Abstract

People spend an increasing amount of time for reading email and deciding whether they are spam or non-spam. Some users spend additional time to label their received spam email for training local spam filters running on their computers. But in the other side email service providers want to relieve users from this burden by installing server-based spam filters in its server. Which this spam filters that has server-based characteristic can't use labeled email from individual user, but from on publicly available sources, such as newsgroup messages or emails received through spam traps. While from that individual email user has each different characteristic when they decide an email is a spam or non-spam. In this final exam, data mining used to decide what if an email which is received by a user is a spam or non-spam and use as a case study for ECML PKDD2006 Discovery Challenge Data Mining Competition with email data that encode into bag-of-word vector space then we doesn't know about the truth about that email. And with Evolving Fuzzy Classifier method used for classification to decide the type of those email with Genetic Algorithm as an evolution for existing Fuzzy Classifier.

Keywords: *data mining, classification, evolving fuzzy classifier, spam, genetic algorithm, non-spam, email, bag-of-word vector space, fuzzy classifier.*