

MEASURING PERFORMANCE OF SPAM FILTER USING GRAHAM'S NAIVE BAYES CLASSIFIER

Wildan Rahman, Julio Adisantoso

Department of Computer Science
Faculty of Mathematics and Natural Sciences
Bogor Agricultural University
julioipb@gmail.com

2010

Abstract

Email spam has become a major problem for internet users and providers. After several failed attempt to filter spam based on heuristic approach such as black-listing or rule-based filtering, content-based filtering using naive Bayes classifier has become the standard for spam filtering today. However, the naive Bayes classifier exists in different forms. This research aims to compare two different forms of naive Bayes which are multinomial naive Bayes using boolean attribute and Graham version of naive Bayes which is popular among several commercial and open source spam filter applications. This research also compares performance of two different methods for data trainings which are train-everything (TEFT) and Train-on-Error (TOE). Finally, this research attempts to identify several hard-to-classify emails. The evaluation result showed that multinomial naive Bayes had better performances compared to Graham naive Bayes. The result also showed that TEFT successfully outperforms TOE in term of accuracy.