New method for spam-filtering

Sass Bálint

HAS Research Institute for Linguistics, H-1068 Budapest, Benczúr u. 33.
joker@nytud.hu

Keywords spam-filtering, document classification, Naïve Bayesian Classifier

Unsolicited emails (*spams*) are becoming one of the most important problems of the internet. One main method for spam is filtering, when incoming mails are divided into two parts: emails are marked as spam or as legitimate on the basis of the content. Thus spam-filtering can be considered as a document classification problem. The so-called Naïve Bayesian Classifier is on of the good document classification methods: the language model is built on the basis of examples of each category (learning-corpus), and then using this model it is determined which category the given document belongs to. The language model consists of the word frequency lists of each category.

NBC is the basis of *Paul Graham*'s spam-filtering method, which was published in 2002 [2]. It considers that spam-filtering is asymmetric: it is not a big trouble if we get one spam, but losing a legitimate email can be a misery.

This method has many advantages: (1) very good filtering perfomance, (2) filter-creation from spam and legitimate corpora is automatic, (3) it can be retrained from time to time, thus it can adapt itself, (4) giving learning-corpora, you can define what means spam for you.

I implemented this method and tested it on my incoming mails in the last six months. Precision was 98.6% and recall was 94.1%.

It is clear that in this case the linguistic processing means only tokenization of emails and creation of word-frequency lists. It was tried to lemmatise text or remove most frequent words, but it did not result in substantial improvement of performance [1]. It seems, that in such realtively simple document classification tasks little linguistic processing can be enough. The algorithm is language-independent, therefore it can be used to filter emails written in any language.

References

1. Androutsopoulos, I. et al.: An Evaluation of Naïve Bayesian Anti-Spam Filtering. In proceedings of the 11th European Conference on Machine Learning. Workshop on Machine Learning in the New Information Age. (2000) 9–17
http://arxiv.org/PS_cache/cs/pdf/0006/0006013.pdf
2. Graham, P.: A Plan for Spam. (2002)
http://www.paulgraham.com/spam.html