



SURVEY ARTICLE

Survey of Spam Filtering Techniques and Tools, and MapReduce with SVM

Amol G. Kakade¹, Prashant K. Kharat², Anil Kumar Gupta³

¹Department of Information Technology, Walchand College of Engineering, Sangli, India

²Department of Information Technology, Walchand College of Engineering, Sangli, India

³C-DAC, Pune, India

¹ agkakade@gmail.com; ² prashant.kharat@walchandsangli.ac.in; ³ anilg@cdac.in

Abstract

Spam is unsolicited, junk email with variety of shapes and forms. To filter spam, various techniques are used. Techniques like Naïve Bayesian Classifier, Support Vector Machine (SVM) etc. are often used. Also, a number of tools for spam filtering either paid or free are available. Amongst all techniques SVM is mostly used. SVM is computationally intensive and for training it can't work with large a dataset, these cons can be minimized by introducing MapReduce framework for SVM. MapReduce framework can work in parallel with input dataset file chunks to train SVM for time reduction. This paper aims at surveying of few such techniques and popular spam filtering tools with scope to introduce MapReduce with SVM.

Keywords - Spam filtering techniques; Naïve Bayes classifier; Support Vector Machine (SVM); Neural Networks (NN); K-Nearest Neighbor (KNN); spam filtering tools; MapReduce