International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-8 E-ISSN: 2347-2693

Abnormal Facebook Multimedia Detection on Facebook using IQR Method

Siddu P Algur[1], Suraj Jain[2], Prashant Bhat[3*]

[1]Dept. of Computer Science, Rani Channamma University, Belagavi, India
[2]JSS Shri Manjunatheshwara Institute of UG and PG Studies, Dharwad, India
[3]*Dept. of Business and Data Analytics, Christ Institute of Management, Pune, India

[*]*Corresponding Author: prashantrcu@gmail.com*

Available online at: www.ijcseonline.org

Received: 04/Jul/2017, Revised: 15/Jul/2017, Accepted: 10/Aug/2017, Published: 30/Aug/2017

Abstract— Recently, discovering outliers among large scale Facebook multimedia have attracted attention of many Facebook mining researchers. There are number of outlier multimedia exists in each category of Facebook multimedia such as- 'Entertainment', 'Sports', 'News and Politics', etc. The task of identifying and manipulate (to remove from the Facebook or to share with others in the Facebook, or to watch/download from the Facebook etc.) such outlier Facebook multimedia have gained significant important research aspect in the area of Facebook Mining Research. In this work, we propose a novel method to detect outliers from the Facebook multimedia based on their metadata objects. Large scale Facebook multimedia metadata objects such as- length, view counts, numbers of comments, rating information are considered for outliers' detection process. The outlier detection method–Inter-Quartile Range (IQR) is used to find outlier Facebook multimedia of same age. The resultant outliers are analysed and compared as a step in the process of knowledge discovery.

Keywords— Outliers, Inter-Quartile Range, Facebook Multimedia Outliers, Facebook, Metadata

I. INTRODUCTION

Face book and YouTube are recognized as most successful user-generated multimedia sharing sites nowadays. Face book and YouTube have 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 multimedia from a large number of multimedia, Face book and YouTube provides different features/metadata objects such as – view counts, rate, ratings, number of comments, favourites, key words, information regarding likes and dislikes etc.

The objective of this study is to detect outlier multimedia among large scale Facebook multimedia using their metadata objects. To succeed in the proposed objective of the work, large scale Face book dataset are extracted from the standard Facebooksite – UCI Machine Learning Repository[5]. This metadata objects includes various attributes such as- 'Category', 'Likes', 'Sharings', 'Number of Comments', 'Ratings' and 'Posting info', etc of each Facebook multimedia.

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

The main contributions of our work are as follows:

- Facebook Multimedia Metadata Object dataset extraction and effective pre-processing for the experiment.
- The analysis and knowledge discovery process from the resultant unsupervised outliers formed by the built IQR outlier model.

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 Facebook multimedia based on their features/metadata objects. The normal Distribution of each numeric metadata attribute is shown in Fig.1.

The rest of the paper is organized as follows: The section 2 represents related works on the clustering of Facebook multimedia, section 3 represents proposed Facebook multimedia outlier 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 Multimedia Metadata Features Dataset

Feature	Type of information	Source	Data type
Posted	Identification	Facebook	Date/time
Permanent link	Identification	Facebook	Text
Post ID			
Post message	Content	Facebook	Text
Type	Categorization	Facebook	Factor: {Link, Photo, Status, Video }
Category	Categorization	Facebook page managers	Factor: {action, product, inspiration }
Paid	Categorization	Facebook	Factor: {yes, no }
Page total likes	Performance	Facebook	Numeric
Lifetime post total reach			
Lifetime post total impressions			
Lifetime engaged users			
Lifetime post consumers			
Lifetime post consumptions			
Lifetime post impressions by people who have liked your page			
Lifetime post reach by people who like your page			
Lifetime people who have liked your page and engaged with your post			
Comments	Performance	Facebook	Numeric
Likes			
Shares			
Total interactions	Performance	Computed	Numeric

II. Related Work

This section represents some related previous works which are implemented to find abnormal Facebook multimedia/ abnormal Facebook multimedia events using metadata objects.

The authors Chueh-Wei Chang, et al. [1], proposed a framework for spatial relationship construction, abnormal event detection and multimedia 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 multimedia 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 multimedia 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 multimedia 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 multimedia 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.. 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 multimedia 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 multimedia 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 multimedia 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 Facebook 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 Facebook fashion as the algorithm examined new data, avoids any matters with concept drift. Investigational results on hours of real world surveillance multimedia and numerous YouTube multimedia showed that, the proposed algorithm might reliably locate the abnormal events in the multimedia 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 multimedia. The interior of the approach was to detect portion of multimedia 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 multimedia, which consists of both normal and abnormal motions.

The authors Du Tran et al [10] proposed to discover spatiotemporal paths for multimedia event detection. This new formulation was accurately found and locate multimedia events in cluttered and crowded scenes, and was vigorous to camera motions. It was significantly developed the overall detection and localization accurateness over the state-of-the-art techniques.

III. METHODOLOGY

In this section we present novel methodology of the proposed abnormal Facebook multimedia detection approach. The Facebook multimedia metadata objects are extracted from standard Facebook multimedia database Facebooksite [11], pre-processed and stored in a database [13]. Then the refined data will be given to proposed outlier detection models as inputs and resultant abnormal Facebook multimedia will be extracted and analysed for knowledge discovery. The system model of the proposed technique consists of the following components:

A) Facebook Multimedia Metadata Objects Collection Process
B) Data Refinement Process
C) Abnormal Facebook Multimedia Detection Process

D) Result Analysis and KDD Process

A) Facebook Multimedia Metadata Objects Collection Process

The different kind of Facebook multimedia metadata objects are extracted using UCI website [5] and Facebook multimedia metadata objects are then pre-processed stored in a disk [13] with CSV or ARFF file format for experimental purpose.

B) Data Refinement Process

The raw Facebook multimedia metadata objects are pre-processed 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 Facebook multimedia metadata object dataset is presented in Table 1. In the Table 1, the attribute 'Type' is nominal and contains 4 different classes (ex- 'Video', 'Status', 'Photo', 'Link') of Face book multimedia [12]. The remaining attributes are numeric and represents meta-objects of each Facebook multimedia (MM).

C) Abnormal Facebook Multimedia Detection Process

The proposed work uses IQR method to identify abnormal multimedia in the Facebook multimedia metadata object dataset. The procedure to detect abnormal multimedia using metadata objects based on IQR method is discussed as follows:

Inter-Quartile Range (IQR) Method

A quartile is each of the three points that divide a range of data into four equal groups which are described as follows:

Lower Quartile (Q1):

The Facebook meta-objects value such that 25% of Facebook multimedia meta-object value less than this value.
The entry at (value at) Q1 can be calculated by:
¼(total number of entries +1)

Upper Quartile (Q3):

The Facebook meta-object values such that 75% of all metadata object entries are less than this value.
The entry at Q3 is calculated by:
¾ (total number of entries +1)

Median (Q2)

It is defined as middle point of Facebook multimedia meta-object data value that divides the middle two quartiles. It is the value such that 50% of all entries are less than this value.
The IQR is a measure of distribution. It is the difference between the lower quartile and upper quartile. i.e.
IQR=Q3-Q1.

The *Inter-Quartile Range* is frequently used to find outliers in datasets. Outliers are observations that fall below Q1 - 1.5(IQR) or above Q3 + 1.5(IQR)

The algorithm to find abnormal multimedia from Facebook is presented in Algorithm 1.

Algorithm 1: IQR ($m_1,m_2,m_3,\ldots m_n$)
Input: Facebook multimedia dataset ($m_1,m_2,m_3,\ldots m_n$)
Output: Outliers ($m_1,m_2,m_3,\ldots m_n$)
Algorithm:

1. Arrange Facebook multimedia data in order.
2. Calculate first quartile (Q1) of each multimedia data
3. Calculate third quartile (Q3)
4. Calculate inter quartile range (IQR)=Q3-Q1
5. Calculate lower boundary= Q1-(1.5*IQR)
6. Calculate upper boundary= Q3+(1.5*IQR)
7. Detect data elements outside the lower and upper boundary as an outlier.

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 Facebook multimedia, using IQR outlier model. At this stage, the resultant outliers will be analysed in depth to find abnormal Facebook multimedia using their meta-objects.

IV. RESULTS AND DISCUSSION

The meta-objects of different categories of 500 Facebook multimedia are extracted, pre-processed and stored in database [12] [13] for abnormal multimedia detection. Then the Inter-quartile range algorithm has been applied on the Facebook multimedia meta-object dataset to detect outliers/abnormal multimedia using WEKA and R programming.

The Table 2 represents results obtained by the IQR method using WEKA. Out of 500 Facebook multimedia 86 were labeled as outliers and remaining Facebook multimedia labeled as non-outliers. It is observed from the experimental result that, the Facebook multimedia category 'Photo' and 'Status' contains more abnormal Facebook multimedia as compared to remaining categories.

In Fig.1, the X-axis represents 4 different categories of Facebook multimedia; each color represents unique Facebook multimedia category and Y-axis represents Outliers and non-outliers sections. The 'Photo', 'Status', categories contains more number of abnormal Facebook multimedia as compared to remaining categories. The Fig.2 represents percentage of abnormality in each Facebook multimedia categories.

Table 2: Category wise abnormality result

Sl.No	Category	Total	Normal	Abnormal	Abnormality percentage
1	Link	22	19	3	13.63
2	Photo	426	367	59	13.84
3	Status	45	25	20	44.44
4	Video	7	3	4	57.14
Total		**500**	**414**	**86**	**17.2%**

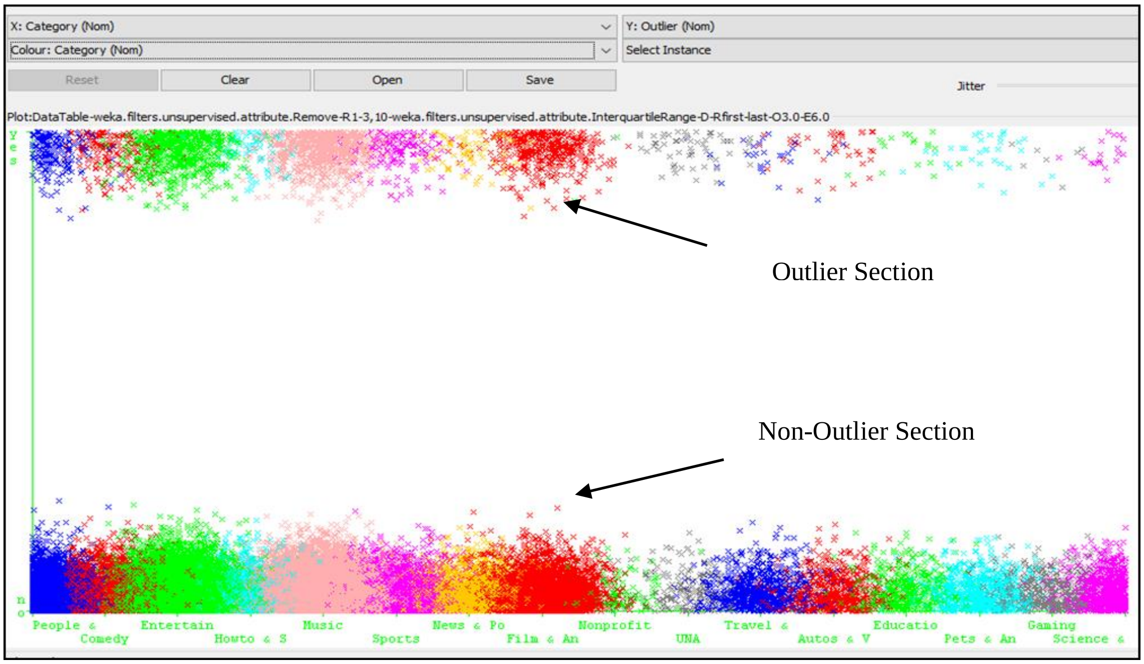

Fig.1: Graphical representation of Outliers/Non-Outliers

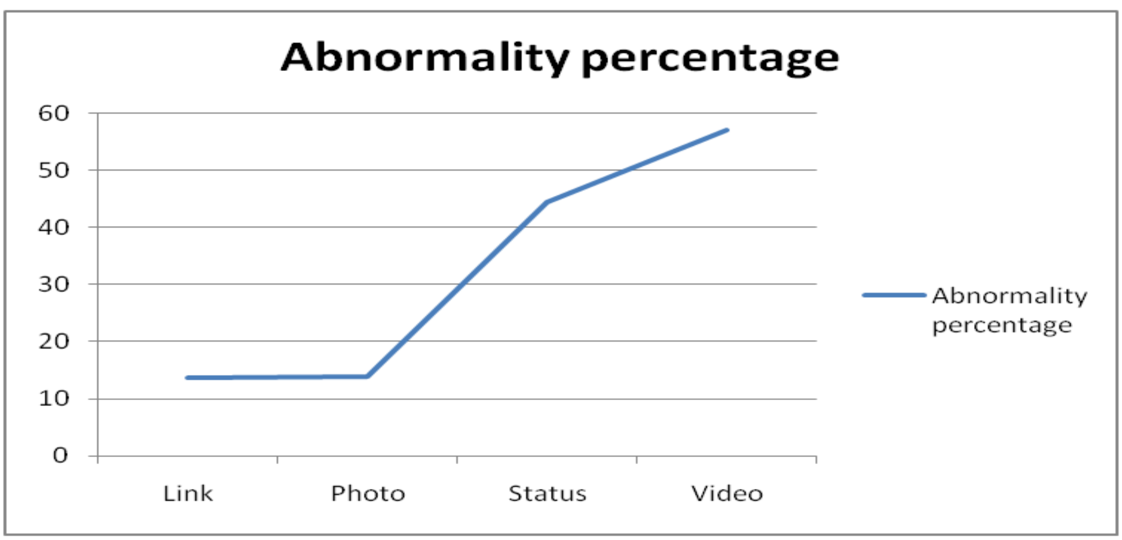

Fig.2: Category wise abnormality percentage

result that, the Facebook multimedia category ‘Photo’ and ‘Status’ contains more abnormal Facebook multimedia as compared to remaining categories.

In Fig.1, the X-axis represents 4 different categories of Facebook multimedia; each color represents unique Facebook multimedia category and Y-axis represents Outliers and non-outliers sections. The ‘Photo’, ‘Status’, categories contains more number of abnormal Facebook multimedia as compared to remaining categories. The Fig.2 represents percentage of abnormality in each Facebook multimedia categories.

V. Conclusion and Future Scope

In this work, novel attempts are made to detect abnormal multimedia from large Facebook multimedia meta-object database. The Inter Quartile Range (IQR) outlier detection technique was employed with large scale data, so that abnormal Facebook multimedia based on their meta-objects are found effectively. The proposed IQR method identified and extracted top 86 abnormal Facebook multimedia and labelled the output dataset with ‘Outliers’ and ‘Non-Outlier’. The future work is to predict abnormal Facebook multimedia based on Facebook multimedia meta-objects.

References

[1] Chueh-Wei Chang, Ti-Hua Yang and Yu-Yu Tsao, “*Abnormal Spatial Event Detection and Multimedia 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 Multimedia by Dynamic Hierarchical Clustering*”, Northwestern University, USA.

[3] Tushar Sandhan et al., “*Unsupervised learning approach for abnormal event detection in surveillance multimedia 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] Y. Cong, J. Yuan, and J. Liu, “*Sparse reconstruction cost for abnormal event detection*,” CVPR 2011, Jun. 2011.

[8] Bin Zhao et al., “*Online Detection of Unusual Events in Multimedia via Dynamic Sparse Coding*”, 2011.

[9] Mahmoudi Sidi Ahmed et al., “*Detection of Abnormal Motions in Multimedia*”, Chania ICMI-MIAUCE’08 workshop, Crete, Greece, 2008.

[10] Du Tran et al., “*Multimedia 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 Multimedia*", http://netsg.cs.sfu.ca/youtubedata/.

[12] Siddu P. Algur, Prashant Bhat, “*Metadata Based Classification and Analysis of Large Scale Facebook Multimedia*”, International Journal of Emerging Trends and Technologies in Computer Science, May-June 2015.

[13] Siddu P. Algur, Prashant Bhat, "*Abnormal Web Video Prediction Using RT and J48 Classification Techniques*", International Journal of Computer Sciences and Engineering, Vol.4, Issue.6, pp.101-107, 2016.