Project Plan

Student Name: Molly Hayward
Supervisor Name: Dr Noura Al-Moubayed
23/10/2019

Project title: An analysis of machine learning and deep learning techniques for the detection of sarcasm in text

I OUTLINE

A Project Description

Numerous studies have looked at machine learning, and more recently, deep learning techniques for sarcasm detection in text corpora. With applications in areas such as opinion mining and sentiment analysis, it has become a popular field of research. The aim of *this* project is to understand how various pre-processing, word-embedding and text classification techniques affect performance on this task, in order to produce a tool that can detect sarcasm with a *high degree* of accuracy.

B Preliminary Preparations

- Collect and evaluate a selection of datasets
- Produce a literature survey documenting previous research in the domain of sarcasm detection and other text classification tasks
- Identify state-of-the-art (SOTA) techniques
- Gain a high-level overview of the following python libraries: NLTK, scikit-learn, spaCy, PyTorch and TensorFlow (including the deep learning library Keras)

C Research Questions

The research questions for this project are as follows:-

- How do machine learning approaches compare to deep learning techniques for sarcasm detection?
- How can a model be used to detect the words that correlate more to sarcastic labels?
- Can the solution perform well on other datasets?

II DELIVERABLES

A Project Deliverables

Table 1: List of Basic, Intermediate and Advanced deliverables

Deliverable	Description
Basic	
Select a high-quality dataset	Compare publicly available datasets and evaluate their quality
Implement a program to clean and segment data appropriately	Research and evaluate the suitability of different text preprocessing techniques, choosing to implement the most suitable option
Experiment with machine-learning archictectures	Conduct research into various machine-learning based classifiers and implement at least one using Python
Evaluate the machine-learning based classifier on the chosen dataset	Compute the precision, recall and F1-score of the classifier on the chosen dataset
Intermediate	
Experiment with deep-learning archictectures	Conduct research into various deep-learning based classifiers and implement at least one using Python
Evaluate the deep-learning based classifier on the chosen dataset	Compute the precision, recall and F1-score of the classifier on the chosen dataset
Produce a solution	Implement a final model using the best approach as determined by the results of former tasks
Experiment with alternative datasets	Apply the solution to other datasets, evaluating its ability to generalise to unseen data
Advanced	
Implement an attention based model	Conduct research into deep learning models such as Hierarchical Attention networks that can be used to highlight attention words (i.e. words that correlate more to the sarcastic label), then implement one such model
Produce a visualization of attention words	Conduct experiments on the chosen dataset to extract attention words, then present the findings in a visual format

B Project Schedule