Twitter Data Preprocessing for Spam Detection

Myungsook Klassen
Computer Science Dept
California Lutheran University
Thousand Oaks, USA
e-mail:mklassen@clunet.edu

***Abstract*—Detecting Twitter spammer accounts using various classification machines learning algorithms was explored from an aspect of data preprocessing techniques. Data normalization, discretization and transformation were methods used for preprocessing in our study. Additionally, attribute reduction was performed by computing correlation coefficients among attributes and by other attribute selection methods to obtain high classification rates with classifiers,such as Support Vector Machine, Neural Networks, J4.8, and Random Forests. When top 24 attributes were selected and used for these classifiers, the overall classification rates obtained were very close in range 84.30% and 89%. There was no unique subset of attributes which performed the best, and there were various different sets of attributes playing important roles.**

Keywords-data preprocessing;spam detection; social network; classification.

I. INTRODUCTION

Twitter6 [17] was started in 2006 by Jack Dorsey as an online social networking and microblogging service for users to send and receive short messages (called tweets) of up to 140 characters. Tweeter's 140-character limit on a message serves modern day busy people's trend of acquiring information in a short and quick way. There is much less mindless minutia to read through short tweets. People can spend 5 to 10 minutes on Twitter to find out fast what is happening in the world.

As a result, within the last few years, Twitter has grown to be one of the most popular social network sites with a half billion daily tweets as of October 2012, up from 140 million per day in early 2011. Along with Twitter's growth, spam activities in Twitter have increased and have become a problem. Spamming has been around since the birth of internet and emails, and is not a unique problem with Twitter, but Twitter simply introduces new kinds of spam behavior. Unlike popular social networking service Facebook, or MySpace, anyone can read tweets without a Twitter account, but must register to post tweets. The fact that most accounts are public and can be followed without the user's consent provides spammers with opportunities to easily follow legitimate users.

A recent spamming activity took place during the Russian parliament election December 4, 2011 [8]. For two days after the election, Twitter users posted over 800,000 tweets containing a hashtag related to elections. It turned out nearly half the tweets were spams with unrelated contents, and spam tweets were sent out through fraudulent accounts purchased by a single person in an attempt to disrupt political conversations following the announcement of the election results.

Twitter currently blocks malware by in-house-built heuristics rules using Google's Safebrowsing application programming interface (API) [18] to filter spam activities described in the "Twitter Rules" posted in its web site. Some of spam definitions in the rule are such as an excessive account creation in a short time period, excessive requests to befriend other users, posting misleading links, and posting unrelated updates to a topic using a hashtag "#". Twitter also checks twitter contents with uniform resource locators (URLs) to see if they are on its known harmful sites blacklist database. Harmful sites can be "phishing" sites, sites that download malicious software onto users' computers, or spam sites that request personal information. However, C. Grier et al. [2] show that it takes a few weeks for URLs posted in Twitter to be on its blacklist. In addition to the fact that Twitter itself does to prevent spamming, Twitter relies on users to report spam. Once a report is filed, Twitter investigates it to decide to suspend an account or not. Currently, much research is going on to find a method to detect Twitter spamming in an efficient and automated way. After all, it is not very reliable for the Twitter community to depend on users to identify spams manually based on previous spam activities.

An example of a tweet is shown in Figure 1. It shows a tweet content of a Twitter user "CLU Career Services" with a Twitter ID "CLUCareer". When a Twitter User name or a Twitter user ID is clicked, its public profile page shows a full name, a location, a web page, a short bio, along with tweet contents, the number of tweets, the total number of followers and their Twitter user names, the total number of people the user is following, and Twitter user names of people the user is following. The tweet content in Figure 1 contains a shortened link URL "bloom.bg/11QHmLM" which points to an article page at [19]. Such shortened URLs allow users to post a message within 140 characters, but hide the source URLs, thus providing an easy opportunity for malicious users to phish and spam. This tweet message also contains a topic "Payrolls" which is identified with hashtag "#" in front of it and a *mention* to "BloomerbergNews" user with "@" symbol in front of it.

All this information can be gathered using Twitter API by crawling the Twitter web site. Using these collected "raw" data, different attributes, either content attributes or user behavior attributes [13] can be created.

 ISBN: 978-1-61208-272-1

Figure 1. Tweet example

The number of followers is an example of user behavior attributes while the number of URLs in tweets is a content attribute.

To identify spam tweets, classification machine learning methods which consist of a training (or learning) process and a testing process can be applied. During the training process, learning takes place to generalize information from a given data set which contains a large number of attributes. Often, using too many attributes may cause overfitting of data during training which can hinder classifiers from classifying "new" data correctly. Using too few attributes may not be powerful enough to generalize characteristics of data.

It is critical to use important and relevant attributes and remove redundant and irrelevant ones for a chosen machine learning algorithm to obtain high classification rates. Detecting spam tweets correctly not only provides a better Twitter social network environment in general, but also reduces the chance to anger legitimate users by mistakenly labeling them otherwise.

A desirable case is to use a small number of attributes which can distinguish data in one class from those in another class. Data with a small number of attributes executes fast during a training process and ultimately allows the machine learning system to make a fast classification decision with new data. This is critical in a real time application situation. It would be desirable to have an automated machine learning system which can detect Twitter spams in real time at a fast speed and alert the Twitter authority.

Normalization and discretization of numeric attribute data are widely used preprocessing methods. Discretization eliminates small data observation variations or errors while normalization is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges [6], thus potentially increasing the performance of classifiers. Data transformation is to map data value into another using a mathematical linear or nonlinear function to capture relationships, if any, between attributes.

In this paper, we analyze an impact of preprocessing of Twitter data for a spam detection task. More specifically the purpose is

- to evaluate the impact of using different attributes of Twitter data on different classifiers.
- to evaluate an impact of using a small number of attributes and a large number of attributes.
- to evaluate if preprocessing steps, such as discretization, normalization and transformation with Twitter data may increase classification rates.
- to evaluate if all these steps are consistently needed for all classifiers used.

The rest of the paper is organized as follows. Section II describes related work on spam detection. Section III describes data used, classifier methods, and evaluation methods. Section IV presents actual experiments and results, followed by the last section, Section V, to summarize the study results and future work.

II. RELATED WORK

In this section, we first discuss previous studies on Twitter social media in general to get insight of Twitter data characteristics, then discuss previous studies on machine learning algorithms approaches for detecting and measuring Twitter spam.

Thomas et al. [3] analyzed over one million suspended Twitter accounts to characterize the behavior and lifetime of spam accounts, to understand how spammers abuse legitimate web services such as URL shortening services by exploring spam affiliate programs and market place of illegitimate programs run by spammers. They report that 77% of spam accounts are suspended within the first day of their first tweet and 92% of accounts within three days. Less than 9% of accounts form follower/following relationships with regular users, 52% of spam accounts use unsolicited *mentions*, and 17% used hashtags in the messages with unrelated contents for trend *topic* search. They also report that 89% of spam accounts have fewer than 10 followers.

Link is an important feature to detect spams or to infer user opinions in sentiment analysis [12]. The computed rank depends on the user's connectivity in the social graph. The more followers a user has, the more likely his/her tweets are to be ranked high. Spammers attempt to use this ranking score by acquiring links in Twitter-they follow other users and try to get others to follow them as a courtesy of "social etiquette". Cha et al. [11] and Ghosh et al. [5] studied links in Twitter and reported that popular users who have many followers are not necessarily influential in terms of spawning retweets or mentions. Kwak et al. [4] collected 41.7 million users to study follower-following topology of Twitter and reported that influence inferred from the number of followers and from the popularity of one's tweet do not match. Their findings also show that ranking users by the number of followers matches with results computed by PageRank while ranking by retweets differs from PageRank and from the number of followers and any retweeted tweet reached an average 1000 users regardless of the number of followers of the original tweet account.

Previous study at U.C. Berkeley [14] shows that 45% of users on a social network site readily click on URLs without doubt. Grier et al. [7] collected over 400 public tweets and reported that 8% of 25 million unique URLs posted to Twitter point to phishing, malware, and scam. They reported that the click through rate is 0.13% which is almost twice higher than the email spam click through rate previously published and 80% of clicks occur within the first day of a spam URL appearing on Twitter.

 ISBN: 978-1-61208-272-1

Alex Wang [13] crawled Twitter and collected 29847 users with around 500K tweets and 49M follower/friends relationships. He manually labeled each tweet either as spam or non spam and found that only 1% is spam account. A graph based attribute *reputation* and content based attributes such as existence of duplicate tweets, the number of HTTP links, the number of replies/mentions, the number of tweets with trending topics are used with Bayesian classifier for spam detection and 89% overall classification rate was reported.

McCord and Chuah [14] collected 1000 Twitter user accounts and extracted the following attributes: distribution of Tweets over 24 hours, the number of friends, the number of followers, the number of URLs, the number of replies/mentions, weighted keywords, the number of retweets, and the number of hashtags and ran 4 different classifiers with theses attribute values. They reported that the Random Forest performs the best among 4 classifiers with an overall precision value of 0.957.

Benevenuto et al. [10] gathered a large Twitter data set related to three trend topics and extracted 39 contents attributes and 23 user behaviors attributes which were used with Support Vector Machine (SVM) classifier to detect Twitter spammers. Further description of data can be found at Section III, since our study was conducted using this set of data. They reported classification rates of 70.1% and 96.4% for spam class and non spam class, respectively.

Twitter spammers are known to employ automation to publish tweets. Zhang et al. [16] presented a technique to detect automated twitter content updates. They tested 19436 accounts and reported that 16% exhibit highly automated behavior and verified accounts, most-followed accounts, and followers of the most followed account all have lower automation rates of 6.9%, 12% and 4.2%, respectively.

III. EXPERIMENTAL SETUP

A. Data set

Data from Benevenuto et al. [10] was used as a basis for this study. In his work, Twitter was crawled to collect tweets with three most trendy topics at the time in August 2009 and 1065 legitimate accounts and 355 spam accounts were used for his study. Data contains thirty nine content attributes and twenty three user behaviors attributes, all numeric values, from the raw tweet information. Content attributes are a fraction of followings: tweets replied, tweets with spam words, tweets with URLs, along with the mean, median, min, and max of the followings: the number of hashtags per words on each tweet, URLs per word on each tweet, characters per tweet, hashtags per tweet, mentions per tweet, numeric characters per tweet, URLs per tweet, words per tweet and times a tweet is retweeted.

Two additional attributes which are computed from existing attributes are added to the data for our study. One is reputation defined by Wang [13] as a ratio of the number of followers to the sum of the number of followers and the number of followees. The second is *Influence factor*, which is defined for this study as a ratio of the sum of a number of times mentioned and a number of times a user was mentioned to the sum of a number of times mentioned , a number of times a user mentioned, and a number of times a user replied.

B. Methods

Four classifiers, SVM, random forest (RF), a multi layer back propagation neural networks and J4.8 decision tree implemented in the open source data mining suites WEKA were used in our experiments. WEKA [20] is a data mining software developed at the University of Waikato, New Zealand. For SVM, a program *grid.py* from the libSVM implementation site [6] was used to select two important parameters, C a penalty parameter of an error term and gamma a RBF kernel function coefficient. These values are used for SVM in WEKA.

C. Evaluations

The ten cross validation is used to measure the generalization performance of classifiers used in this research. The method first partitions data into 10 equal sized segments and in each iteration, 9 different segments are used for training and 1 remaining segment is used for testing. This repeats 10 times and an average of 10 results from testing segment is computed.

Classifier performance results are discussed using values derived from a confusion matrix. TABLE I shows a confusion matrix of two classes.

TABLE I: CONFUSION MATRIX

	Predicted Class1	Predicted Class2
Actual Class1	a	b
Actual Class2	c	d

True Positive (TP) for class 1 is a/(a+c) and False Positive(FP) for class 1 is c/(c+d). Precision for class 1 P= a/(a+c) is the ratio of the number of data predicted correctly to the total predicted as class 1. Recall for class 1 R= a/(a+b) is the ratio of the number of data correctly predicted to the number of data in class 1. TP, FP, P and R for class 2 are similarly defined. A classification rate or an average weighted TP rate in WEKA is defined as the ratio of the number of correctly predicted data to the total number of data in both classes, (a+d)/(a+b+c+d). F-measure is a weighted average of the precision P and recall R to measure of a test's accuracy and is defined as 2* P*R/(P+R).

IV. EXPERIMENTS AND RESULTS

A. Normalization

The original data set has a vast range of attribute values. Seventeen attributes such as a *fraction of tweets replied* are

 ISBN: 978-1-61208-272-1

in a range between 0 and 1 while most of content based features such as the number of followers are in a range 0 and over 40000. The age of an account and an elapsed time between tweets are measured in seconds so values range between 0 and 87,000,000. We investigated if attributes in greater numeric data ranges dominate their significance in learning and produce inaccurate classification. Experiments without normalization and with normalization in a range -1 and 1 were performed with libSVM and results shown in TABLE II. It also shows that without normalizing data, the spam class was predicted very poorly with only 6.8% TP rate and the overall classification rate of 68.9%. With data normalization, not only the classification rate went up significantly to 88.3%, but also more importantly the spam class TP values went up to 75.2%! When data is normalized between 0 and 1, similar results to those with normalization between -1 and 1 are obtained.

Our next experiment is to evaluate sensitivity of classifiers with data normalization to obtain high classification rates. Classification rates obtained with 4 classifiers using both original data set and normalized data set and results are presented below in TABLE III. Without data normalization, both the multi-layer neural networks and libSVM show a low performance with 68.68% and 70.42% classification rates respectively, while J48 and Random Forests consistently perform well regardless of data normalization.

B. Manual atttributes selection

A manual attribute selection process used for this study is based on the notion that an attribute with a high correlation with the "class" attribute, but with a low correlation with other attributes is a "good" attribute. Correlation between the "class" attribute and an attribute being reviewed is computed for all attributes. For instance, the "*existence of spam words in the screen name"* attribute has -0.01085 correlation values with the "class" attribute, so this attribute is considered not very useful and is eliminated from the attribute list. Attributes *min of the number of URLs per tweet* and *max of the number of URLs per tweet* showed a similar trend and as a result, they are eliminated from the attribute list. Similar steps were taken with all other attributes for selecting good attributes.

Redundancy of attributes is evaluated by their correlation values which are shown in TABLE III. When there is high correlation between two attributes, one attribute is eliminated and if there is no strong correlation between two attributes, both are kept. In the case of the number of words per tweet and the number of characters per tweet, there is a little correlation so both are kept. After this manual process, 24 attributes are selected.

C. Using WEKA attribute selection methods

Chi Squared Attribute selection, Filtered Attribute selection, Info Gain Attribute evaluation, Gain Ratio Attribute Eval and oneR AttributeEval with Ranker selection method were used to rank original 62 attributes. Top 10 ranking from Filtered Attribute Evaluation and Info Gain Attribute Evaluation are quite similar, but are quite

TABLE II. BETTER PERFORMANCE OF NORMALIZED DATA WITH SVM

Without scaling				
	TP rate	FP rate	Precision	Recall
No spam class	1	0.932	0.682	1
Spam class	0.068	0	1	0.068
Classification rate	0.689	0.622	0.788	0.689
Data scaled [-1,1]				
	TP rate	FP rate	Precision	Recall
No spam class	0.947	0.248	0.885	0.947
Spam class	0.752	0.053	0.876	0.752
Classification rate	0.883	0.183	0.882	0.883

TABLE III. PAIRED-T-TEST OF F MEASURE OF CLASSIFICATION RATES

	libSVM	**MultiLayer NN**	**J48**	**Random Forest**
Original data	68.68	70.42	83.22	88.55
Normalized data	87.59	87.47	85.54	88.16

TABLE IV. CORRELATION COEFFICIENTS OF SOME PAIRED ATTRIBUTES

Two Attributes selected for correlation inspection	**Correlation coefficients**
number of hashtags per word on each tweet(mean). number of hashtags per tweet(mean).	0.886909
number of posted tweets per day(mean). number of posted tweets per week(mean).	0.742563
number of followees of a user's followers. number of followees.	0.903385
number of followees of a user's followers. number of followers.	0.862586
number of words per tweet (mean). number of characters per tweet(mean).	0.349

Different from those obtained by 3 others and top 10 rankings from these 3 methods have only a few attributes in common. Simply there is not a group of top 10 attributes common for all selection methods. TABLE V shows three attributes, *the number of followers per followees, fraction of tweets replied, and the number of times the user replied* are on the top 10 ranking attributes for all 5 selection methods. When a comparison is made for top 24 attributes, a situation is worse-there is less percentage of common attributes selected by all 5 methods.

However, despite different rankings of attributes, when top 24 attributes were used with classifiers, all of them showed compatible classification rates, slightly lower than results obtained when manually selected 24 attributes were used as shown in TABLE V.

 ISBN: 978-1-61208-272-1

D. Data discretization

This process reduces the number of possible values for attributes with continuous values. Equal-interval binning was performed. An objective is to measure an effectiveness of data discretization on classifier performance. With the original data set of 62 attributes, attributes with a very large data range such as the number of followees, the number of followers, number of tweets, and age of the user account are discretized into 10 bins and results such as TP, P and F values of spam class and no spam class along with the overall classification rate are as shown in TABLE VII. Classifier performances of all classifiers are compatible to those obtained from data normalization.

V. CONCLUSION AND FUTURE WORK

In this paper, we explore attribute reduction and data preprocessing such as data normalization and discretization from the aspect of Twitter spam detection using various machine learning algorithms. Four classifiers SVM, back propagation multilayer neural network, decision tree J4.8 and random forest.

TABLE V. COMMON TOP TEN RANKING ATTRIBUTES SELECTED

	attributes
Attributes ranked in a top 10 by all selection methods.	• Number of followers per followees. • Fraction of tweets replied. • Number of times the user replied.
Attributes ranked in a top 10 by 4 selection methods.	• Fraction of tweets with URLs. • Average number of URLs on each tweet. • Number of times user replied. • age of the user account.
Attributes ranked in a top 10 by 3 selection methods.	• Average number of hashtags per word on each tweet.

TABLE VI. CLASSIFICATION WITH DIFFERENT ATTRIBUTES OF DIFFERENT CLASSIFIERS

	libSVM	**MultiLayer NN**	**J48**	**Random Forest**
Manually picked 24 attributes	87.90	87.72	85.98	88.39
ChiSquare	84.57	87.83	86.19	87.95
GainRatioAttributeEval	86.83	87.71	86.39	87.90
InfoGainAttributeEval	86.12	87.23	86.12	88.91
OneRAttributeEval	86.23	87.72	86.35	87.42

Using correlation coefficients, good attributes are selected manually and are compared to those obtained with 5 WEKA attribute selection methods. With correlation coefficient and along with Twitter data structure information, a total of 24 attributes were selected and results obtained with 4 classifiers were very close, or slightly higher values obtained with attributes selected by WEKA attribute selection methods.

TABLE VII. CLASSIFICATION RESULTS (%) FROM DATA DISCRETIZATION

	Overall TP	**Nospam TP**	**Spam TP**	**Nospam P**	**Spam P**	**Nospam F**	**Spam F**
libSVM	87.32	0.962	0.696	0.863	0.901	0.91	0.785
ML NN	87.32	0.942	0.735	0.877	0.864	0.908	0.795
J4.8	86.29	0.944	0.701	0.863	0.862	0.902	0.773
RF	88.26	0.965	0.718	0.873	0.911	0.916	0.803

The newly introduced attribute *influence factor* has a much higher correlation coefficient (> 0.35) with the 'class" attribute while its component attributes, *the number of times mentioned, the number of times the user was replied, the number of times the user replied,* has correlation coefficient values 0.11449, 0.149, and 0.256 respectively with the class attribute. This means the better attribute "influence factor" was created through a linear data transformation of existing three attributes to boost the classifier performance.

When the original data was used initially, the overall classification rate obtained with libSVM was 68.9% with TP value of no spam class= 1 and TP value of spam class =0.068 which is extremely low. This is a much skewed classification result that most of spam accounts were misclassified as non spam while non spam accounts were 100% correctly classified. The probable explanation for this is that attributes which represent characteristics of the spam class have small data ranges compared to other attributes and thus can't contribute to correctly classify the spam class. With back propagation multi layer neural network, the similar results and explanation can be applied. Both libSVM and multilayer back propagation neural network consider and add up all attributes and as a result, scaling becomes important. In tree decision algorithm such as J48 and Random Forest, each attribute is individually considered for information gain, so normalization is not an important factor and such is the case with twitter data.

Data discretization produced similar results to those obtained by data normalization process, classification rates of libSVM and multi layer neural networks increased significantly and TP value of the spam class increased dramatically over to 0.7 from 0.068. We also report that when the number of bins was increased from 10 to 15, classification performance changed less than 5%.

Regardless of all methods applied, the highest TP of the spam class is relatively low with a value of 0.752 while the highest TP obtained in our experiments for the non spam class is quite high with a value of 0.964. This is a slightly higher value than 0.701 reported by Benevenuto [10] where data for our study came from.

A close comparison with two other works [13][14], which used similar attributes and reported higher spam class TP rates is further discussed to hopefully understand why our data shows low non spam class TP rates and to continue this study in the future work.

 ISBN: 978-1-61208-272-1

McCord [14] collected data at random without considering trendy topics while Benevenuto did with three specific popular topics at the time Twitter data was gathered. The two unique attribute used in McCord's work are the *word weight metric* which is a difference between the weight of spam words and the weight of legitimate words in tweets. The sum of all weights is used as a "word weight metric" attribute. This weight parameter controls a probability that a word can be in a list of spam words and a probability that it can be in the regular word list. And the other unique attribute used for their study is the time of a day a tweet was posted. The rationale of this attribute is that spammers work at night.

Wang [13] reported TP value 0.89 of the spam class and the overall classification rate 91.7% and the recall value R 0.917 with Naïve Bayesian classifier. With SVM, neural networks, and J48 decision tree, 100%, 100%, and 66.7% overall classification rates respectively reported. But the reported recall rates for three classifiers are very low, 0.333, 0.417, 0.25 for decision tree, neural network, SVM respectively. In his work, like McCord's work, random Twitter accounts were collected without considering trendy topics. And the unique attributes used in his work are reputation which we used for our study and the total number of duplicate tweets which is computed by using Levenshtein distance. The rationale is that spammers use different user names to post the same contents. An observation worth mentioning is that the number of hashtags for the spam class in his study is lower than those for the non spam class. Many spam accounts have less than or equal to 2 average hashtags while non spam accounts have anywhere from 0 to 20 (an estimate of an average number of hashtags by a quick visual inspection of Figure 7d in his work is about 7).

This is quite contrast to what Benevenuto's data shows regarding the hashtags that spammers post much higher fraction of hashtags per tweet. This contrast may come from the fact that data crawled from Twitter by Benevenuto et al. was using trendy topics while Wang's was gathered at random without using any trendy topic. This comparison suggests that spammers use more hashtags to capture legitimate users attention when there are hot trendy topics being discussed but when there is little hot trendy topics being discussed among Twitter users, usage of hashtags among spammers and among non spammers is not much different .

In conclusion, our study shows that normalization, transformation and discretization improve Twitter spam/no spam classification rates, especially the spam class when libSVM and back propagation multi layer neural networks were used. And our study demonstrates that when using a smaller number of attributes selected manually using correlation coefficient, equally high classification rates were obtained. This is an important finding for a real time detection of spams. Further investigation of Twitter characteristics is needed to understand why the spam class TP value is not as high as we hope to.

REFERENCES

[1] D. Terdiman, "Report: Twitter hits half a billion tweets per day," http://news.cnet.com, December 2012. Retrieved: January, 2013.

[2] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: The underground on 140 characters or less," Proceedings of CCS' 10 Conference, pp. 27-37, October 2010.

[3] K. Thomas, C. Grier, V. Paxson, and D. Song, "Suspended Accounts in Retrospect: An analysis of Twitter spam," Proceedings of IMC conference, pp. 243-256, November 2011.

[4] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter,a social network or a news media?" Proceedings of WWW2010 conference, pp. 591-600, April 2010.

[5] S. Ghosh, B. Viswanath, F. Kooti, N. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. Gummadi, "Understanding and combating link farming in the twitter social network," Proceedings of WWW2012 conference, pp. 61- 70, April 2012.

[6] C. Hsu, C. Chuan, and C. Lin, "A Practical guide to support vector classification," http://www.csie.ntu.edu.tw/~cjlin. Retrieved: January, 2013.

[7] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam:the underground on 140 characters or less," Proceedings of CCS' 10 conference, pp. 27-37, October 2010.

[8] http://www.icsi.berkeley.edu/icsi/gazette/2012/05/twitter-spam. Retrieved: December, 2012.

[9] W.S. Sarle. Neural Network FAQ, 1997, ftp://ftp.sas.com/pub/neural/FAQ.html. Retrieved: December, 2012.

[10] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting Spammers on Twitter," Proceedings of CEAS 2010 conference, pp. 21-30, July 2010.

[11] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter: The million follower fallacy, "Proceedings of ICWSM 2010 conference, pp. 10-17, May 2010.

[12] J. Rabelo, R. Prudencio, and F. Barros, "Using link structure to infer opinions in social networks," Proceedings of SMC conference, pp. 681- 685, October 2012.

[13] A. Wang, "Don't follow me: spam detection in twitter," Proceedings of 5th international conference on security and cryptography, pp. 142-151, July 2010.

[14] M. McCord, and M. Chuah, "Spam detection on twitter using traditional classifiers," Proceedings of international conference on autonomic and trusted computing (ATC) conference, pp. 175-186, September 2011.

[15] U. C. Berkeley. Twitter Spam. International computer science institute (ICSI) ICSI Gazette, May 2012.

[16] C. Zhang, and V. Paxson, "Detecting and analyzing automated activity on twitter," Proceedings of PAM conference, pp. 102- 111, March 2011.

[17] http://www.twitter.com. Retrieved: December, 2012.

[18] https://developers.google.com/safe-browsing/. Retrieved: January, 2013.

[19] http://www.bloomberg.com/news/2012-12-07/. Retrieved: December, 2012.

[20] http://www.cs.waikato.ac.nz/ml/weka/. Retrieved: December, 2012.

 ISBN: 978-1-61208-272-1