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

Abnormal Web Video Detection Using Density Based LOF Method

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

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

www.ijcseonline.org

Received: Mar/19/2016 Revised: Mar/26/2016 Accepted: Apr/12/2016 Published: Apr/30/2016

***Abstract*—** Recently, discovering outliers among large scale web videos have attracted attention of many web mining researchers. There are number of outlier/abnormal videos exists in each category of web videos such as- ‘Entertainment’, ‘Sports’, ‘News and Politics’, etc. The task of identifying and manipulate (to remove from the web or to share with others in the web, or to watch/download from the web etc) such outlier web videos have gained significant important research aspect in the area of Web Mining Research. In this work, we propose novel methods to detect outliers from the web videos based on their metadata objects. Large scale web video metadata objects such as- length, view counts, numbers of comments, rating information are considered for outliers' detection process. The outlier detection method –Local Outlier Factor (LOF) with different nearest neighbor values (with K=3, K=5 and K=7) are used to find abnormal/outlier web videos of same age. The resultant outliers are analyzed and compared as a step in the process of knowledge discovery.

***Keywords*—**Outliers, Lcal Outlier Factors, Inter-Quartile Range, Web Video Outliers, Clustering, YouTube

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]. 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 detect outlier videos among large scale web videos using their metadata objects. To succeed in the proposed objective of the work, large scale web video metadata objects are extracted from the standard YouTube dataset website [5]. This metadata objects includes various attributes such as- ‘Category’, ‘View Counts’, ‘Rate’, ‘Number of Comments’, ‘Avg Ratings’ and ‘Length’ 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 resultant unsupervised outliers formed by the built three different (K=3, K=5 and K=7) LOF outlier models.

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

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 detect 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 outlier detection of web videos, section 3 represents proposed outlier/abnormal web video detection methodology, section 4 represents performance evaluation analysis of outlier 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
1	People & Blogs	217.0	1157.0	3.6	5.0	3.0
2	Comedy	426.0	667.0	4.0	4.0	4.0
3	Entertainment	237.0	1063.0	0.0	0.0	1.0
4	Comedy	294.0	274.0	1.0	1.0	2.0
5	Comedy	109.0	48.0	5.0	2.0	1.0
6	People & Blogs	263.0	62.0	5.0	4.0	5.0
7	Howto & Style	34.0	3437.0	4.7	20.0	1.0
8	Comedy	50.0	385.0	2.71	7.0	9.0
9	Comedy	251.0	91.0	5.0	1.0	1.0
10	Comedy	145.0	209.0	5.0	1.0	0.0
11	Comedy	150.0	21.0	0.0	0.0	0.0
12	Comedy	19.0	189.0	5.0	1.0	0.0
13	Entertainment	14.0	91.0	0.0	0.0	0.0
14	Comedy	284.0	1146.0	4.5	2.0	1.0
15	Music	210.0	8483.0	4.85	13.0	7.0
16	Entertainment	134.0	3524.0	4.95	57.0	39.0
17	Music	181.0	498.0	5.0	2.0	0.0
18	Music	460.0	2514.0	4.55	11.0	5.0
19	Entertainment	196.0	1606.0	4.94	32.0	20.0
20	Music	136.0	11838.0	4.85	27.0	32.0
21	Music	31.0	6586.0	4.73	15.0	8.0
22	Entertainment	397.0	1893.0	4.95	22.0	6.0
23	Entertainment	50.0	6966.0	4.0	6.0	7.0
24	Entertainment	261.0	726.0	5.0	1.0	13.0

II. RELATED 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. [1], proposed a framework for spatial relationship construction, abnormal event detection and video content searching with respect to visual surveillance applications. The proposed system [1] 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 [2], 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 [3] 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 [4], 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 [4] discussed a hybrid method for abnormal event detection which allows to remove several false detection cases.

The authors Yang Cong et al [5] 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 [5] 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 [5] 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 [6] 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 [6] arrived at high detection rates on benchmark datasets at a rate of 140~150 frames per second.

The authors Yang Cong et al [7] 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 [7] 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 [7] provides a unified solution to detect both local abnormal events (LAE) and global abnormal events (GAE). The experiment of [7] 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 [8] 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 [8] 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 [9] 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 [9] 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 [10] 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 [10] 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 [10] 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 of the proposed abnormal web video detection approach. The web video metadata objects are extracted from standard web video database website [11], preprocessed and stored in a database [13]. Then the refined data will be given to proposed density based LOF outlier detection models as input and resultant abnormal web videos will be extracted and 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 Metadata Objects Collection Process
B) Data Refinement Process
C) Abnormal Web Video Detection 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 [14] and web video metadata objects are then preprocessed stored in a disk [13] 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. 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 [12]. The remaining attributes are numeric and represents meta-objects of each web videos.

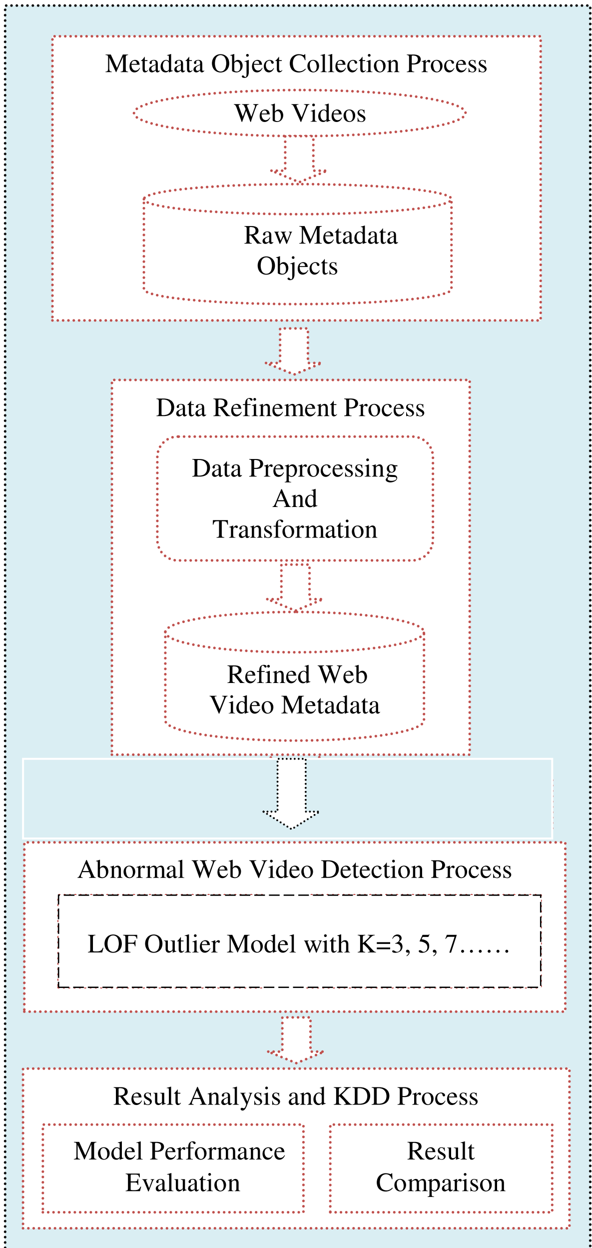

Fig. 1: System model of the proposed methodology

C) Abnormal Web Video Detection Process

The proposed work uses LOF method to identify abnormal videos in the web video metadata object dataset. The procedure to detect abnormal videos using metadata objects based on LOF method is discussed as follows:

- *Local Outlier Factor (LOF) Method*

The Local Outlier Factor (LOF) is a technique to detect anomalous/abnormal data points by computing the local deviation of a given data point with respect to its neighbors. The local outlier factor is based on a concept of a local density, where locality is given by K nearest neighbors, whose distance is used to estimate the density. By comparing the local density of a web video metadata object to the local densities of its neighbor's web videos, it is possible to identify regions of similar density, and points that have a substantially lower density than their neighbors web videos. Such web videos are considered to be outlier web videos. The local density of a web video metadata object will be estimated by the typical distance at which a point can be "reached" from its neighbors.

In LOF method, the local density of a web video metadata object value is compared with that of its neighbors. If the previous is significantly lower than the latter (with an LOF value greater than one), the point is in a sparser region than its neighbors, which indicates that it is an outlier. The function lofactor() calculates local outlier factors using the LOF algorithm, and it is available in packages *DMwR* and *dprep* of R.

Algorithmic steps to detect top N abnormal web videos

Algorithm: *To detect top N abnormal web videos*
Input: *Web video metadata object dataset*
Output: *Top N abnormal web videos*
Algorithm:

1. Import DMwR library of R

2. Import web video metadata object dataset D

// Remove categorical feature 'Category' of web video metadata object dataset D.

3. D1 ← D – 'Category'

4. Outlier_Score ← lofactor(D1, K)// K=3, 5, 7,……

5. Density_Factor ← density (Outlier_Score)

// Pick top K web videos as outliers/abnormal web videos

6. Outliers ← order (Outlier_Score, decreasing=T) [1: N]

// Display outliers as abnormal web videos

7. Print (Outliers)

The proposed method is able to extract top N outliers from the dataset by using the value of K-nearest neighbor objects. This paper demonstrates the results for K=5 and K=3 with top 10% outliers/abnormal web videos in section 4.

Also to be more accurate in the detection of abnormal/outlier web videos, the experimental attempts will be extended to identify and extract the abnormal/outlier web videos which are intersection of resultant outlier/abnormal web videos of LOF with K=5, LOF with K=3 and LOF with K=3 (i.e., [results of LOF with K=3] ∩ [results of LOF with K=5] ∩ [results of LOF with K=7).

D) Result Analysis and KDD Process

In Data Mining strategy, the performance evaluation and result analysis are significant steps to discover the knowledge. In this component of the proposed model, we are discovering abnormal web videos, using LOF outlier models. At this stage, the resultant outliers will be analyzed in depth to find abnormal web videos using their meta-objects.

IV. RESULTS AND DISCUSSIONS

The meta-objects of different categories of 47660 web videos of same age (1400 days) are extracted, preprocessed and stored in database [12] [13] for abnormal video detection. Subsequently, the LOF algorithm has been applied on the web video meta-object dataset to detect outliers/abnormal videos using R programming.

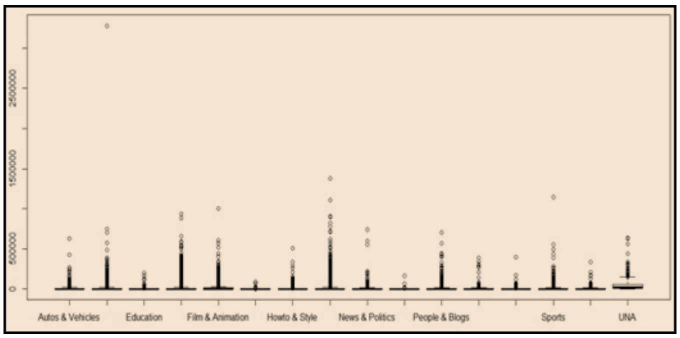

Fig. 2. (a) Box plot representation of ‘Views’

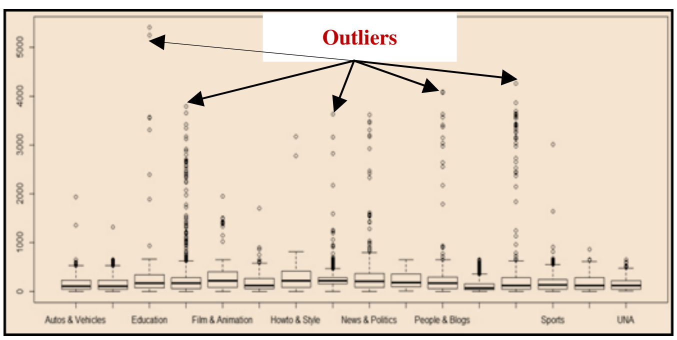

Fig. 2.(b) Box plot representation of ‘Length’

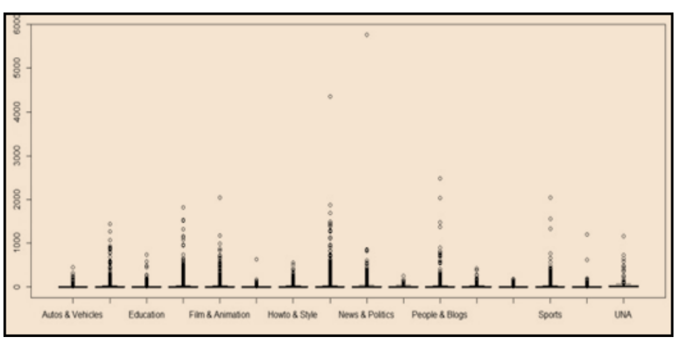

Fig.2.(c) Box plot representation of ‘Comments’

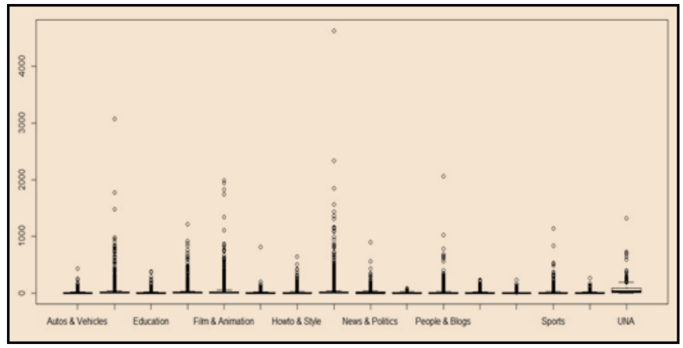

Fig.2.(d) Box plot representation of ‘Ratings’

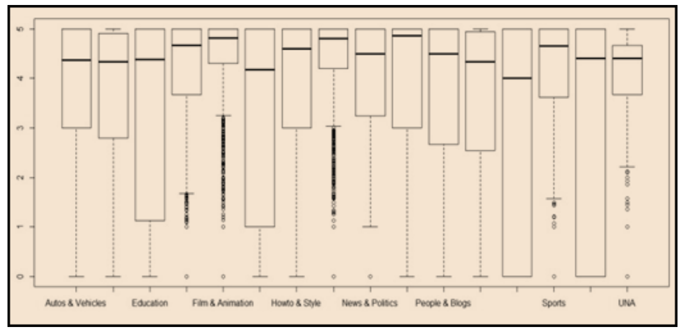

Fig. 2.(e) Box plot representation of ‘Rate’

The Fig 2(a) to Fig. 2(e) shows ‘Category’-wise box plot representation of web video metadata attributes- ‘Views’, ‘Length’, ‘Comments’, ‘Rate’ and ‘Ratings’ respectively. It is observed from the box plot graphical representation, the web video dataset contains several outliers in each of the numeric metadata attribute with respect to different categories of web videos. Sample outliers are addressed in Fig. 2(b). In view of this an attempt is made to find the outlier density factor and is represented in Fig. 3. The Fig.3 reveals that, the dataset contains outliers (abnormal videos) and the density of the outlier factor is positively skewed with bandwidth 0.016.

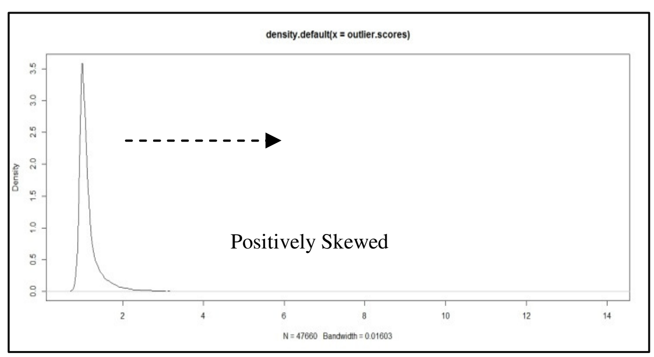

Fig.3: Outlier Density Factor

The Table 3 shows structure of results obtained by the LOF method which contains Video ID/Row ID, length of the web videos (duration), view count of the web videos, average ratings and number of ratings given by the web users, and finally number of comments. Since all videos of same age with 1400 days, we obtained some interesting patterns from the web video meta-object datasets.

There are 2 cases: Web videos those who have become viral in a specified period, and web videos, those who have completely neglected by the web users in a specific period. In this experiment, these two cases are addressed as abnormal web videos. For example, as the results described in the Table 3, the web video ID 25951 has become viral in the specific period, as it has very high view counts, high average rate, ratings and comments, meaning that it contains awesome/very interesting contents and attracting web users in rapid way. But the web video with ID 23179 completely neglected by the web users in a specific period because it might not contain interesting contents. Hence, such kinds of web videos considered to be abnormal videos. Also these abnormal web videos are represented in the form of scatter plot matrix in Fig.5 in which the red marks indicates abnormal web videos and black mark indicates normal web videos.

In the similar way, the top 10% (4700) abnormal web videos were detected and extracted with K-nearest neighbour value 5 and 3 seperately. Also the same experiment is extended to detect the common abnormal web videos (ie., intersection) of both LOF outlier detection models. The Table 4 describes category wise result comparision of LOF method with K=3, K=5 and K=7, and results of (K=3 ∩ K=5 ∩ K=7).

Experimental results with LOF(K=3):

Eventhough the abnormality resluts of LOF with K=3 followed the results of LOF with K=5, the bulk in abnormality has found in diverse web video categories including ‘Comedy’, ‘People and Blogs’, ‘Entertainment’, ‘Education’, ‘Sports’, ‘Music’, ‘News and Politics’, ‘Non Profit and Activism’, ‘Films and Animation’and ‘Gaming’.

Experimental results with LOF(K=5):

The majority in abnormality has found in various web video categories including ‘Comedy’, ‘People and Blogs’, ‘Entertainment’, ‘Education’, ‘Sports’, ‘Music’, ‘News and Politics’, ‘Non Profit and Activism’, ‘Films and Animation’and ‘Gaming’. Among these web videos, the experimental observation attracted by the statistics of abnormalities in ‘Non Profit and Activism’ and ‘News and Politics’ categories.

Since, out of 130 web videos of ‘Non-Profit and Activism’ category, 30 (23.07 %) web videos were identified as abnormal web videos because, those web videos were completely neglected by the web users.

Table 3: Structure of top 10% abnormal web videos in decreasing order

Sl.No	ID	Length	Views	Avg Rate	Ratings	Comments
1	25951	586	3281256	4.89	3076	1266
2	2919	98	142806	4.06	898	5772
3	25567	3637	58389	4.64	140	76
4	46991	3125	34097	3.53	34	21
5	21424	1002	4688	4.56	18	9
6	46993	2407	31924	3.96	26	12
7	35275	2172	25499	4.11	239	0
8	20072	1023	3422	4.66	38	45
9	19854	235	86855	4.87	1775	1446
10	23179	821	93	5	1	1
11	19306	1358	13002	4.84	38	38
12	33915	1067	3982	4.5	2	3
13	37140	568	3535	2.9	42	627
14	46994	2566	19294	4.07	15	8
15	16184	1126	8409	4.82	17	0
16	25523	1228	24206	4.84	102	184
17	25834	656	7698	4.93	535	353
18	21409	1546	14661	4.5	46	59
19	36853	446	149759	3.98	353	1531
20	18125	1104	894	5	6	11
21	16175	1073	19350	4.15	13	0
22	46992	3170	42653	3.62	21	10
23	12089	11	458	4.2	10	203
24	928	171	35868	4.52	692	714
25	30940	1000	6150	4	4	0
26	2258	651	2731	2.94	166	381
27	8477	50	2514	4.08	60	64
28	4614	1133	2038	4.65	17	8
29	47639	21	7987	4.55	194	162
30	2422	970	6951	4.61	28	33
31	24688	175	131534	4.85	1844	1687
32	43794	869	496	3	2	0
33	4884	1389	59686	4.85	394	161
34	47031	1197	9636	4.67	6	6
35	16186	1304	15427	4.67	40	0
36	38928	1499	41179	4.76	147	47
37	21259	360	28803	4.9	641	495
38	38947	1500	118490	4.93	853	661
39	2415	1431	27797	4.73	134	174
40	2417	893	11257	4.84	38	39
41	16045	1218	11527	5	15	0
42	33439	51	2161	3.7	148	41
43	24701	480	37815	4.93	469	824
44	16753	1389	74769	4.93	589	408
45	4967	39	130088	3.42	1943	10
46	3568	811	1239	4.81	36	35
47	34090	589	10340	4.84	171	288
48	40919	1439	44867	4.74	23	4
49	17471	179	1091	4.87	85	78
50	20735	545	4363	4.9	260	185

Table 4: Category wise result comparision of LOF method with K=3, K=5 and K=7

Category	No.of Videos	Abnormal Web Videos				Abnormality Percentage			
		K=3	K=5	K=7	K5 ∩K3 ∩ K7	K=3	K=5	K=7	K5 ∩ K3 ∩ K7
Auto and Vehicle	686	32	24	22	12	4.66	3.50	3.21	1.75
Comedy	2885	353	405	389	250	12.24	14.04	13.48	8.67
Education	532	67	69	79	51	12.59	12.97	14.85	9.59
Entertainment	11474	1175	1156	1155	706	10.24	10.07	10.07	6.15
People and Blogs	3637	399	425	421	290	10.97	11.69	11.58	7.97
How to and Style	2017	289	293	298	198	14.33	14.53	14.77	9.82
Music	13974	1031	988	1009	596	7.38	7.07	7.22	4.27
Sports	2821	196	176	172	106	6.95	6.24	6.10	3.76
News and politics	1559	277	288	285	212	17.77	18.47	18.28	13.60
Films and Animation	4631	521	525	523	328	11.25	11.34	11.29	7.08
Non Profit and Activism	130	27	30	29	21	20.77	23.08	22.31	16.15
UNA	238	26	25	29	15	10.92	10.50	12.18	6.30
Travel and Events	878	53	51	50	31	6.04	5.81	5.69	3.53
Pets and Animals	878	102	97	98	70	11.62	11.05	11.16	7.97
Gaming	429	76	73	71	47	17.72	17.02	16.55	10.96
Science and Technology	891	76	75	70	46	8.53	8.42	7.86	5.16
Total	**47660**	**4700**	**4700**	**4700**	**2979**	**11.50**	**11.61**	11.66	7.67

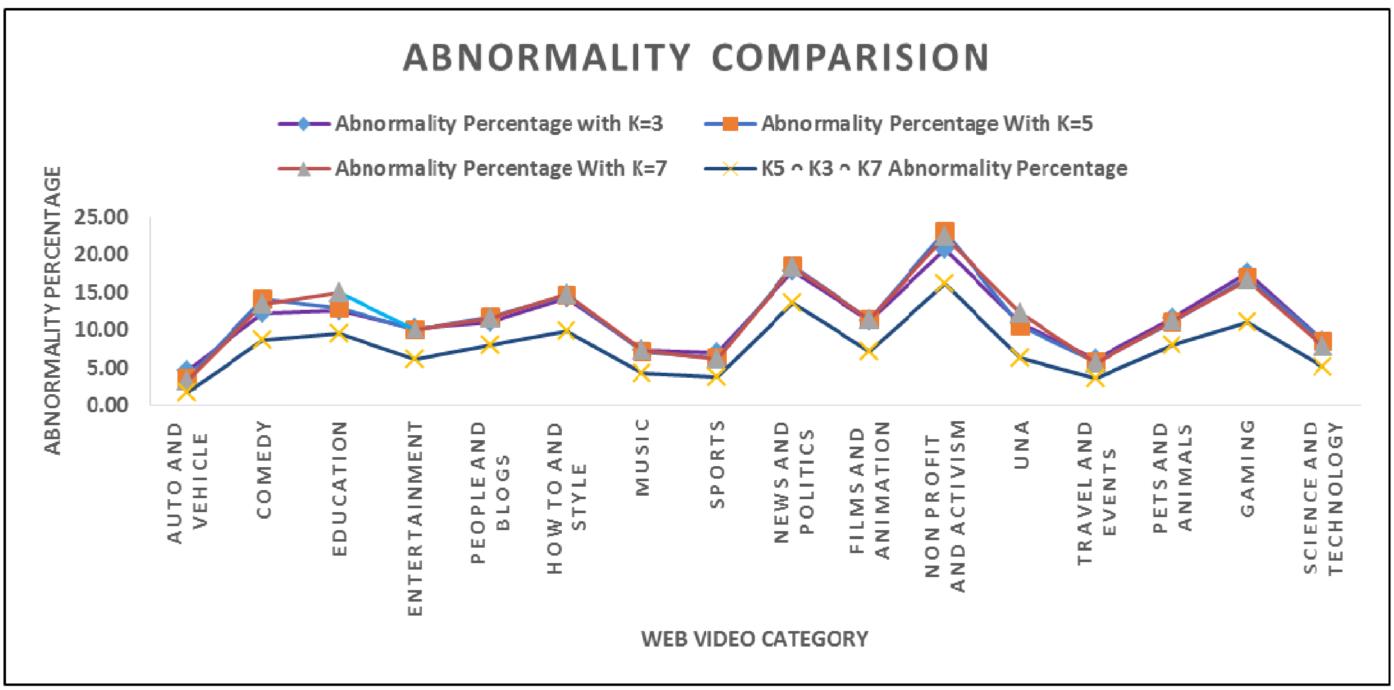

Fig. 4: Category wise result comparision of LOF method with K=3, K=5 and K=7

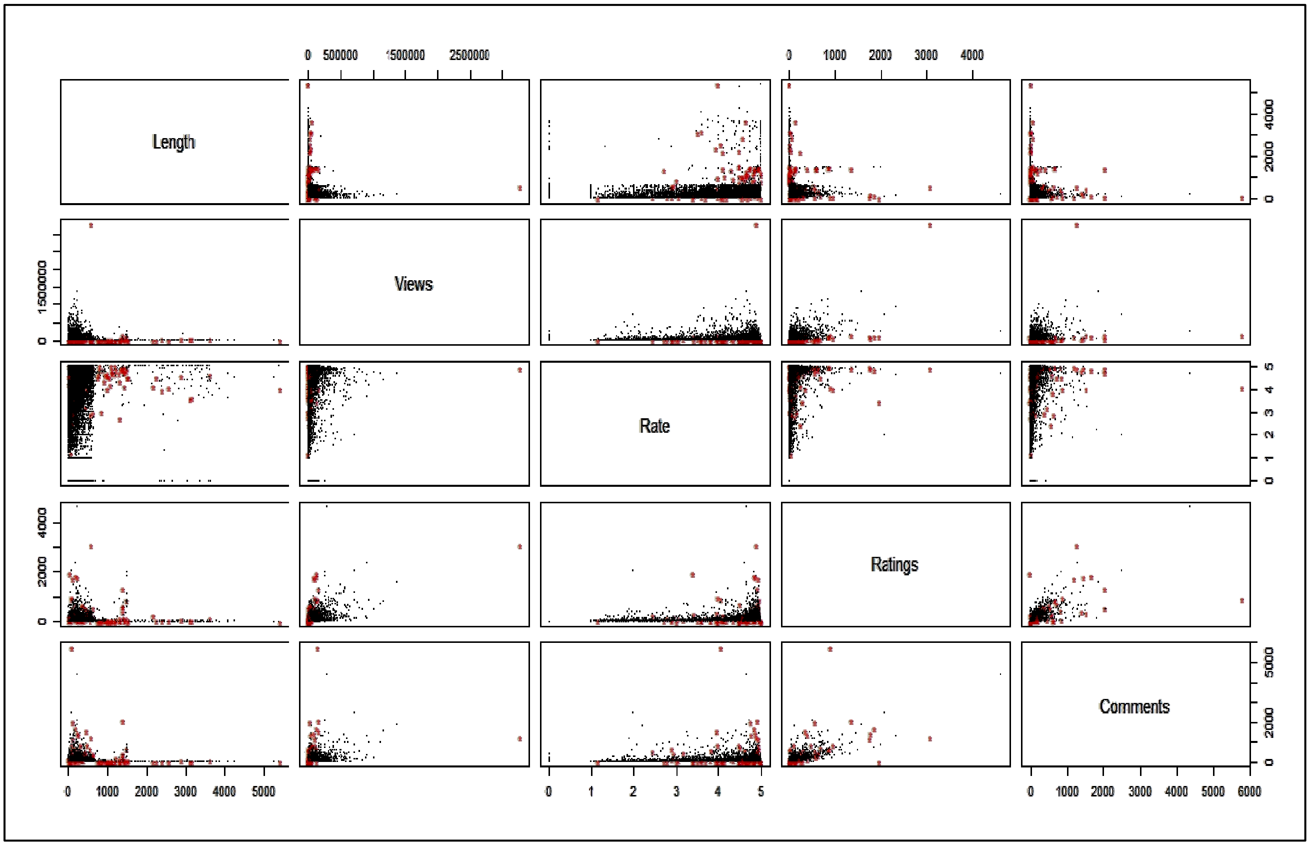

Fig.5: Scatter Plot Matrix representation of outliers

Among 1559 web videos of ‘News and Politics’ web video category, 288 (18.47 %) web videos were identided as abnormal web videos which contains both the cases of viral web videos and neglected web videos. And also it has been noted that, the web videos of ‘Travel and Events’ and ‘Auto and Vehicle’ etc, contains very less percentage of abnormalities, because such kind of web video categories usage are very less as compare to other categories.

Experimental results with LOF(K=7):

The abnormality results with LOF K=7 are described in the Table 4. The statistical reflections reveals that, the web videos of ‘Auto and Vehicle’, ‘Travel and Events’, and ‘Science and Technology’ are found with less number of abnormalities and web videos of ‘Comedy’, ‘Education’, ‘How to and Style’, ‘News and Politics’, ‘Non-Profit and Activism’, and ‘Gaming’ are found with more than 10% abnormalities.

Percentage of Abnormalities with LOF ($K=5 \cap K=3 \cap K=7$):

Further, the experinetal observation attracted the statistical comparison of abnormalities in various categories of abnormal web videos with LOF K=3, LOF K=5 and LOF K=7. It has been observed that, the percentage of abnormalities in various web video categories were nearly same with all the three experimental LOF results.

Hence, to be more accurate in the detection of abnormal web videos, an attempt is made to detect abnormal web videos which are detected by the combination of LOF K=3, LOF K=5 and LOF K=7 outlier detection models, and is described in the Table 4. Also the abnormality comparision is graphically represented in the Fig.4.

The perecntage of abnormalities detected in web videos of ‘Auto and Vehicle’, ‘Music’ and ‘Sports’ categories, and in categories of ‘News and Politics’, ‘Non-Profit and Activism’ and ‘Gaming’ found with more abnormalities as compared to other categories of web videos. The abnormal web videos of ‘News and Politics’ category contains viral videos as majority, however, the abnormal web videos of ‘Non-Profit and Activism’ and ‘Gaming’ categories contains neglected web videos in majority. The common abnormal web videos which are exist in all the three LOF models are found approximately 75%. This fact concludes that, the method we chosen for the detection of abnormalities from the web video metadata object dataset is excellent. In the resultant abnormal web videos, the viral and neglected web video ratio has been found 65:35.

V. CONCLUSION AND FUTURE WORK

In this work, novel attempts are made to detect abnormal videos from large web video meta-object database. The Local Outlier Factor (LOF) with two different cluster size approach were employed with large scale data, so that abnormal web videos based on their meta-objects are found effectively. The proposed LOF method identified and extracted top K abnormal web videos, and labeled the input dataset with 'Outliers' and 'Non-Outlier'. The future work is to build an efficient prediction model for abnormal web video based on web video meta-objects, which should predict the abnormal web videos with less computation time as compared to the proposed LOF method.

REFERENCES

[1] 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.

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

[3] 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

[4] 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.

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

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

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

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

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

[10] 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.

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

[12] Siddu P Algur, Prashant Bhat, "Metadata Based Classification and Analysis of Large Scale Web Videos", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) , Volume 4, Issue 3, May - June 2015 , pp. 111-120 , ISSN 2278-6856.

[13] 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, Volume 4, Issue 2, February-2015.

[14] 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 about to complete 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.

