CS 410 Project Proposal

Group: LiveDataLab Admins

Team Members

- Yanbo Chen, ychen380
- Linfei Jing, ljing2
- Huaminghui Ding, hding14
- Captain: Yanbo Chen

Project Topic

Text classification competition
We choose the classification competition and are fully prepared to learn state-of-the-art neural network classifiers. We have learnt about some classical machine learning algorithms such as Naive Bayes, Decision Tree. We also have basic knowledge about pattern mining from text files using Apriori algorithm and FP-Growth. Our team member used to write those algorithms from scratch without using predefined libraries to classify animals in a zoo's data-set. Those will be helpful for text classification algorithm design in the project.

Programming Language

Python.

Problem to solve

Given a training set containing tweets from Twitter, each with a label "SARCASM" or "NOT_SARCASM", train a classifier to predict the label for each tweet in the test set.

Methods to use

1) Naive Bayes Classifier.
Before we dive into neural network classifiers, we decide to apply Naive Bayes Classifier first and see if it performs well on the problem.

2) Neural Network Classifiers
If Naive Bayes Classifier does not work well on predicting the label, we are prepared to learn some neural network classifiers such as LSTM and BERT.
3)Deep Learning Frameworks
We plan to use some popular frameworks such as Keras library in Tensorflow.