Text Classification on News Articles

Sean Soutar STRSEA001[a], Fabio Fehr FHRFAB001[b]

[a]*UCT Statistics Honours, Cape Town, South Africa*

[b]*UCT Statistics Honours, Cape Town, South Africa*

Abstract

This proposal outlines a research objective to maximise mutliclass, single label document topic classification accuracy and computational efficiency. Text classification is an important area of research due to the dramatic increase in data generation within the last decade. This proposal provides a short overview of the current methodology used in text classification as well as some exploratory analysis on the News24 article data which will be used in our project. Finally we propose a timeline for our project and a work allocation going forward.

Keywords: Text classification, machine learning, natural language processing

JEL classification

1. References

Email addresses: `sean.soutar@gmail.com` (Sean Soutar STRSEA001), `FHRFAB001@myuct.ac.za` (Fabio Fehr FHRFAB001)