

Youtube Spam Detection Using Ensemble Method

Nor Azman Mat Ariff*, Mohd Zaki Mas'ud, Norharyati Harum, Erman Hamid and Nazrulazhar Bahaman

Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

*Corresponding e-mail: nazman@utem.edu.my

Keywords: Youtube spam detection; n-gram; SVM;

ABSTRACT – The extensive usage of youtube among the community is one of the reason in the increase of youtube spam. Since youtube spam is difficult to detect, various techniques have been proposed. In this paper, we present youtube spam detection using ensemble method. N-gram extraction from 1-gram to 5-gram is applied to generate feature vector. Two Support Vector Machine(SVM) kernels, linear and RBF are used to train single classifiers. These two kernels are deployed as these kernels are efficient in classifying linear and non-linear separable problems. Ensemble method is constructed by combining probability outputs of each n-gram single classifier to produce a better classification. From the experiment, our proposed ensemble method produces a better predictive performance compared to a single classifier.

1. INTRODUCTION

Popularity of the internet has adversely affected activity visualization in our lives. The transition from the physical world to the digital domain is undeniably enhanced by the significant quality of life, considering the convenience and time savings offered [1]. YouTube is among the most popular video-sharing platforms for social media [2]. YouTube is currently the world's wide video content provider which has become a large medium for the spreading of interactive media information [3]. However, due to the popularity of YouTube, it has been targeted for malicious attack. Recently, various unwanted threats have affected online social networks. The malicious users often post links, advertisements and fraudulent information on the phishing website in the comment section that can transmit viruses or malware.

Filter-based approaches are commonly used for text classification since youtube spam is considered as text classification problems. Aziz et al. [4] compare SVM and k-NN classifiers where SVM give a better performance with 91.49% while k-NN classification rate is 90.59%. In other research, Sharmin and Zaman [5] used k-NN, Naive-Bayes and SVM where Naive-Bayes gave the highest accuracy with 94.57%. Aiyar and Shetty [6] have shown that single classifier has their own domain of competence, therefore is not an optimal approach to solve all problems. Since youtube spam is consider as text classification, most of the proposed solutions to enhance youtube spam detection is to use n-gram or words as a feature to produce a single classifier. In this research, ensemble method is deploy since it can make better prediction than any single classifier. Several n-gram values is considered to be the features to

generate multiple classifiers before these classifiers are used in ensemble method.

2. METHODOLOGY

In this section, we will describe the methodology used in this study. Figure 1 shows the process flow of our methodology. First phase involves classification using single classifier and the second phase involves our proposed ensemble method.

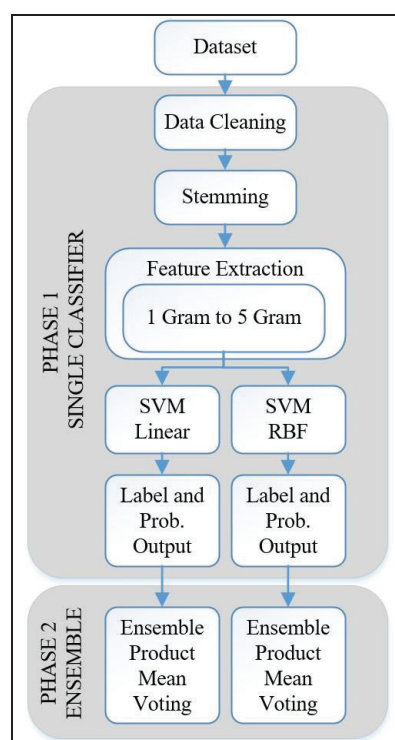


Figure 1 Process Flow of the proposed methodology

Phase 1 starts with data cleaning where stop words, and non-ASCII characters are removed. All words are changed to lowercase to reduce the size of dictionary. Stemming is applied to the words and only root words are extracted. Stemming improves the learning speed and efficiency of the text classification system. N-gram feature extraction then applied to the words start with 1-gram to 5-gram. This extraction will create five datasets, 1-gram, 2-gram, 3-gram, 4-gram and 5-gram datasets. Each dataset will be splitted to 80% and 20% for train set and test set. These train set and test set will go through classification process. The classification process is performed 2 times, the first using linear kernel and the second using RBF kernel. This is to identify which kernel give the best

classification performance. Linear and RBF are deployed as these two kernels are widely used and efficient in classifying linear and non-linear separable problems. In this study, classification will be fine-tuned to produce the class probability outputs.

In phase 2, after the best kernel is identified, the proposed ensemble method will construct probability outputs from 1-gram to 5-gram and combine them before feed to the combination method. Three combination methods are selected namely product rule, mean rule and majority voting.

3. RESULTS AND DISCUSSION

The dataset used in this study is UCI Machine Learning Repository's YouTube Spam dataset. The datasets include five datasets youtube video comments. There are a total of 1005 spam comments and 951 legitimate comments. There are only four features in the dataset namely comment_id, author, date, and content. The evaluation using 10-repeated train-test split procedure where the size of train set and test set are 80% and 20% respectively. In this experiment, SVM with linear and RBF kernels is used as the classifier. Table 1 shows the average classification accuracy (%) using single classifier with linear and RBF kernels. Referring to table 1, classification using linear kernel outperformed RBF kernel for all of the datasets. The best results are illustrated in bold characters. The best classification performances for eminent(99.97%), psy(92.12%), shakira(92.06%), and katy perry(89.43%) are achieved through 3-gram except for LMFAO(96.00%) at 4-gram.

Table 1 The average classification accuracy (%) of Single Classifier, L=Linear Kernel and R=RBF Kernel

Dataset	1G	2G	3G	4G	5G
Eminem(L)	99.09	99.32	99.97	97.95	97.27
Eminem(R)	99.32	90.45	94.55	87.27	50.23
Psy(L)	82.42	90.61	92.12	92.12	90.91
Psy(R)	74.85	74.55	63.33	54.24	57.27
Shakira(L)	86.47	92.06	92.06	90.59	89.71
Shakira(R)	80.59	54.17	54.17	50.00	50.00
LMFAO(L)	86.25	93.50	95.00	96.00	95.75
LMFAO(R)	85.75	84.75	83.25	86.50	75.00
Katy Perry(L)	84.29	88.57	89.43	89.14	88.57
Katy Perry(R)	84.86	65.14	58.29	55.14	49.71

Table 2 shows classification performances for the proposed ensemble method. Since linear kernel gives the best classification performances for all datasets in the earlier experiment, a set of probability outputs from single classifier using linear kernel from 1-gram to 5-gram are constructed and combine them to achieve a strong classification ability. Three combination methods are used namely product rule, mean rule and majority voting. Referring to table 2, all combination methods give almost the same classification performance. Eminem best classification performance is 100% using majority voting, psy best classification performance is 93.03% using product rule, shakira classification performance is 92.35% for all combination methods, LMFAO best classification performance is 96.25%

using mean rule and katy perry best classification performance is 90.86% for both, mean rule and majority voting. Classification results from our proposed ensemble method outperformed all single classifiers.

Table 2 The average classification accuracy (%) of Ensemble Method Using Product Rule, Mean Rule and Majority Voting

Dataset	Product	Mean	M. Voting
Eminem	99.77	99.77	100.00
Psy	93.03	92.73	92.73
Shakira	92.35	92.35	92.35
LMFAO	96.00	96.25	95.00
Katy Perry	89.71	90.86	90.86

4. CONCLUSIONS

In this study, classification of single classifiers using linear gives best performance results from all datasets compare to classification using RBF kernels. The probability outputs of these linear kernel single classifiers from 1-gram to 5-gram then are combined to construct an ensemble method. The classification result shows that the ensemble method outperformed classification performance of single classifier.

ACKNOWLEDGEMENT

Authors are grateful to InforsNet Research Group of Universiti Teknikal Malaysia Melaka (UTeM) for the support.

REFERENCES

- [1] A. Bendovschi, Cyber-Attacks – Trends, Patterns and Security Countermeasures, *Procedia Economic and Finance*, vol. 28, no. December 2015, pp. 24–31, 2015.
- [2] A. K. Uysal, Feature Selection for Comment Spam Filtering on YouTube, *Data Science and Application*, vol. 1, no. 1, 2018.
- [3] M. Wattenhofer, R. Wattenhofer, and Z. Zhu, The youtube social network, *ICWSM 2012 - 6th International AAAI Conference on Weblogs and Social Media*, pp. 354–361, 2012.
- [4] A. Aziz, C. F. M. Foozy, P. Shamala, and Z. Suradi, Youtube spam comment detection using support vector machine and K-nearest neighbor, *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 2, pp. 607–611, 2018.
- [5] S. Sharmin and Z. Zaman, Spam detection in social media employing machine learning tool for text mining, *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems*, vol. 2018-Janua, pp. 137–142, 2018.
- [6] S. Aiyar and N. P. Shetty, N-Gram Assisted Youtube Spam Comment Detection, *Procedia Computer Science*, vol. 132, no. Iccids, pp. 174–182, 2018.