Abtract

Many algorithms can be used to overcome the problem of spam email, such as Naive Bayes. However, using only a single classifier then filtering system can be easily attacked by spammers. Therefore, Symbiotic Naive Bayes algorithm is applied using local filters of different users to improve the performance of filtering on a personal level.

Do some testing to see the effect of word weighting mechanism, the ratio of training data and testing data, and the number of datasets to the filtering system performance as measured by the value of precision, recall, and f - measure. These three things are done to find a better weighting mechanism, the ratio condition training data and testing data such as what, and how the ideal number of datasets so Symbiotic Naive Bayes can produce high performance.

Based on the test results can be concluded that the filtering system that uses data transformation term results of the weighting mechanism produces a better performance compared with term frequency. At the time of the training data ratio larger than the data testing the filtering system will result in better performance. And increasing the number of datasets can improve filtering system performance as well.

Keywords: Naive Bayes, Symbiotic Naive Bayes, Classification, Dataset, Data Training, Data Testing, Term Frequency, Term Transformation, Ratio Data Training and Data Testing, Precision, Recall, F - measure