Table 5 are nearly same.

Classification errors of RT and J48 classification models are graphically shown in Fig. 3 and Fig.5 respectively. It has been observed from the experiment and Fig. 3 and Fig.5; the classification error rate is reduced in the simulation of J48 classification/prediction result as compared to RT classification/prediction results.

In order to study the classification errors and percentage of correctly classification by the RT and J48 classification models, we draw confusion matrix and is represented in Table 6. Out of 44582 normal web videos, 44552 web videos were correctly labeled as ‘Normal’ by the RT classification model and 30 web videos were mislabeled as ‘Abnormal’ by the RT classification model. And also, out of total 3115 actual abnormal videos, 3076 web videos were correctly labeled as ‘Abnormal’ by the RT classification model and 39 web videos were mislabeled as ‘Normal’.

In the experimental simulation of J48 classification model, Out of 44582 normal web videos, 44574 web videos were correctly labeled as ‘Normal’ by the J48 classification model and 8 web videos were mislabeled as ‘Abnormal’ by the J48 classification model. And also, out of total 3115 actual abnormal videos, 3107 web videos were correctly labeled as ‘Abnormal’ by the RT classification model and 8 web videos were mislabeled as ‘Normal’. Observe that, in the both case of classification/prediction of class labels ‘Normal’ and ‘Abnormal’, the J48 tree classification model produced negligible amount of error rate.

V. CONCLUSION AND FUTURE WORK

In this work, novel attempts are made to classify web videos as ‘Normal’ and ‘Abnormal’ based on their meta-object values such as- length, view counts, rate, ratings, and number of comments as a part of knowledge discovery from large scale web videos. The web video meta-objects were extracted from standard website and stored in a database for classification. The RT and J48 classification algorithms were chosen to classify/predict the class labels. The classification results of RT and J48 classification models were compared and found J48 classification model with predictive analysis is more efficient for classify web videos using meta-object values.

REFERENCES

[1] Siddu P. Algur, Prashant Bhat, "Abnormal Web Video Detection Using Density Based LOF Method", International Journal of Computer Sciences and Engineering, Volume-04, Issue-04, Page No (6-14), Apr - 2016, E-ISSN: 2347-2693

[2] 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, Yokohama, JAPAN.

[3] Dataset for "Statistics and Social Network of YouTube Videos", http://netsg.cs.sfu.ca/youtubedata/.

[4] Fan Jiang, Ying Wu, Aggelos K. Katsaggelos, "Abnormal Event Detection from Surveillance Video by Dynamic Hierarchical Clustering", Northwestern University, USA.

[5] Tushar Sandhan et al., "Unsupervised learning approach for abnormal event detection in surveillance video by revealing infrequent patterns", IEEE 28th International Conference on Image and Vision Computing, 2013- New Zealand

[6] Thi-Lan Le and Thanh-Hai Tran, "Real-Time Abnormal Events Detection Combining Motion Templates and Object Localization", Advances in Intelligent Systems and Computing 341, DOI 10.1007/978-3-319-14633-1_2, Springer International Publishing-2015, Switzerland.

[7] Yang Cong et al., "Abnormal Event Detection in Crowded Scenes using Sparse Representation", Pattern Recognition, January 30, 2013

[8] Cewu Lu et al., "Abnormal Event Detection at 150 FPS in MATLAB", The Chinese University of Hong Kong.

[9] Yang Cong et al., "Sparse Reconstruction Cost for Abnormal Event Detection".

[10] Bin Zhao et al., "Online Detection of Unusual Events in Videos via Dynamic Sparse Coding", 2011.

[11] Mahmoudi Sidi Ahmed et al., "Detection of Abnormal Motions in Video", Chania ICMI-MIAUCE'08 workshop, Crete, Greece, 2008.

[12] Du Tran et al., "Video Event Detection: From Subvolume Localization to Spatiotemporal Path Search", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, 2014.

[13] Siddu P. Algur, Prashant Bhat, "Metadata Based Classification and Analysis of Large Scale Web Videos", International Journal of Emerging Trends and Technologies in Computer Science, May-June 2015.

[14] Siddu P. Algur, Prashant Bhat, Suraj Jain, "The Role of Metadata in Web Video Mining: Issues and Perspectives", International Journal of Engineering Sciences & Research Technology, February-2015.

[15] Chirag Shah, Charles File, "InfoExtractor – A Tool for Social Media Data Mining", JITP 2011.

Authors

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. He was also Director, PG Programmes, 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 45 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. His research interest includes Data Mining, Web Mining, web multimedia mining and Information Retrieval from the web and Knowledge discovery techniques, and published 20 research papers in International Journals. Also he has attended and participated in International and National Conferences and Workshops in his research field.