

A Simple Machine Learning Approach for Identifying Promotional Short Message Service (SMS) Messages

D.S. Dias (dulandias@outlook.com)¹, N.G.J. Dias²

¹ Department of Computer Science, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka.

² Department of Computer Systems Engineering, Faculty of Computing and Technology, University of Kelaniya, Sri Lanka.

Abstract

Mobile phones play an integral part in the modern lives of humans. Short Message Services (SMS) Messages have become a popular mode for simple communication. Its' simplicity, cost-effectiveness and large audience has attracted the attention of advertising industry to send targeted promotional messages to mobile phones. In Sri Lanka, a survey conducted in Colombo, yielded that 3 out of 5 SMS messages received our promotional messages. Even though extensive research has been carried out in detecting junk SMS messages, the amount of research conducted on filtering promotional SMS messages is rare.

The purpose of this research is to evaluate the success and accuracy of utilizing a simple machine learning algorithm to identify promotional SMS messages. Here, we have used a feed-forward neural network based on a statistical model, which was trained with a training data set consisting of promotional as well as non-promotional messages. Each test message was broken down in to individual words and filtered through by cleaning to form keywords which will have consist of a weight and probability value. With each message that is used to train, these values will be updated according to whether it is a promotional or a non-promotional message. When a message is tested through this neural network, the words of the message will be matched against the keyword's weight and probability, which will then calculate a resultant probability. By setting a par-value, we can classify the test as a promotional or a non-promotional message.

The proposed model yielded a 100% accuracy when tested within the given test data set. In order to get successful results for broader test data sets, the model has to be trained comprehensively with proper amount of promotional and non-promotional messages. Optionally, the results obtained from the feed forward neural network for incoming messages, can then be fed back in to the feed forward neural network for further training.

As future work, we intend to take this solution to an android-based mobile application that extracts promotional messages from the incoming SMS messages as well as from a server, and display them to the user based on his preferences.

Keywords: *SMS filtering, Smart Phones, Machine Learning, Promotional, Marketing, Neural Network, Statistical Model*