The Spam Email Filter Based on the Bayes Method

Zihao Zhou *

Department of Foreign Language, Beijing Institude of Technology, Beijing, China

* Corresponding Author Email: 1120230124@bit.edu.cn

Abstract. There are many spammers send a large amount of spam information to the public. The situation that spammers send spam email is common. The research from AA science have asked 55852 people, all of them claimed that spam email is becoming one part of their live. Nowadays they become used to open their mailboxes and delete the spam email. To deal with this situation people provide some filters to help people indicate which is spam email. Bayes method is one of earliest method, and up to now it still has its importance. This paper introduces the Bayes method, about the Bayesian theory and how it works in the filter to deal with the spam email. And then to analyze the outcomes, about the data and its accuracy. In this paper there also includes some small details that should be mentioned and the advantages and disadvantages of the method.

Keywords: Bayesian; Datasets; Spam email; Filter.

1. Introduction

Now-a-days, communication through email has become one of the cheapest and easy ways for the official and business users due to easy availability of internet access [1]. At the same time spam emails frequently appear in people's lives. Many classification techniques used for identifying spam emails, treat spam filtering as a binary classification problem. That is, the incoming email is either spam or non-spam [2]. Almost all the people have received the spam email. Spammers always send the spam email, that may bring a lot of trouble. Spam can be divided into different types depending on how it is sent and the type of information it contains [3]. Spam mails are also said to be Unsolicited Bulk Emails and Its another part is Unsolicited Commercial Emails. Hacker. Phishers and malicious attackers are frequently using email services to send false kinds of messages by which target user can loss their money and social reputations [4]. For the people who get spam email, their computers would be affected by the virus and they may not be able to open their computers, and they may open some spam website and spill their information without realization. And for people get the advertises, some of them may believe the advertisements and spend their money on it, some of them would be disturbed by it. In all it would have some harmful effect on people's lives. To avoid such effect people shouldn't open the spam emails and they shouldn't spill their ID and even any of their secret code on the website.

Spam is considered an invasion of privacy [5]. As the research shows that at the beginning of 2022, up to the end of 2021, the spam emails have caused an 18.84 billion loss of RMB in China, that is a great loss. On average, each Chinese have received 330 spam email in the past 2021, the sum of the spam email is about 69.4 billion. What should mention was that this number is only 50.0 in 2020, so it has increased by 38.8% at that year. That is really a horrible number. In addition, according to another research, every 3 second a new spam email company would be established. That caused a large amount of trouble in people's daily lives.

Once you do train the filter, no more training is needed as each new document classified additionally trains the filter by simply being classified [6]. Filters of this type have so far been based mostly on manually constructed keyword patterns [7]. The filter would check the text of the email, and the topic of the email. And next is to get the probability of the words and get the main topic of the text. The Bayes method is one of the original methods that identify the spam email, and nowadays it still has effect. It is accurate and precise, and it can adapt to the new data, and it can be trader made to the users.

On the next part, this paper first tells the theory of the experiment and the method, of why people should use such method and how it works, the basic theory of it and the formulas. Then the next part

is to get the code and the data, introduce the code and the step, and then use the data to do the experiment. And the next step is to get the outcome and analyze the data. From the data, there are the conclusions, then the last step is to analyze the conclusions and get the advantages and disadvantages of this method.

2. Method

When a compound used in spam email is separated into single words, it becomes words that are usually used in both spam and ham emails. This leads to the difficulty for the system to filter spam emails [8]. There isn't any mathematical structure in the email, so if people want to use algorism to deal with the email, they must try to use some methods to turn the email into the data. After pre-processing, a vector space model should be made, in order to turn it into a space model. Some feature words, feature phrases, or other attribute features (e. g. special symbols, etc.) were used as the feature Item, and the feature item set was defined as T ={t1,..., tm}. Then each message text can be represented as a vector using this method. The key ingredients to a Bayesian analysis are the likelihood function, which reflects information about the parameters contained in the data, and the prior distribution, which quantifies what is known about the parameters before observing data [9]. And as the Bayesian theory claim that the probability of the text belonging to one kind is

$$\frac{P(C|X) = P(C) * P(X|C)}{P(X)} \quad (1)$$

As the P(X) doesn't rely on the kind of the variables, the only thing should be done is to find the maximum P(C)*P(X|C), and the X belongs to it. Then a threshold of the data is the thing needed, for example, a threshold y, and for most of the situation, the y is 0.5. The next step is to define spam email as C1, and not spam email as C2, then put them in the formula.

$$\frac{P(C1) * P(X|C1)}{P(C1) * P(X|C1) + P(C2) * P(X|C2)} \quad (2)$$

And then calculate it, get the solution, if it is greater than y, then it is spam email. There are several methods for people to get the solution of P(X|C1), as there are several Bayes methods, but they all obey assumption, which is to assume that each feature variable X is independent under a given category variable.

Some existing problems are regarding accuracy for email spam filtering that might introduce some error [10]. It may not be entirely accurate to use algorism to deal with the words and email, because it is not the same kind of information and the algorism may not be able to identify the real meaning of the words. So, for different methods, the accuracy and error of the outcome should be considered.

3. Experimental results

The experiment shows the accuracy, error, precision and recall of the outcome, which tells the advantages of the Bayes method on the accuracy.

It should be mentioned that in the experiment the proportion of the training and test dataset is 4:1,but in some situation, it may make the conclusion over fitting, that may make it difficult to classify the test dataset, but in another situation if the proportion is too low, the degree of classifier learning may not achieve the desired effect, and for this situation, if the amount of the special dataset is too low, and if people analyze the large volumes of mail, the tilter may lose its effect. So, the proportion of the two sets should be considered and make some progresses.

The first step is to get some data, several emails, and put all of them in the CSV file and let it read from the python. And what should be done is to read the data from the CSV file and stored in the data variable.

Then the next step is to deal with the data, divide the data into two parts as two samples of the experiment. The data was disrupted, that is to improve the rigor of the experiment. Then the disrupted data is divided into two parts, training set and test set, the proportion is 4:1.

Then Bayes model should be turned into use. The first step was to fit the model. The training set includes a large amount of the data and they are used to finish this step. The next step is to make prediction. The test set, including a small amount of the data, can be used for it. The output part was defined as y_hat, from the output part, as the conclusion of the prediction. Then in order to get the accuracy, method of accuracy was used, to calculate it, and output the accuracy of the model.

Then it comes to the last step, integration model. The integration of the model was initialized. The next part is to ensemble, and specified the thresholds given. Then, the initial data that has not yet been separated was used for fitting. That is all the steps. In this experiments, 60 samples were tested and the predicted results were all stored in y_hat, the conclusion of the prediction.

This diagram shows the outcome of the experiment.

Table1. Performance measurement

Bayesian Classifier	Accuracy (%)	Error (%)	precision	recall
Dataset1	94.02	5.98	0.93	0.81
Dataset2	95.92	4.08	0.95	0.92

In Table 1 performance measurement can be seen that the three datasets get almost the similar outcome. The accuracy is all about 95% while the error is all about only 5%, The presision is from 0.93-0.95. What is different is the recall rate, The two dataset get the different recall rate. In all the outcome is precise.

This method has many advantages. For many methods, they may draw the conclusion by analyzing some of the words, so the error maybe pretty much. But for this method, as can be seen in Table 1, the rate of error is only about 5% and the rate of accuracy is about 95%, based on the test. That is because the Bayes method analyzes each word of the email, that makes the filter works more accurately and leave less leak for people who send the spam email to escape from the detect of the filter.

Also, this method can adapt to every new spam emails. It means that every time the filter identifies the new spam, the new data would be added to the filter, and the next time it calculates, it would be adapted to it rapidly. This method can also be trader made to the users. That is because the filter entirely depends on the user's data, and classified the data into spam emails and not-spam emails before training the model. Also, people can make processes on the choose feathers, to make it adapt to the data. However, each coin has two sides, that means that people should provide their training data to the models anytime they use this method, which causes a lot of time and made a lot of troubles. That makes it difficult to train it.

There are some designed words when people use this method, that also makes some bad guys use some special method, design some words and send the spam email. As those words are special, it would be identified as the not-spam emails, and that may cause something wrong. People who send spam email would find the words frequently appears in the not-spam email. Well, the filter can adapt to it as time goes by, but it takes time to learn, that is also one thing to be considered.

Also, the images would influence the method, unless people have specially designed to made it identify the images. It is hard for the filter to identify the images whether it is spam email or not, so in many situations it may be ignored.

4. Conclusion

In conclusion, the Bayes method is an accurate and precise method to deal with the spam email. It can turn the emails into the data, and divide them into two parts. Then it uses one part to be the training part, and it get the data, and learn how to deal with the data, then it uses the test part to do

the test and get the conclusion. However, the method is accurate as the accuracy of the conclusion is about 95% and the rate of error is only about 5%. Also, it can adapt to the new data, as the new data is added into the filter and it can learn from the new data and it would adapt to the new method. This would also adapt to the users as they can give the new data and the filter would depend entirely on the data. However, it may cause some troubles, as people should provide their training models every time they use it, and bad guys may find the leak from the filter and sent the spam email without being found by the filter. And it is hard for this method to identify the images from the spam emails. The proportion of the training data and the test data should also be considered.

References

[1] Agarwal K, Kumar T. Email spam detection using integrated approach of Naïve Bayes and particle swarm optimization. Second International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2018: 685-690.

[2] Zhou B, Yao Y, Luo J. A three-way decision approach to email spam filtering. Advances in Artificial Intelligence: 23rd Canadian Conference on Artificial Intelligence, Canadian AI 2010, Ottawa, Canada, May 31–June 2, 2010. Proceedings 23. Springer Berlin Heidelberg, 2010: 28-39.

[3] Sharabov M, Tsochev G, Gancheva V, et al. Filtering and Detection of Real-Time Spam Mail Based on a Bayesian Approach in University Networks. Electronics, 2024, 13(2): 374.

[4] Rathod S B, Pattewar T M. Content based spam detection in email using Bayesian classifier. International Conference on Communications and Signal Processing (ICCSP). IEEE, 2015: 1257-1261.

[5] Abu-Nimeh S, Nappa D, Wang X, et al. Bayesian additive regression trees-based spam detection for enhanced email privacy. Third International Conference on Availability, Reliability and Security. IEEE, 2008: 1044-1051.

[6] Eberhardt J J. Bayesian spam detection. Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal, 2015, 2(1): 2.

[7] Androutsopoulos I, Koutsias J, Chandrinos K V, et al. An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. 2000: 160-167.

[8] Ebadati O M E, Ahmadzadeh F. Classification spam email with elimination of unsuitable features with hybrid of GA-naive Bayes. Journal of Information & Knowledge Management, 2019, 18(01): 1950008.

[9] Glickman M E, Van Dyk D A. Basic bayesian methods. Topics in Biostatistics, 2007: 319-338.

[10] Rusland N F, Wahid N, Kasim S, et al. Analysis of Naïve Bayes algorithm for email spam filtering across multiple datasets. conference series: materials science and engineering. IOP Publishing, 2017, 226(1): 012091.