

# Artificial Intelligence and Its Application: Assignment

## A Classification to Spam

Name: Yuchen Song

Student ID: 201830360498

### Introduction

Nowadays, As the technology developing, convenient email takes the place of traditional handwriting mails and plays a more and more significant role in our life. However, the spams are quite annoying. Although there are quite a lot of filters to spams, the spams have become trickier to pass through the detective method at the same time (like change the order of a word or deliberately misspell a word), which makes it harder to detect.

In this paper, however, I aim to do the basic work to detect the spams, i.e. develop a classifier to spam based on the common features of spams.

### Method

I used the spam database (1999) from UCI machine learning repository at <http://archive.ics.uci.edu/ml/datasets/Spambase>. It is quite old, but still make sense to most characters on how to detect a spam.

The data set contains 4601 instances (with 1813 spam instances), and there are 58 columns of attributes, including 57 features of spams and 1 label. The 57 features consist of 48 features on frequent words, 6 on frequent characters(marks), and 3 on capital letters.

I developed four simple classifiers (DT, Bayes, MLP and KNN) and four ensemble classifiers (Random Forest, voting, boosting and bagging) to spams based on the database, comparing the accuracy and choosing the most suitable classifier.

### Experiment

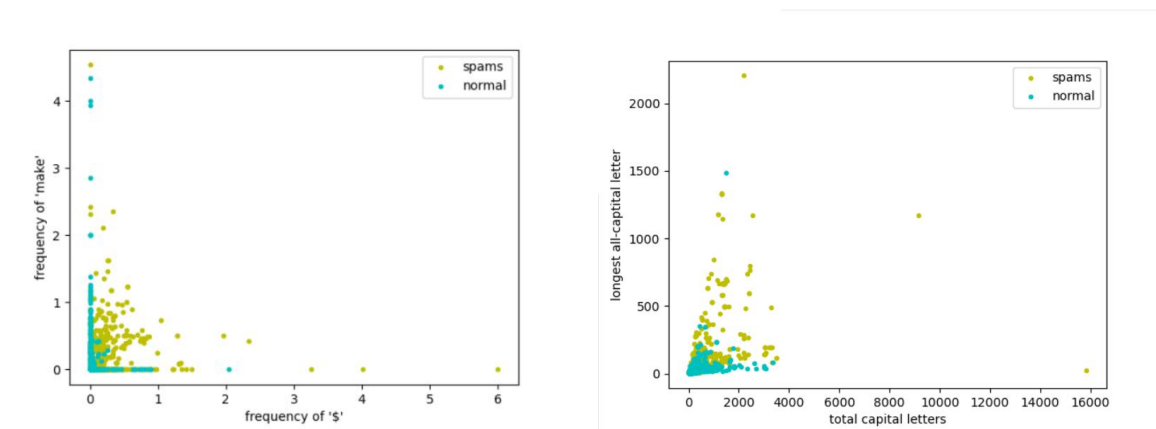
I divided the data set into two parts, training set (2301, even-numbered sets) and testing set (2300, odd-numbered sets). Each part contains about the same amount of the data (~40% spam).

At first, I decided to use one single classifier to solve the problem, but their testing accuracy are not even higher than 90% (with MLP highest at 0.89), so I applied ensemble method instead.

For ensemble part, the Random Forest reached the highest accuracy with 1.0 training accuracy plus 0.945 training accuracy. Moreover, for all spam labeled sets, the testing accuracy increases to 0.972, that is to say, only about 3% of spams will pass through my filter.

accuracy	K-NN	DT	Bayes	MLP	RF	voting	bagging	boosting
training	0.886	0.869	0.819	0.910	1.00	0.928	0.994	0.973
testing	0.752	0.862	0.810	0.897	0.947	0.913	0.925	0.937

On the other hand, We can see from the figure that although there is a slight border do distinguish spams using only two features, in most cases it is not clear at all. So my classifier takes in all features from the data.



## Conclusion

In this paper, I discussed how to model a classifier to spams. The result of experiment shows that Random Forest ensemble classifier is the best one to this problem, with 1.0 training accuracy plus 0.945 training accuracy.

Here are some topic I can think of for further research (out of my ability):

- Could we remove some features and have a better training model?
- How can we build up our own spam database? (Maybe use regex to capture frequent words?)
- How can we visualize multidimensional data classifier?

## Acknowledgments

Special thanks to the database from Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.