Journal of Physics: Conference Series

Spam Filter Based on Naive Bayesian Classifier

Spam Filter Based on Naive Bayesian Classifier

Teng Lv[1], Ping Yan[2,] *, Hongwu Yuan[1] and Weimin He[3]

[1]School of Information Engineering, Anhui Xinhua University, Hefei 230088, P. R. China
[2]School of Science, Anhui Agricultural University, Hefei 230036, P. R. China
[3]Department of Computing and New Media Technologies, University of Wisconsin-Stevens Point, 2100 Main Street, Stevens Point, WI 54481, USA

Email: LT0410@163.com, want2fly2002@163.com(*corresponding), yuanhongwu@axhu.edu.cn, whe@uwsp.edu

Abstract.. Spam now accounts for over 70% of all emails and the harm to users is increasing, such as waste a lot of network bandwidth to transfer and space to store, has large quantity, repetitive, fraud, and unhealthy content, etc. As spam is usually embedded in normal e-mails, it is difficult to identify them. This paper analyzed the main technologies to identify and block spam, such as information or content filtering technology, blacklist and white list technology, intention and behaviour analysis technology, etc. A model to determine an e-mail is a spam or not based on naive Bayesian classifier is presented in the paper. The test result shows that the model is effective.

1. Introduction

Spam generally refers to unsolicited e-mail and e-mail that the recipient cannot reject, such as mail from the email address blacklisted by the addressee, the subject line or content contains wrong, misleading or false information, and use, relay or send mail through the Internet device of a third party without consent [1], etc. Spam emails now account for over 70% of all email messages [2]. Research of spam has gained interests of research community for the last several decades [3]. Spam senders use spam and other related technologies to embed spam in normal e-mails, so the form of spam is various, and the harm to users is increasing [4].

Related work. At present, anti spam technologies include image recognition, intention analysis, sender feature recognition, AI-based methods, and other related technologies. The details are as follows:

(1) Information filtering technology, also known as content filtering technology, is a technology used to block and deny access to annoying information in e-mails. Ref.[5] used a Naive Bayesian classifier to build up the content features filter. Ref.[6] proposed an anti-image spam technique that

uses image file size information to distinguish between image spam and legitimate e-mails.

(2) Blacklist and white list technology. The workflow of a spam filtering system with blacklist and white list is that all traffic flow from a white list is automatically passed while that from a blacklist is automatically blocked [7].

(3) Intention or behavior analysis technology. We can monitor and analyze the data in e-mails to set up a set of intention or behavior features of spam by comparing with normal e-mails, so as to determine which category the received e-mail belongs to[8,9].

Organizations of the work. The following is organized as follows. Section 2 introduces the basic concepts used in the paper, such as Bayes theorem and Naive Bayesian classifier. Section 3 gives a filter model based on Naive Bayes classifier. Section 4 concludes the paper and point out the future directions of the work.

2. Conditional Probability and Bayes Theorem

Suppose (Ω, F, P) is a probability space, where Ω is the sample space of experiment F, and event $B \in F$. if P(B)>0, then for any event $\forall A \in F$, we have the following conditional probability formula:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

where P(A|B) is the conditional probability of A after the occurrence of B, P(B|A) is the conditional probability of B after the occurrence of A, P(A) is the prior probability of A, and P(B) is the prior probability of B.

Bayes Theorem is a kind of inverse operation of conditional probability. We can infer the probability of new events according to the existing probability. In general, the probability of event E under event F is different from that of event F under event E. However, there is a relationship between them, and Bayesian Theorem is used to show the relationship as follows:

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A)}$$

where $B_i \cap B_j = \Phi$ ($i \neq j, i, j = 1,2,\cdots,n$) and $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$. $P(B_i \mid A)$ is the conditional probability of B_i after the occurrence of A, also known as the posterior probability due to the value obtained from A, $P(A \mid B_i)$ is the conditional probability of A after the occurrence of B_i, also known as the posterior probability of A due to the value obtained from B_i, $P(B_i)$ is the prior probability of B because it doesn't take into account any aspect A, and P(A) is the prior probability of A because it doesn't take into account any aspect B. In fact, P(A) can be calculated by Total Probability Formula the as following:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$$

.

3. Spam Filtering Model Based on Naive Bayesian Classifier

A classifier is a kind of software or hardware device that assigns a category name to each classification mode. Bayes classifier is a classifier designed according to Bayes Theorem and is the most basic statistical classification method [10]. Naive Bayesian classifier [11] is a classification method based on Bayes Theorem and independent hypothesis of characteristic conditions.

Naive Bayesian classification algorithm is widely used in spam filtering. The formal process of Naive Bayes classifier is following:

1. Suppose an e-mail can be represented as a vector $d = \{w_1, w_2, \cdots, w_n\}$ and each e-mail can be classified according to a class space $c = \{c_1, c_2\}$, where $w_1, w_2, \cdots, w_n$ is feature attributes of *d* and c_1, c_2 is class categories of *c* indicates an e-mail is a spam or not.
2. Compute the probability of vector *d* belonging to each category $c_j (j = 1,2)$ by Bayesian Theorem

$$P(c_j \mid d) = \frac{P(d \mid c_j)P(c_j)}{P(d)}),$$

where $P(d \mid c_j) = P(w_1, w_2, .., w_n \mid c_j) = \prod_{i=1}^{n} P(w_i \mid c_j)$ as Naive Bayes hypothesis suppose that occurrence of all attributes of a text are independent to each other, and $P(d) = P(d \mid c_1)P(c_1) + P(d \mid c_2)P(c_2)$ by Total Probability Formula.

3. Let $d \in c_k$ if $P(c_k \mid d) = \max\{P(c_1 \mid d), P(c_2 \mid d)\}$.

Ling-spam is used to verify the proposed Naive Bayesian classification algorithm in the paper. Ling-spam has two datasets: a test set with 260 e-mails and a training set with 702 e-mails in which spam and non-spam account for 50% respectively. The file name of e-mail including string “spmsg” indicates that the e-mail is a spam. We use spam probability 0.5 as the threshold to determine whether an e-mail is a spam or not.

Table 1 is the test result of the model, where diagonal elements represent the number of correct classification, while non-diagonal elements represent the wrong classification. From Table 1, we can see that the model has good classification effect. In ideal environment, Naive Bayes model has the minimum error probability compared with other classification models and the classification effect of naive Bayesian classifier will be good when the correlation between attributes is independent, so it is often used to filter spam in reality.

Table 1. Test results of the model

	spam	ham
spam	129	1
ham	0	130

4. Conclusions

This paper analyzed the main technologies to identify and block spam and presented a model to determine an e-mail is a spam or not based on Naive Bayesian classifier. The experiment results show that the model is effective in Ling-spam dataset. Compared with other classification methods, Naive Bayesian model has the minimum error rate in theory, but it is not practical that the attributes assumed by Naive Bayes model are independent from each other in real situation, which has a certain impact on the correct classification of naive Bayesian model. Another problem is that the prior probability comes from the assumption, so the prediction result may be different in different prior probability assumption.

5. Acknowledgments

The work is supported by Science Research Team of Anhui Xinhua University (No.kytd201902), Anhui Province Quality Engineering (No.2019jxtd119), and Introduction of Talents Foundation of Anhui Xinhua University(No.2015kyqd002).

6. References

[1] Tencent mail service. 2019 What is SPAM mail. https://service.mail.qq.com/cgi-bin/help?subtype=1&id=16&no=82.
[2] SENSATA. 2019 Email Security. http://www.sensata.co.uk/email-spam/.
[3] Cranor L. F. and LaMacchia B. A. 1998 Spam. Communications of ACM, 41(8): 74-83.
[4] Bernik J. 2007 The harm behind spam. Bank Technology News(September).
[5] Yu Y. and Chen Y. 2012 A novel content based and social network aided online spam short message filter. Intelligent Control & Automation. IEEE.
[6] Uemura M. and Tabata T. 2008 Design and Evaluation of a Bayesian-filter-based Image Spam Filtering Method. International Conference on Information Security & Assurance. IEEE Computer Society.
[7] Cai Y., Qutub S. S, and Sharma A. 2006 Spam white list. http://www.freepatentsonline.com/y2006/0168033.html
[8] Mao C. H., Lee H. M., and Yeh C. F. 2011 Adaptive e-mails intention finding system based on words social networks. Journal of Network and Computer Applications, 34(5), 1615-1622.
[9] Yang Y. U. and Yu. C. 2013 Analysis and application of social behavior in offline spam message filter. Journal of Chinese Computer Systems, 34(8), 1877-1881.
[10] Stanlee N. and Patil. A. 2018 Mitigating Spam Emails Menace Using Hybrid Spam Filtering Approach. International Conference on Emerging Research in Computing, Information, Communication and Applications. Springer, Singapore.
[11] Az-Zahra H. M. 2017 Spam detection framework for Android Twitter application using Naïve Bayes and K-Nearest Neighbor classifiers. International Conference on Software & Computer Applications. ACM, USA.