Abstract

E-mail is the most prevalent methods for correspondence because of its availability, quick message exchange and low sending cost. Spam mail appears as a serious issue influencing this application today's internet. Spam may contain suspicious URL's, or may ask for financial information as money exchange information or credit card details.

Here comes the scope of filtering spam from legitimate e-mails. Classification is a way to get rid of those spam messages. Various researches are proposed for spam filtering by classifying them into labels of spam and business messages.

Bayesian classification based spam filtering technique is a popular method. Also SVM based classifications are also used. K-nearest neighbour classification is simple, straightforward and easy to implement and has high F-measure compare to Bayesian and SVM classification. But accuracy of traditional KNN is lower than Bayesian classification.
In this work a detection of spam mail is proposed by using K-nearest neighbour classification method by combining Spearman's correlation coefficient as distance measure rather than traditional Euclidean distance. Experimental results present a significant improvement in accuracy with higher F-measure compare to traditional algorithms.

References

4. MEI paper on Spearman’s rank correlation coefficient. December 2007. "Spearman’s rank correlation"
5. Volume 14, Supplement 1, August 2015 "Privacy-preserving email forensics"
6. T. Pranav Bhat, C. Karthik A Privacy Preserved Data Mining Approach Based on k-Partite Graph Theor Volume 54, 2015
7. Volume 27, Issue 1, January 2015" Clustering and classification of email contents"
8. Ommera jan ,heena khana “An analysis of misclassification error detection in mails using data mining techniques”MAY 2015
9. Tarjini vyas ,payal prajapati “A survey and evaluation of supervised machine learning techniques for spam E-mail filtering” 978-1-4799-608S-9/1S/$31.00©201S IEEE
10. Mehran sahani ,susan dumais “A Bayesian approach to filtering junk E-mail”
13. Dr.sanjeev dhawan, jyoti verma “social networking spam detection using R package and k-nearest neighbor classification”www.iasir.net
15. Tao ban “An online malicious spam mail detection system using resource allocating network with locality sensitive hashing” Received 25 February 2015; accepted 20 April 2015; published 22 April 2015

Index Terms

Computer Science Software Engineering
Keywords

Bayesian classification, SVM Classification, spam, Email, KNN classification, Spearman correlation, Spam Filtering, Accuracy, F-measure.