CS410 Project Proposal: Management Sentiment Analysis in Company SEC Filings

Team Members

- Venkat Rao Bhamidipathi (vrb3)
- Wang Chun Wei (wcwei2)

Introduction

The project aims to extract management sentiment across large capitalization American firms (eg. Dow Jones Industial Average constituents or S&P 500 constituents), whose quarterly (10Q) and annual reports (10K) are available in SEC (US Securities and Exchange Commission) filings. Using the 'manager discussion and analysis' section (MD&A) of these reports, we are able to analyze management sentiment across different firms and across time.

SEC's EDGAR Database

The EDGAR (Electronic Data Gathering, Analysis, and Retrieval) database is managed by the SEC and contains public disclosure data regarding company quarterly (10Q) and annual reports (10K) as well as substantial holdings notices, and mutual fund disclosures. We focus on the 10Q and 10K reports that outline the financial performance of the firm. In particular, we focus on the management discussion and analysis (MD&A) section which is the most subjective section of these accounting reports. Unlike financial statements, the MD&A section conatins valuable information on management's comments regarding the future of the business.

Project Proposal Details

Function

We aim to create a Python tool that allows users collect and analyze management sentiment for any input firm (or list of firms) that has SEC filings.

Who will benefit from such tool?

We envisage this to be useful for portfolio managers or private investors who wish to keep track of management sentiment. This will be helpful for investors that may not have the time to read through all the documents, but simply want a high level gauge on market wide and firm wide sentiment.

Similar tools that exist

Websites such as Last10K and Whalewisdom have utilized the scraping and analysis of company SEC filings for investors, alerting them of significant changes in manager holding, significant changes in reported accounting figures and general report sentiment.

Our tool is slightly different because:

- It focuses specifically on sentiment in the MD&A section.
- It allows users to analyse a large batch of stocks at once and compare and rank sentiment across stocks.

Resources, Techniques and Algorithms used

We will use bag-of-words representation on the MD&A section.

For calculating sentiment, we will leverage off existing dictionaries and sentiment word lists in Loughran and McDonald (Journal of Finance, 2011) (https://sraf.nd.edu/textual-analysis/resources/).

This will allow us to measure sentiment categories in Loughran and McDonald: such as negative, postive, uncertainity, litigious, modal and constraining.

Further, we can use some machine learning classification algorithm to determine a final positive and negative prediction using the sentiment scores determined above.

We will evaluate our tool by tabulating a confusion matrix and determining the accuracy of our classification algorithm.

Steps and Deliverables

The activities required to complete the project is as follows:

- Aided by EDGAR's master index file, we can download the 10Q and 10K files of a given company name or ticker and date range.
- Extract the MD&A component from the 10Q and 10K files
- Conduct sentiment analysis on the MD&A sections using “Bag-of-words” approach
    - Input: Name of the firm
    - Output: Sentiment score The process identifies positive and negative words (or a string of words) within an article. For this, it makes use of a large dictionary which contains words that carry sentiment. Each word in this dictionary can be assigned a weight. The sum of the positive and negative words is the final sentiment score generated by the model.

A rough timeline is provided below:

- Design: October 28th
- Development: November 18th
- Testing: December 1st