William Lyon
CSCI 555
Graduate Student Project Proposal
09.13.12

Project Title: Implementing An Autonomous Trading Agent Using Natural Language Processing of Financial News Headlines

Description of topic: I plan to use Natural Language Processing to analyze financial news headlines and generate buy/sell signals that can be used to place (simulated) trades in the financial markets. In finance, the efficient-market hypothesis claims that prices in the financial markets reflect all publicly available information and that prices are adjusted as new information becomes available. Daskalopoulos [4] uses a classification method to classify frequently occurring words in financial news headlines to generate buy/sell signals. For further research he suggests that filtering the word list by part of speech would yield higher accuracy. I plan to use a similar method and tag words by part of speech to factor in this information as well.

Sources:

(1) Zhang, Wenbin, et al. Financial Analysis Using News Data. Unpublished whitepaper: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.7073&rep=rep1&type=pdf

(2) Bird, Steven, et al. *Natural Language Processing With Python*. O'Reilly Media: 2009. http://nltk.org/book

(3) Python Natural Language Toolkit. http://nltk.org (python library for working with natural language processing)

(4) Daskalopoulos, V. *Stock Price Prediction From Natural Language Understanding of News Headlines.* Presented at Information Conference on Machine Learning (iCML03). http://www.cs.rutgers.edu/~mlittman/courses/ml03/iCML03/papers/daskalopoulos.pdf

(5) Sebastiani, Fabrizio. *Machine Learning in Automated Text Categorization.* ACM Computing Surveys. Vol 34:1. March 2002

(6) Peramunetilleke, D., et al. *Currency exchange rate forecasting from news headlines.* Australian Computer Science Communications. Vol 24:2 Jan-Feb 2002. pp131-139.

(7) Skiena, S., et al. *Trading Strategies To Exploit Blog and News Sentiment. Fourth Int. Conf. on Weblogs and Social Media (ICWSM 2010), Washington DC, May 23-26, 2010. http://www.cs.sunysb.edu/~skiena/lydia/blogtrading.pdf*

(8) Lavenko, V., et al. *Language Models For Financial News Recommendation.* Proceedings of the ninth international conference on Information and knowledge managment (CIKM 2000) pp389-396

(9) Barbosa, R., et al. *Algorithmic Trading Using Intelligent Agents.* 2008 International Conference on Image Processing, Computer Vision, and Pattern Recognition.

Description of program: My program will collect a sample of news headlines from a Google Finance RSS feed (for example: http://www.google.com/finance/company_news?q=NASDAQ:AAPL&output=rss&num=500) and historical stock quotes corresponding to the time frame the article was released. The agent will be trained on a subset of the data sample (leaving the rest for testing) to classify frequently occurring words as indicators of an increase/decrease in price (or possibly volatility) for the time period examined. The Python Natural Language Toolkit (http://nltk.org) will be used to tokenize and tag words by parts of speech (and possibly further analyze more grammatical structure). By classifying frequently occurring words (phrases?) as consistent with an increase/decrease in price for that time period the agent will be able to generate buy/sell signals. After training the agent using a subset of the sample data, the remaining data will be used to test the accuracy (profitability) of these buy/sell signals.

Timeline:

09/13-09/30 Further research
- watch NLP Stanford videos
- read more papers on topic
- write literature review
- Confirm data availability
  - write data collection prototype
  - build dataset

10/01-10/15 Programming / finalize method
- prototype program
- fully define classification method / models

10/15-10/31 Complete program / begin testing
- complete fully functional program
- begin training agent

11/01-11/11 Testing
- conduct testing to determine accuracy of buy/sell signals
- tweak program as necessary
- begin first draft of paper

11/20 First draft of paper completed

11/20-12/04 Final paper
- complete paper
- prepare presentation

12/04 **Project Due**