Abstract

Electronic mail is one of the important means of communication. Thus, this useful tool has invaded by invaders for different purposes. One such Invasion is the posting of useless, unwanted e-mails known as spam or junk e-mails. Several methods of spam detection exist, but each has certain weaknesses. This paper address these weaknesses by implementing and describing a spam detection system in text classification mode, which uses Bayesian method vs. PCA to filter out written spam mails from the user’s mail box. In the proposed method first extract all tokens that exist in body of emails for classifying emails based on them. But sum of these tokens aren’t useful. Sum of them are repeated in two categories spam and non-spam mails equally, so they aren’t appropriate for distinguishing two types of emails. So proposed method finds best tokens as main features using feature selection methods such as genetic
algorithm (GA), forward and backward feature selection methods.

References


5. T. Liang, Y. Pi, "On Spam Detection Based on Cognitive Pattern Recognition", International Conference on Computational Intelligence and Security Workshops, 2007


Index Terms

Computer Science  Artificial Intelligence

Keywords

Spam, Electronic Emails, Genetic Algorithms, Text Classification, Forward, backward, feature selection, Naïve Bayesian