Instiúid Teicneolaíochta Cheatharlach

INSTITUTE of
TECHNOLOGY
CARLOW

At the Heart of South Leinster

Classification of incoming email for an insurance company
Alexander O'Connor
Supervisor Dr Greg Doyle
M.Sc. in Data Science

Introduction

The company receives a considerable amount of email on a variety of topics. It is important that the emails reach the appropriate person as quickly as possible. Disposition of incoming emails is a manually intensive process which requires an operator to open each email, read it and forward to the appropriate end recipient. This filtering process delays emails reaching their final destination and has the potential for error. Those staffing this role need to be trained to quickly and accurately disposition each email.

1 Research Objectives

This research will identify a method for the classification of incoming email so as to eliminate the requirement for manual disposition. This paper will also estimate the accuracy, repeatability and precision of the proposed solution.

2 Literature Review

A considerable body of work exists in the field of email classification. Particulary in relation to Spam detection. Much of the literature discusses the "Bag of Words" with stemming and leammatisation. There is consensus in the literature that the techniques showing best results are Neural Network, Support Vector Machine, Naive Bayes and K Nearest Neighbour

3 Methodology

This paper adopted the approach of Preprocessing -> Learning -> Classification. This is a supervised learning project on unstructured data. In particular the main techniques examined in the this paper are Support Vector Machine, Decision Tree, Naive Bayes and Neural Networks.

4 Technology

5 Next Steps

Most email classification systems work only with text. "Only two out of the 98 studies considered images for spam email classification" Email Classification Research Trends: Review and Open Issues 2017. However it is not uncommon for emails to contain images and scans of documents.The consideration of image analyis along with text analysis is an area of potential future research.

References

Ghulam & Shuib, Liyana & Raj, Ram & Majeed, Nahdia & Al-Garadi, Mohammed. (2017). Email Classification Research Trends: Review and Open Issues. IEEE Access. PP. 1-1. 10.1109/ACCESS.2017.2702187.

Kumar, Aman & Assistant, Sharma & Sahni, Suruchi. (2011). A Comparative Study of Classification Algorithms for Spam Email Data Analysis. International Journal on Computer Science and Engineering. 3.

Senthamarai Kannan Subramanian and N. Ramaraj, 2007. Automated Classification of Customer Emails via Association Rule Mining. Information Technology Journal, 6: 567-572. DOI: 10.3923/itj.2007.567.572

Contact Details

Alexander O'Connor | axoconnor@gmail.com | 0872310717 | C00265469@itcarlow.ie