Threat Prediction Based on Opinion Extracted from Twitter

Zerihun Tolla
HiLCoE, Computer Science Programme, Ethiopia
wanofii@gmail.com

Tibebe Beshah
HiLCoE, Ethiopia
School of Information Science, Addis Ababa University, Ethiopia
tibebe.beshah@gmail.com

Abstract

The availability of technology and infrastructure create opportunities for citizens to publicly voice their opinions over social media, but this has created serious problems when it comes to making sense of these opinions. Government and companies don't yet have an effective way to make sense of this users' conversation and interact importantly with thousands of others. Huge amount of opinions are posted and tweeted on the Web. Such opinions are a very important source of information for governments and companies. A lot of researchers describe that users are relying on online opinions to make effective use of user opinions. Unfortunately, due to the increasing number of social media users, many government-opposing groups use access of the web to initiate people in anti-government protests and for violent actions. Many researches are being conducted in the domain and many attempts to approach the topic have been presented. However, researches conducted so far, using opinion mining for threat prediction are rare. In this research, we present how to model and automatically monitor trends and detect opinions which are considered as threats on Twitter. Specifically, we propose a supervised machine learning algorithm which classifies opinions using n-bag of word features.

Keywords: Opinion Mining; Text Classification; Machine Learning

1. Introduction

The advance of technology to access the Internet and the availability of enormous data on the web has motivated most of the research groups [1, 2]. The significant role of analyzing social media and web networks to improve our knowledge of information sharing, maintaining communication, opinion formation, and dissemination has been accepted [3, 4, 5]. But, quantitative studies of the social media content, especially on opinion mining and information technology management, remain rare [5]. Considerable obstacle to social media usage is lack of efficient methodology for selecting, collecting, pre-processing, and analyzing contextual information obtained from the web. However, in the field of opinion mining, many companies have developed their own or proprietary text mining systems for data analysis [6], and researchers in the field of big data analytics have developed expert systems for fraud detection, spam detection and sentiment analysis [3].

Because of the widely accessible and available data in electronic format on social media, developing a systematic approach is necessary, as it helps researchers, organizations, and governments to understand the commonality in various online text data. Using the extracted information from social media, researchers can acquire valuable perceptions into the beliefs, values and attitudes of social media users with regard to the utility of user-written opinion and formation [7, 8]. The available information on social web can help governments to monitor the awareness of people regarding violent actions taken using social networks and aid governments in strategic planning.

The grounded theory approaches studied in [9] analyze social media content and identify the underlying factor structure of collected information to address the gap between the availability of user-written raw text and the contextual information of aggregated data. Nowadays many threats to people and infrastructures can be related to the activities of

people on social media, blogs and forums. Social media platforms like social networks and micro blogs allow users to share messages with others. The amount of data on social media is rapidly increasing and it is difficult to monitor the continuous flow of tweets, blogs, opinions and comments and on websites. Looking at the growth of Internet users and the activities of people on websites, researchers are motivated to do research in the area of opinion mining. Overall the intent of this paper is to develop predictive model that takes people's opinion on twitter blogs and classifies the tweets as active threat, past threat and normal message.

2. Related Work

In the field of opinion mining, several efforts have been made to predict or classify threatening or offensive texts. Prediction of offensive text methods are described in [10] for detecting offensive languages in social media. Weak words and strong offensive words, combined with text mining techniques, are also used in sentiment analysis appraisal approach, like n-grams, Bag-of-Words. They also try to classify users as being offensive or not.

In [11] approaches for text mining to detect cybercrime like Internet predation and cyber bullying are discussed, both via rule-based and statistical approaches. Data mining techniques have been applied to detect and alert suspicious e-mails, coming from terrorists, using machine learning algorithms, emphasizing initially on creating the feature space, and then applying different feature selection techniques as stated in [12]. Setting supervised learning using Decision Tree, which seems to outperform methods like supervised vector machine and Naïve Bayes are also used to classify threat e-mails [13].

Kim *et al.* [8] give an approach for sentiment analysis in case of using twitter list from a corpus. In this context, lists are groups of users who share a common interest. Tweets contain enough information to express identifiable characteristics, interests and sentiments.

It is shown that machine learning is a good and vital tool for sentiment analysis for product and movie reviews from a corpus. For such a case three known machine learning algorithms are applied: Naïve Bayes, maximum entropy, and support vector machines. The threat or failure cascaded across infrastructures has been identified as a key challenge for governments [14]. Infrastructure security could be increased by automated detection of deviant behaviors as stated in [15].

There are opportunities for social media such as Facebook, Twitter and weblogs to help the timely, comprehensive and transparent spreading of information [16]. However, the automatic analysis of social media requires other methods than usual opinion or text analysis. The social media provide huge amounts of visual data which can be analyzed to use textual information for anomaly detection [17].

The authors in [18] tried to predict upcoming events in the future based on micro blog messages from social media. They developed a model that selects the most relevant information during big events and incidents. A lot of researches have been done so far on opinion mining related to improving business by reviewing user opinion commented on products. To the best of our knowledge and reviewing related works, using user opinion for predicting is rare except the work in [19] which tried to develop a model to predict early threats from user tweets in Dutch language. In this paper, we use the previous work on opinion mining as a guide and different data mining techniques as methodology to develop threat predictive model from users' opinion in English language.

3. The Proposed Solution

Our proposed method to classify opinion as a threat or not is an implementation of supervised machine learning using SVM Classifier. We have used a twitter developer API and Rstudio to scrawl user opinions from Twitter. After opinions are extracted and preprocessed we used grammar and tense to label the training data sets. Finally,

supervised classifiers built in RTextTools are trained using n-bag of words as features sets. We also set a parameter to select the algorithm with the best performance and accurately classify test data. Over all the complete proposed system consists of the components shown in Figure 1.

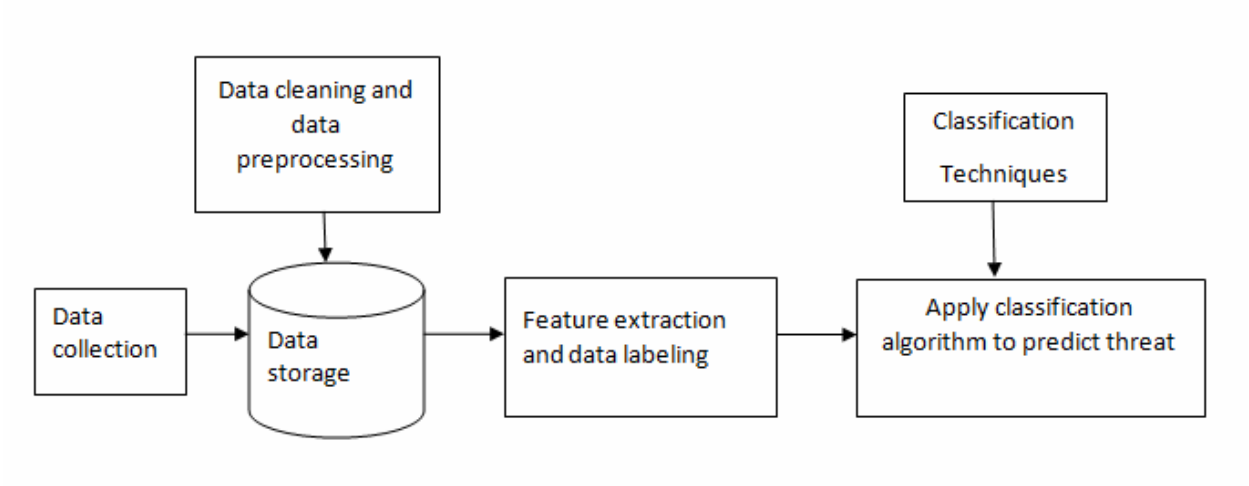

Figure 1: Overview of the Research Process

4. Experiment and Results

4.1 Overview

We experimentally tested our approach using datasets that we collected from Twitter. We converted the datasets into a relational database to make it easier to process and extract features. We extracted only the message from the body of tweets because the rest of attributes are not important to classify the text for our research objective. In order to see whether the extracted tweets can be predicted or not, we prepared a data set containing 500 tweets labeled as past threat activity, 500 tweets having no threat words which are purely normal tweets and 500 tweets labeled as active threats. Five selected classifiers are trained based on the n-bag of words features that we extracted.

4.2 Experimental Setting

As the first step in developing predictive classification model, we selected the actual modeling technique to be used. We used RTextTools which is a machine learning package for automatic text classification that makes it simple for users having a little knowledge of object oriented programming to get started with machine learning, while allowing experienced users to easily experiment with different settings and algorithm combinations. The package includes nine algorithms for ensemble classification (svm, slda, boosting, bagging random forests, glmnet, decision trees, neural networks, and maximum entropy), comprehensive analytics, and thorough documentation. For conducting our experiment, we selected five classification algorithms to classify opinions to three class labels. The class labels are coded as -1, 0, 1 representing past threat, normal message and active threat respectively.

All experiments are first run on the dataset using 10-fold cross validation (cross_validate() function in RTextTools library) to test the validity of our data sets on each algorithm and 75/25 percentage split. Finally, we get the result shown in Table 1.

Table 1: Summary of Experiments

Experiment	*Algorithm*	*Test Technique*	*Accuracy %*
1	SVM	10 fold cross validation	0.92
2	MaxEnt	10 fold cross validation	0.01
3	Random forest	10 fold cross validation	0.88
4	NNETWORK	10 fold cross validation	0.73

Experiment	*Algorithm*	*Test Technique*	*Accuracy %*
5	TREE	10 fold cross validation	0.73
6	GLMNET	10 fold cross validation	3
7	SVM	75/25 percentage split	97.87
8	NNETWORK	75/25 percentage split	73.8
9	Random forest	75/25 percentage split	85.6

4.3 Results and Discussion

The main objective of this paper is building a threat predicting model that helps organizations to predict threat based on user opinion collected from Twitter. To study the domain and achieve the objective, a tool to collect and conduct experiments has been identified. Finally, a predictive model which predicts threats is built.

The classifiers that we adopted in this work are: SVM, MaxEnt, Random Forest, NNETWOR, TREE and GLMNET. As each algorithm uses different parameters and techniques to learn from the training data, we did several experiments. In order to find the optimal classifier which correctly classifies our data, we performed all experiments with cross-validation and make sure that the parameters are not optimized for one particular test set and performed the experiment using 75/25 percentage split of the data sets by doing 10-fold cross validation experiment which works on full data set. Finally, according to the criteria stated above and evaluation technique, to test our model on actual data, we conducted one experiment using new data. For this purpose, we prepared 10 instances of tweets and check if our model predicts the instance to predefined class labels. The results achieved by applying the selected data mining algorithm (SVM) for classification on the collected data reveal that our model has an overall accuracy of 97.87%.

5. Conclusion and Future Work

The growing use of opinion on social media needs text mining, machine learning, and natural language processing techniques and methodologies to organize and extract pattern and classifying user opinion. This paper focused on the existing literature and explored document representation and analysis of opinion extracted from Twitter. We presented a method that can automatically detect threatening and abnormal activities in the real world based on information collected from Twitter. We showed a way to extract user opinion from Twitter and defined features that analyze the content of messages, such as active threat, past threat and pure messages. These features are trained on messages that were selected with a short list of query keywords. This list can easily be modified and extended to refine the existing features or to define new categories for another domain. The grammar and vocabulary used in a sentence separates the type of activities. In combination with our post-processing steps, we are able to report threats and demonstrate activities that have a great impact on a nation's security. We experimentally tested our approach using datasets that we collected from Twitter. The datasets were collected using the Twitter steaming API. We converted the datasets into a relational database to make it easier to process data and extract features. The existing classification methods are compared and contrasted based on their accuracy.

We use n-bag of words feature of tweets to train the classifier. We will extend it by extracting several user-based and tweet-based features from the body of tweets and the users who published the tweet. Furthermore, we will add additional features from the network of the users which is believed to be very informative.

In this work we examined several classifiers independently. One interesting extension to our work would be to implement fusion and boosting methods to combine all the classifiers and benefit from the advantages of all.

Another extension to our work would be to implement some feature engineering methods such as feature extraction to see if more efficient and accurate classifies can be trained. Also techniques such as query expansion can be applied to exploit additional auxiliary information, labeling of the text and developing a dictionary of a threat word to improve the performance of our classifiers.

References

[1] Csorgo, M., Horváth, L., "Limit Theorems in Change-Point Analysis", Wiley, 1997.

[2] Denning, D., "An Intrusion-Detection Model", IEEE T. Softw. Eng. 13(2), pp. 222-232, 1987.

[3] Abrahams, A. S., J. Jiao, G. A. Wang, and W. G. Fan, "Vehicle Defect Discovery from Social Media," Decision Support Systems, Vol. 54, No. 1:87-97, 2012.

[4] Airoldi, E. M., X. Bai, and R. Padman, "Markov-blanket and Meta-heuristic Search: Sentiment Extraction from Unstructured Text," Lecture Notes in Computer Science, Vol. 3932:167-187, 2006.

[5] Bai, X., "Predicting Consumer Sentiments from Online Text," Decision Support Systems, Vol. 50, No. 4:732-742, 2011.

[6] Arnold, E., A. Bruton, and C. Ellis-Hill, "Adherence to Pulmonary Rehabilitation: A Qualitative Study," Respiratory Medicine, Vol. 100, No. 10:1716-1723, 2006.

[7] Karimov, F. P., M. Brengman, and L. Van Hove, "The Effect of Website Design Dimensions on Initial Trust: A Synthesis of the Empirical Literature," Journal of Electronic Commerce Research, Vol. 12, No. 4:272-301, 2011.

[8] Kim S. B., Rim H. C., Yook D. S., and Lim H. S., "Effective Methods for Improving Naïve Bayes Text Classifiers", LNAI 2417, 2002, pp. 414-423.

[9] Strauss, A. and J. Corbin, "Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory", CA, Thousand Oaks: Sage Publications, 1998.

[10] Chen, Y., Zhu, S., Zhou, Y., Xu, H., "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety", ASE/IEEE Int. Conf. Social Computing SocialCom, 2012.

[11] Kontostathis, A., Edwards, L., Leatherman, A., "Text Mining and Cybercrime", Text mining – Applications and Theory, 2010.

[12] Nizamani, S., Memon, N., Wiil, and U. K., Karampelas, P., "Modeling Suspicious E-mail Detection Using Enhanced Feature Selection", Int. J. Modeling and Optimization 2 (4), 2012.

[13] Appavu alias Balamurugan, S., "Learning to Classify Threatening e-mail", IEEE Int. Conf. Modeling and Simulation AICMS, 522-527, 2008.

[14] Eeten, M. van, Nieuwenhuijs, A., Luiijf, E., Klaver, M., and Cruz, E., "The State and the Threat of Cascading Failure across Critical Infrastructures: The Implications of Empirical Evidence from Media Incident Reports", Public Administration 89(2), 381-400, 2011.

[15] Burghouts, G. J., Hollander, R., Schutte, E. A., "Increasing the Security at Vital Infrastructures: Automated Detection of Deviant Behaviors", Proc. SPIE 8019, 2011.

[16] Kleij, R. van der, Vries, A. de, Faber, W., "Opportunities for Social Media in the Comprehensive Approach", NATO RTO-MP-HFM-201, 2012.

[17] Schavemaker, J, Eendebak, P., Staalduinen, M., and Kraaij, W., "Notebook Paper: TNO Instance Search Submission 2011", Proc. TRECVID, 2011.

[18] Weerkamp, W., and Rijke, M. de, "Activity Prediction: a Twitter-based Exploration", SIGIR Workshop on Time-aware Information Access, 2012.

[19] Bouma, H., Raaijmakers, S., Halma, A., and Wedemeijer, H., "Anomaly Detection for Internet Surveillance", Proc. SPIE, Vol. 8408, 2012.