

Abstract

Email characteristics of low cost, high-speed, and convenient impact on the unappropriate use to send messages to a large number of people. These messages are called unsolicited bulk email, junk mail, or spam. Because of this spam, many users who have email or ISP gain much financial loss and also the development of spam to email traffic is increases every year. Therefore, users need to develop automatic classifier that can distinguish between spam and non-spam (ham). In this Final task which the topics are “ Analyse and implementation Personalized Spam Filtering Using Artificial Neural Network ” Data Mining is used to decide whether the e-mail is a spam or a ham and as the case study ECML-PKDD 2006 Discovery Challenge Data Mining Competition with e-mail that has been encoded as bag-of-word vector space so we don't know the content of that e-mail, and using information gain as data preprocessing for the feature selection and Artificial Neural Network (ANN) as the text classification and genetic algorithm for training ANN's weights.

Keywords: *spam filtering, text classification, data preprocessing, information gain, Artificial Neural Network, Genetic Algorithm.*