CS 532: Pattern Recognition Project Proposal

Members:

- Apul Jain
- Ankit Aggarwal
- Ameya Raul

Topic: Sentiment Analysis using Deep Neural Networks

Motivation:

Sentiment Analysis is an interesting application involving the combined use of NLP and Neural Networks for tasks like classifying whether an article has a positive flair or not. Using this technique we can classify several news articles to see if they contain positive content or not. Also, since this approach leverages structure of the sentences, we could use this to determine if a piece of text is composed of "factual" sentences or just vague. This could be potentially an interesting application to determine the quality of articles published by various news agencies.

Challenges:

One of the challenges we face is to tweak the traditional neural network approach to solve this problem. One of the papers we found (http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf) illustrates an approach using Recursive Neural Tensor Networks to achieve very high prediction accuracy. We plan to explore this approach to see if we could apply a similar idea to our goal.

We plan to gather test datasets for sentiment analysis from Wikipedia articles, twitter feeds, newspaper articles etc. We will then attempt to train our modified version of neural networks, to learn from the word vector representation of these sentences and classify whether they represent a positive or negative emotion.

Core concepts:

- Neural Networks
- sentence representation using word vectors
- scraping sentences from news articles or Wikipedia etc
- use of the Lasso or Tikhonov Regularization

Related Papers, Datasets:

You can find related datasets and papers on this site:
http://nlp.stanford.edu/sentiment/
Kaggle dataset and a tutorial to get started:
https://www.kaggle.com/c/word2vec-nlp-tutorial
Microsoft Research-Beijing paper:
http://www.aclweb.org/anthology/S14-2033
This is a tutorial we found online. Its not Neural net based but its a high level overview:
http://deeplearning4j.org/sentiment_analysis_word2vec.html