Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 177 (2020) 593–598

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Artificial Intelligence for Natural Language Processing (IA&NLP 2020)
November 2-5, 2020, Madeira, Portugal

Analysis and Classification of User Comments on YouTube Videos

K.M. Kavitha*, Asha Shetty, Bryan Abreo, Adline D'Souza, Akarsha Kondana

Department of Computer Science & Engineering, St Joseph Engineering College, Vamanjoor, Mangaluru 575028, India

Abstract

We categorize the user comments posted on YouTube video sharing website based on their relevance to the video content given by the description associated with the video posted. Comments are analysed for polarity and are further segregated as positive or negative. A comparative analysis of classifier using the Bag of Words and Association List approaches is presented.

Keywords: YouTube Comments; Classification; Video Description; Association list; Bag of words

1. Introduction

YouTube is recognized as the second most popular website in the world by Alexa Internet[1] [1]. The commenting system in YouTube video sharing website allows users to post comments on videos and these comments represent opinions or queries regarding the videos, appreciations to the contributor of the video, or expressions of displeasure towards the video and/or the video contributor. Malicious users nevertheless employ the commenting system as a medium for sharing irrelevant comments. Such comments are commonly referred to as 'spam' as they embody information that are irrelevant in the context of the uploaded video. Automated bots disguised as a user often contribute to spam [1].

Irrelevant comments could be accidental and a result of error in typing or intentional as with malicious links, off-topic hate comments, opinionated comments on controversial topics, etc. Substantial rise in off-topic comments generally deviate genuine participants from their topics of interest. For instance, religion-specific

* Corresponding author. Tel.: +91-9741538886
E-mail address: kavitham@sjec.ac.in

[1] https://try.alexa.com/competitive-website-analysis

1877-0509
10.1016/j.procs.2020.10.084

hate comments distracts user's attention in watching the video further. As irrelevant comments outnumber the relevant ones, identifying specific, relevant and important comments become difficult. User comments seeking clarifications on video content, comments posing queries on the posted video are a few instances of such relevant comments. Identifying and segregating comments that are totally irrelevant to the topic from those that are relevant and important is hence essential.

It is a generally observed gesture among the YouTube user community to appraise the contributors of the video with complemetary comments, when the posted video is heavily liked. Likewise, when the posted video is disliked, a common tendency among users is to bombard the video contributors with negative comments. Typically in a scenario with significant number of user views, the presence of a significant number of positive or negative opinionated comments could highly likely hide the most informative user comments reflecting the video content. Therefore, we consider it important to identify and separate those 'positive' and 'negative' comments from the content specific 'relevant' comments. Thereby, the entire collection of video comments is classified into four categories, namely 'relevant', 'irrelevant', 'positive' and 'negative'. In our classification experiments, we assume that any video posted has an associated description pertaining to the video uploaded. In the said context, the objective of our work is to develop a classifier-based tool for segregation of YouTube comments into afore-mentioned categories with specific attention to the video description.

2. Related Work

In this section, we discuss the related work on spam detection and comment classification exclusively in the context of YouTube Comments.

Language and tokenization independent metric such as content complexity have been employed in feature identification for comment spam detection. Content complexity is estimated over groupings of messages with same author, sender IP, included links [2]. Network motif profiling has been recommended for tracking spam campaigns over time. By conducting content analysis over YouTube user comments, classification schema have been created for comment categorization, revealing as many as ten broad categories, and 58 subcategories reflecting the wide-ranging use of the YouTube comments facility [3].

Severyn et al. focus on opinion mining on YouTube comments by premodeling classifiers for predicting the polarity of opinions and the type of YouTube comments. Tree kernel technology were used to automatically extract and learn features, augmented with robust shallow syntactic structures to improve model adaptability [4]. Asghar et al. provide a review of various techniques to analyze user opinions about a particular YouTube video [5]. Distinguishing feature selector (DFS) and Gini Index (GI) feature selection methods have been reported to provide best classification results in spam filtering task on YouTube [6]. The authors recommend Decision Tree classifiers over Naïve Bayes (NB) for spam filtering on YouTube [6]. Naïve Bayes based multilabel classifier for sentiments of the commenters has been used in understanding the behaviour and response of individuals towards Youtube Video [7]. An extractive frequency-based summarization technique with redundancy control, SumBasic, has been empoloyed in capturing the main concerns reflected in viewers' comments. For emotion classification on Indonesian Youtube comments, word embedding with CNN has been reported to provide best accuracy [8]. Performance of ensemble classifier over single classifier algorithm is analyzed in sorting out spam comments on YouTube videos from the legitimate one [9].

A comparative analysis of common YouTube comment spam filtering techniques show that high filtering accuracy (>= 98%) can be achieved with low-complexity algorithms [10]. This study uses features based on the Edge Rank algorithm and are based on experiments employing nine different learning classifiers such as decision trees, Bayesian nd function-based. Standard machine learning algorithms such as Random Forest, Support Vector Machine (SVM), Naïve Bayes along with N-grams, K-Nearest Neighbour classifiers, Logistic Regression have been used to detect spam comments on YouTube [11] [1] [12] [13] [14] [15]. Studies on content-based analysis techniques have shown that augmenting YouTube comments by incorporating the mood feature help in improve social spam filtering results [16]. In segregating the comments as fake, meta-fake and genuine reviews, Jawaid et al. [17] employ Sentiment Analysis, Negative Ratio Checking and Cosine Similarity for detection of fake reviews and spam content along with other examinations. Fernando et al. [18] use the bag-of-words representation with Multi-nomial Naive-bayes method in classifying 2019

Indonesian Election YouTube comments. In improving the relationship of coding lesson video creators and their viewers, Jain et al. [19] proposed an approach that allows classification and summarization of YouTube comments of only genuine viewers. Bansal's study provides performance-based comparison of deep learning approaches employed in detection of offensive YouTube comments [20].

YouTube allows users to sort the comments based on top comments and newest first. While in the newest first method, sorting relies on the date on which a video was posted, the top comments method considers three features such as the dislike ratio, the number of replies, and the contributor of the video. In our current work, we focus on categorizing the YouTube comments based on content relevance to video description and hence focuses on textual content analysis. In the sections that follow, we present the approach involved in YouTube comment classification and provide the experimental results with discussion.

3. Approach

The steps involved in YouTube comment classification task are outlined in the subsections that follow.

3.1. Data Extraction

The URL for the video provided by the user is first verified for correctness. The video id following '?=v' in the URL should contain exactly 11 characters. YouTube video should have an associated description. As the proposed approach compares a comment with the video description, the description about the video is mandatory and this requirement is verified from the client side. After verifying the video URL entered by the user, the comments, description of the video and author details are extracted using the YouTube API[2].

3.2. Pre-processing

Tokenization and Stopword Removal are performed prior to the feature extraction. During tokenization, each comment is split into individual words with space as the delimiter. Stopwords such as 'the', 'to', 'a', 'an' and so forth were eliminated using the NLTK toolkit. Although comments ocassionally involved words in multiple languages, only stopwords specific to English were eliminated.

3.3. Feature Extraction

Bag of words and association word list based features were explored in our experiments. In the bag of words model, the features extracted represent most frequent words in the video description. The word occurrence in test comment is scored based on its presence or absence in the feature set. In the association word list representation scheme, initially a bag of words is created for video description and thereafter association lists for description and comments are compiled using the Synset[3] provided in the WordNet[4].

3.4. Comment Classification

The user comments are classified into one of the four classes namely 'relevant', 'irrelevant', 'positive' and 'negative'. Classification using the association list approach involves the below steps:

1. Create two separate lists of positive and negative sentiment words using the dataset from Keenformatics[5]. This vocabulary serves as features for positive and negative class.
2. Score the word occurrence in test comment based on its presence or absence in the sentiment word lists.

[2] https://developers.google.com/youtube/v3/docs/comments
[3] http://www.nltk.org/howto/wordnet.html
[4] https://wordnet.princeton.edu/
[5] http://keenformatics.blogspot.com

3. Label the test comment as 'positive' or 'negative' based on the scoring.
4. Create bag of words using the video description and compile its association list using Synset.
5. Extract features by choosing the most frequent words in the association list for description.
6. Compile the association list for test comment.
7. Score the word occurrence in test comment based on its presence or absence in the feature set.
8. Label the comment as 'relevant' based on the scoring. This overrides the classification done in step (3).
9. If the test comment does not satisfy the scoring conditions in steps (3) and (8), label it as 'irrelevant'.

The bag of words model follows a similar approach, with the exception that features are extracted by choosing the most frequent words in the video description represented as a bag of words rather than choosing the most frequent words in the association list for video description.

4. Experimental Setup and Results

4.1. Data Set

Videos from "MS Word" domain comprising of 46 comments were employed in our initial experiments. Among the 46 comments considered, 35 were relevant and 11 were irrelevant comments. Relevant comments were internally further classified as most relevant, positive and negative with counts 22, 12 and 1 respectively. Also, we conducted experiments using user comments corresponding to the videos from "Natural Language Processing", "Sports" and "Movie" domains.

4.2. Results and Evaluation

We extracted the comments using the video URL and manually categorized them into four classes. Experiments were carried out to automatically categorize the extracted comments. The classified comments were evaluated using Precision (P), Recall (R) and Accuracy (A) metrics, estimated as below.

$$Precision = TP/(TP + FP) \tag{1}$$

$$Recall = TP/(TN + FN) \tag{2}$$

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \tag{3}$$

Precision is calculated as the ratio of the number of true positives to the total number of comments classified to be true. Recall represents the ratio of the number of true positives to the total number of comments that are true and accuracy respresents the total comments that are classified as true to the total number of comments that are classified. For relevant comments,TP indicates the number of relevant comments that are correctly classified as relevant, FP indicates the number of non-relevant comments that are incorrectly classified as relevant, TN represents the number of non-relevant comments that are correctly classified as non-relevant comments and FN represents the number of relevant comments that are incorrectly classified as non-relevant comments.

In the first experiment, "Ms Word" data set mentioned in Section 4.1 was used. Results using the bag of words model showed that among the 22 relevant, 12 positive, 1 negative and 11 irrelevant comments, 24 were classified as relevant, 10 as positive, 1 negative and 11 were classified as irrelevant. With the association list approach, 27 were classified as relevant, 8 as positive, 0 negative and 11 were calssified as irrelevant. Figure 1 depicts the relation between the expected and current outcome using association word list and bag of words based models in classifying the comments. The Y-axis in figure 1 shows the number of comments.

The evaluation results for the association word list and the bag of words based comparison of comments is shown in Table 1. Association word list enabled better precision and recall in recognising irrelevant comments, while bag of words model scored better in classifying relevant and positive comments. We repeated the experiments using the video comments extracted from three different domains. Table 2 shows the experimental result obtained in classifying comments for videos from "NLP" (E1), "Sports" (E2) and "Movie"

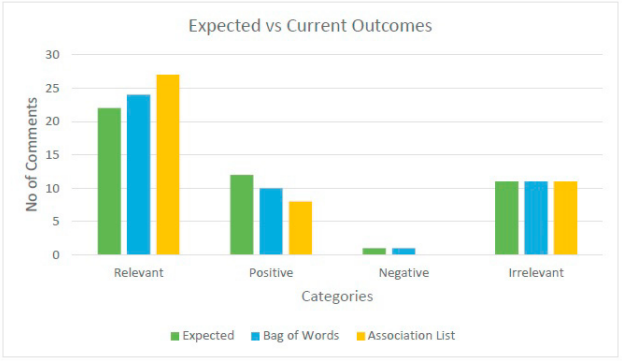

Fig. 1. Comparison of Comment Classification Results using Bag of Words and Association List based Approaches

Table 1. Comment Classification through Bag of Words and Association List based Comparison

	Association Word List							Bag of Words						
Category	TP	FP	TN	FN	P (%)	R (%)	A (%)	TP	FP	TN	FN	P (%)	R (%)	A (%)
Relevant	22	5	19	0	21.48	100	89.13	20	4	20	2	83.33	90.90	86.95
Positive	8	0	34	4	100	66.66	91.3	10	0	34	2	100	83.33	95.65
Negative	0	0	45	1	0	0	97.8	0	1	44	1	0	0	95.65
Irrelevant	11	0	35	0	100	100	100	10	1	34	1	90.90	90.90	95.65

(E3) domains. The comparative analysis of the precision and recall obtained for each category in different experiments conducted are depicted in Figure 2. X-axis in the figure shows various experiments that were conducted and the Y-axis shows the precision values in percentage. Representing comments as association list enables classification of irrelevant comments with precision of 100% in each of the experiments.

Table 2. Comment Classification Results for Videos from Various Domain using Association List and Bag of Words Approaches

		Association Word List							Bag of Words						
	Class	TP	FP	TN	FN	P (%)	R (%)	A (%)	TP	FP	TN	FN	P (%)	R (%)	A (%)
E1	Relevant	88	2	26	1	88	98	97	84	1	27	5	98	94.38	94.8
	Irrelevant	16	0	100	1	100	94.1	99	16	2	98	1	88	94	97.4
	Positive	9	1	106	1	90	90	98	11	3	103	0	78.60	100	97.40
	Negative	0	0	117	0	0	0	100	0	0	117	0	0	0	100
E2	Relevant	22	5	19	0	81.50	100	89.10	20	4	20	2	83.30	90.90	86.95
	Irrelevant	11	0	35	0	100	100	100	10	1	34	1	90.90	90.90	95.65
	Positive	8	0	34	4	100	66.70	91.30	10	0	34	2	100	83.33	95.65
	Negative	0	0	45	1	0	0	97.80	0	1	44	1	0	0	95.65
E3	Relevant	28	5	17	0	84.80	100	90	27	4	18	1	87	96	90
	Irrelevant	5	0	45	0	100	100	100	4	0	45	1	100	80	98
	Positive	10	0	37	3	100	76.92	94	11	1	36	2	91.70	84.60	94
	Negative	2	0	2	46	100	50	96	3	0	46	1	100	75	98

5. Conclusion and Future Work

A classifier-based tool for automatic classification of YouTube comments is discussed in this paper. Comments are segregated into one of the four categories as relevant, irrelevant, positive and negative considering the relevance of the comments to the video content given by description associated with the video posted. In the current work, bag of words and association list based approaches for feature extraction were com-

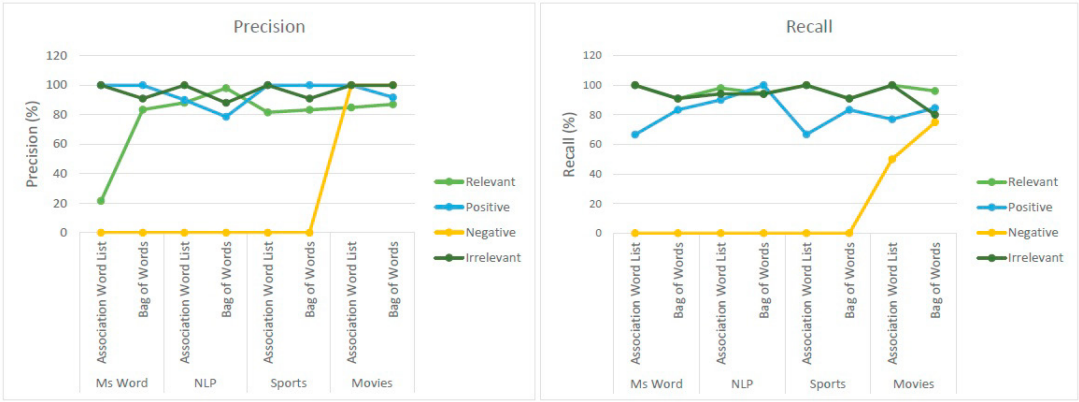

Fig. 2. Precision and Recall for Classification of Comments for Videos from Various Domain

pared. As future work, we intend to classify the YouTube comments involving multilingual phrases or words. Non-contiguous phrases needs to be explored as features in future work.

References

[1] S. Aiyar, N. Shetty, N-gram assisted youtube spam comment detection, Procedia Computer Science 132 (2018) 174–182, 2018 International Conference on Computational Intelligence and Data Science, ICCIDS 2018.
[2] A. Kantchelian, J. Ma, L. Huang, S. Afroz, A. Joseph, J. Tygar, Robust detection of comment spam using entropy rate, in: Proceedings of the ACM Conference on Computer and Communications Security, 2012, pp. 59–70.
[3] A. Madden, I. Ruthven, D. McMenemy, A classification scheme for content analyses of youtube video comments, Journal of documentation (2013).
[4] A. Severyn, O. Uryupina, B. Plank, A. Moschitti, K. Filippova, Opinion mining on youtube (2014).
[5] M. Z. Asghar, S. Ahmad, A. Marwat, F. M. Kundi, Sentiment analysis on youtube: A brief survey, arXiv preprint arXiv:1511.09142 (2015).
[6] T. C. Alberto, J. V. Lochter, T. A. Almeida, Tubespam: Comment spam filtering on youtube, 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA) (2015) 138–143.
[7] A. U. R. Khan, M. Khan, M. B. Khan, Naïve multi-label classification of youtube comments using comparative opinion mining, Procedia Computer Science 82 (2016) 57–64.
[8] J. Savigny, A. Purwarianti, Emotion classification on youtube comments using word embedding, in: 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), 2017, pp. 1–5.
[9] S. Sharmin, Z. Zaman, Spam detection in social media employing machine learning tool for text mining, 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (2017) 137–142.
[10] A. O. Abdullah, M. A. Ali, M. Karabatak, A. Sengur, A comparative analysis of common youtube comment spam filtering techniques, 2018 6th International Symposium on Digital Forensic and Security (ISDFS) (2018) 1–5.
[11] E. Poché, N. Jha, G. Williams, J. Staten, M. Vesper, A. Mahmoud, Analyzing user comments on youtube coding tutorial videos, in: 2017 IEEE/ACM 25th International Conference on Program Comprehension (ICPC), 2017, pp. 196–206.
[12] A. Aziz, C. F. M. Foozy, P. Shamala, Z. Suradi, Youtube spam comment detection using support vector machine and k–nearest neighbor, 2018.
[13] Burhanudin, Y. Musa'adah, Y. Wihardi, Klasifikasi komentar spam pada youtube menggunakan metode naïve bayes, support vector machine, dan k-nearest neighbors, 2018.
[14] N. M. Samsudin, C. F. M. Foozy, N. Alias, P. Shamala, N. F. Othman, W. Din, Youtube spam detection framework using naïve bayes and logistic regression, Indonesian Journal of Electrical Engineering and Computer Science 14 (2019) 1508.
[15] G. Kaur, A. Kaushik, S. Sharma, Cooking is creating emotion: a study on hinglish sentiments of youtube cookery channels using semi-supervised approach, Big Data and Cognitive Computing 3 (3) (2019) 37.
[16] E. Ezpeleta, M. Iturbe, I. Garitano, I. V. de Mendizabal, U. Zurutuza, A mood analysis on youtube comments and a method for improved social spam detection, in: HAIS, 2018.
[17] A. Jawaid, S. Dev, R. Sharma, Predilection decoded: Web based spam detection and review analysis for online portals, 2019.
[18] J. R. Fernando, Udayawibawamukti, Klasifikasi spam pada komentar pemilu 2019 indonesia di youtube menggunakan multinomial naïve-bayes, 2019.
[19] S. Jain, D. M. Patel, Analyzing user comments of learning videos from youtube using machine learning, 2019.
[20] P. Bansal, Detection of offensive youtube comments, a performance comparison of deep learning approaches, 2019.