INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ENGINEERING AND TECHNOLOGY (IJARET)

ISSN 0976 - 6480 (Print)
ISSN 0976 - 6499 (Online)
Volume 5, Issue 6, June (2014), pp. 40-44
© IAEME: http://www.iaeme.com/IJARET.asp
Journal Impact Factor (2014): 7.8273 (Calculated by GISI)
www.jifactor.com

IJARET
©IAEME

SPAM DETECTION IN SOCIAL BOOKMARKING SYSTEM– EMPIRICAL EVALUATION

Mittal Sejpal[1], Priyank Thakkar[2]

Department of Computer Science & Engineering, Institute of Technology
Nirma University, Ahmedabad, 382481, Gujarat, India

ABSTRACT

Social Bookmarking web sites have recently become popular for collecting and sharing of interesting web sites among users. People can add web pages to such sites, as bookmarks and allow themselves as well as others to work on them. One of the key features of the social bookmarking sites is the ability of annotating a web page when it is being bookmarked. The annotation usually contains a set of words or phrases, which are collectively known as tags that could reveal the semantics of the annotated web page. Efficient and effective search of web pages can then be achieved via such tags. However, spam tags that are irrelevant to the content of web pages often appear to deceive other users for malicious or commercial purposes. There are users who intentionally assign spam tags. Manual detection of such users is very difficult.In this paper, main focus is on the detection of spam users in Social Bookmarking System. Experimental Evaluation is done using ECM PKDD discovery challenge 2008 data set. Experimentation using naïve Bayes and K-Nearest Neighbour classifiers on all three Information Retrieval (IR) models (Boolean, bag-of-words and TFIDF) gives promising results.

Keywords: Spam Detection, Social Bookmarking Systems, Feature Selection.

I. INTRODUCTION

Social Bookmarking Systems have gained high popularity now a day. With this growing popularity, the spam users have started exploiting tags used during annotating process for malicious and commercial purposes. In social bookmarking websites, users can store and access their bookmarks online through a web interface. The stored information is sharable among users, allowing for improved searching. Any system that is highly dependent on user-generated content is vulnerable to spam in one form or another [3]. Social bookmarking websites provide a large and continuously growing pool of potential customers. In fact, a spam post is more attractive than a non-spam post [5].

Spamming in Social Bookmarking Systems has become a crucial challenge affecting both users and service providers. Users face spamming obstacles while doing activities like web based searching [10]. The main challenge is that the characteristics and behavior of spam user's change over time, hence, maintaining the rules for detecting spam is a very difficult task. It is very difficult to manually detect spam users because of huge number of user's data etc. [10].

The data set used in this work is taken from the ECML PKDD discovery challenge 2008 composed of tags, bookmarks and bibtex post of users. Researchers with the help of this data set can now conduct supervised learning and reliable evaluation of the task.

In this paper, the objective is to automate the task of spam user detection in social bookmarking systems. The paper is organized as follows. In Section II, related work is described. In Section III, brief description of classification methods used is given. Section IV consists of implementation methodology. The paper finally ends with conclusions in section V.

II. RELATED WORK

Spam Detection in Social Bookmarking system is a relatively new research area and the literature is still sparse. A brief review of related work and the recent shift of attention toward social spam are discussed here.

A method for spam detection using language models based on the intuitive notion that similar users and posts tend to use the same language was proposed in [3]. Authors in [5], proposed a novel, fast and accurate supervised learning method as a general text classification algorithm for linearly separated data for the task of spam detection. A machine learning-based approach to automate spam detection was proposed in [6].

An algorithm to identify spammers from the collaborating systems by employing a spam score propagating technique was proposed in [7]. The problem of learning to classify texts was addressed by exploiting information derived from both training and testing sets in [8]. To accomplish this, clustering was used as a complementary step to text classification, and was applied not only to the training set but also to the testing set [8].

III. CLASSIFICATION METHODS

In this paper, K-Nearest Neighbor (KNN) and Naïve Bayes (NB) is used for the task of spam user detection in social bookmarking system.

KNN Classifier

KNN is an algorithm that is extremely easy to see yet works extraordinarily well in practice. Likewise, it is astonishingly adaptable and it has already been applied in various domains. For the test instance, it predicts the class based on the class of k-nearest neighbours from the training set. That is, k instances from the training set which are most similar to the testing instance are utilized to predict the class label of test instance.

To find the distance or similarity between instances, one can use different distance measures such as Euclidean distance, Hamming distance, Cityblock distance, cosine similarity etc. Details can be found in [12].

Naïve Bayes Classifier

The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. It considers supervised learning from a probabilistic point of view. The task of classification can be regarded as estimating the class posterior probabilities given a test example.

Naive-Bayes classifier assumes class conditional independence. Given test instance X, Bayesian classifier predicts the probability of belonging to a specific class C_i. To anticipate probability it utilizes idea of Bayes' theorem. Bayes' theorem is valuable in that it gives a method for computing the posterior probability,$P(C|X)$, from $P(C)$, $P(X|C)$, and $P(X)$. Bayes' theorem states that

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

where,

$$P(X|C) = P(x_1, x_2, x_3, \dots x_n \,|\, C) = \prod_{i=1}^{n} P(x_i|C)$$

Here, $x_1, x_2, \dots, x_n$ are features describing X. An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters necessary for classification. Based on the type of data, appropriate distribution is fitted to the data to evaluate $P(x_i|C)$. Details can be found in [11], [12].

IV. EXPERIMENTAL METHODOLOGY

Performance Parameters

Accuracy is used as the parameter for evaluating the performance of the classification algorithms. Accuracy of a classifier on a given test set is the percentage of test set items that are correctly classified by the classifier. In this paper, non-spammer user is considered as positive class and spammer user is considered as negative class. Accordingly, True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) are defined as under [11].

TP (True Positives): It refers to the number of positive items that are correctly labelled by the classifier.

FP (False Positives): It refers to the number of negative items that are incorrectly labelled as positive.

TN (True Negatives): It refers to the number of negative items that are correctly labelled by the classifier.

FN (False Negatives): It refers to the number of positive items that are mislabelled as negative.

Base on the above interpretations, accuracy is defined by the following equation.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Dataset

The assessment of the proposed method is carried out using ECML PKDD Discovery Challenge 2008 Data set.The Data set consists of 7171 unique users. The tables in the data set are:

tas, tas_spam
bookmark, bookmark_spam
bibtex, bibtex_spam
user

The user table contained flag denoting whether the user is spammer or non-spammer. This is helpful to us for supervised learning.

Preprocessing

Pre-processing of data set is necessary to improve the quality of data thereby helping to improve subsequent classification processes. Each of the user is described by means of the tags posted by him. First of all, all the special characters are removed and all the characters are converted to lower-case letters. Stop words which are not important in describing tag profile of the user are than removed. For each of the user, Boolean, bag-of-words and TFIDF profile is prepared. Details can be found in [12].

V. RESULTS AND DISCUSSIONS

Table 1 shows the accuracy of NB and KNN for the task of bookmark spam user detection. It can be seen that best results are achieved in case of naïve-Bayes classifier when multivariate Bernoulli distribution is fitted to the data. The other observation is that KNN classifier works best for all the representations while naïve-Bayes works best for the Boolean representation of the user vector.

Table 1: Accuracy of all three IR models considering all 71911 attributes

Classifier	Accuracy for Different IR Models		
	Boolean	Word Count	TFIDF
Naïve Bayes	97.544	78.297	84.395
KNN	97.423	96.931	96.984

VI. CONCLUSIONS AND FUTURE WORK

This paper focuses on the task of detecting spam users in social bookmarking system. The problem is modelled as classification task. Naïve Bayes and K-Nearest Neighbor classifiers are used as the classification techniques. Experiments are carried out with three IR models (Boolean, bag-of-words and TFIDF).It is evident from the results that KNN classifier works well for all the three IR models while naïve-Bayes works the best for Boolean representation of the users.

The future work would be to work on combining the interrelationship between tags with feature selection for better classification performance.

REFERENCES

[1] Benjamin Markines, Ciro Cattuto, Filippo Menczer, “Social Spam Detection”, Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web, pp. 41-48, 2009.

[2] Stephan Doerfel, Andreas Hotho, Robert Jaschke, Folke Mitzlaff, Juergen Mueller, “ECML PKDD Discovery Challenge”, 2008.

[3] Toine Bogers, Antal Van Den Bosch, "Using Language Modelling for Spam Detection in Social Reference Manager Websites", Proceedings of the 9th Belgian-Dutch Information Retrieval Workshop, pp. 87-94, 2009.

[4] Steven Young, Itamar Arel, Thomas P Karnowski, Derek Rose, "A Fast and Stable Incremental Clustering Algorithm", Information Technology: New Generations (ITNG), pp. 204-209, 2010.

[5] AnestisGkanogiannis and Theodore Kalamboukis, "A Novel Supervised Learning Algorithm and its use for Spam Detection in Social Bookmarking Systems", ECML PKDD Discovery Challenge, 2008.

[6] Chanju Kim and Kyu-Baek Hwang, "Naive Bayes Classifier Learning with Feature Selection for Spam Detection in Social Bookmarking", ECML PKDD Discovery Challenge, 2008.

[7] Ralf Krestel and Ling Chen, "Using Co-occurrence of Tags and Resources to Identify Spammers", ECML/PKDD Discovery Challenge Workshop, pp. 38-46, 2008.

[8] Antonia Kyriakopoulou and Theodore Kalamboukis, "Combining Clustering with Classification for Spam Detection in Social Bookmarking Systems",ECML/PKDD Discovery Challenge Workshop, pp. 47-54,2008.

[9] Beate Krause, Christoph Schmitz, Andreas Hotho, Gerd Stumme, "The Anti-Social Tagger – Detecting Spam in Social Bookmarking Systems", Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web", pp. 61-68, 2008.

[10] Amgad Madkour, Tarek Hefni, Ahmed Hefny, Khaled S Refaat, "Using Semantic Features to Detect Spamming in Social Bookmarking Systems", ECML PKDD Discovery Challenge Workshop, 2008.

[11] J. P. Jiawei Han, Micheline Kamber, "Data Mining Concepts and Techniques". Morgan Kaufmann, 3 Edition, July 2011.

[12] Zdravko Markov, Daniel T Larose, "Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage", John Wiley & Sons, 2007.

[13] Radhika Kotecha, Vijay Ukani, Sanjay Garg, "An empirical analysis of multiclass classification techniques in data mining", NUiCONE 2011, pp. 1-5, 2011.

[14] Rutu Joshi and Priyank Thakkar, "Experimental Evaluation of Different Classification Techniques for Web Page Classification", International Journal of Advanced Research in Engineering & Technology (IJARET), Volume 5, Issue 5, 2014, pp. 91 - 101, ISSN Print: 0976-6480, ISSN Online: 0976-6499.

[15] Priyank Thakkar, Samir Kariya and K Kotecha, "Web Page Clustering using Cemetery Organization Behavior of Ants", International Journal of Advanced Research in Engineering & Technology (IJARET), Volume 5, Issue 1, 2014, pp. 7 - 17, ISSN Print: 0976-6480, ISSN Online: 0976-6499.

[16] Prajakta Ozarkar and Dr. Manasi Patwardhan, "Efficient Spam Classification by Appropriate Feature Selection", International Journal of Computer Engineering & Technology (IJCET), Volume 4, Issue 3, 2013, pp. 123 - 139, ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375.

[17] C.R. Cyril Anthoni and Dr. A. Christy, "Integration of Feature Sets with Machine Learning Techniques for Spam Filtering", International Journal of Computer Engineering & Technology (IJCET), Volume 2, Issue 1, 2011, pp. 47 - 52, ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375.

[18] R. Manickam, D. Boominath and V. Bhuvaneswari, "An Analysis of Data Mining: Past, Present and Future", International Journal of Computer Engineering & Technology (IJCET), Volume 3, Issue 1, 2012, pp. 1 - 9, ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375.