Final Project

Sentiment Analysis - Applied to Stock Market

Joe Gorfinkle
Taehoon Kim
Wenjie Gao
Bharathan M.Sridharan
May 12, 2018

Overview / Motivation

Our team consisted of four people grouped as two independent but cooperative teams. Team Red consisted of Joe Gorfinkle and Taehoon Kim, Team Blue consisted of Wenjie Gao and Bharathan Sridharan. The idea for Sentiment Analysis applied to common stocks was introduced by Wenjie, a Financial Engineering student. Bharathan has a similar interest as he is taking a Financial Analysis course as part of his MBA curriculum. Taehoon was interested in insuring whatever the team choose would reinforce his learnings in IST 341 and specifically the use of libraries. My interest as a former Financial Advisor, was to create a model as one ingredient of a multi-factor model, which would help in predicting future stock prices. Sentiment Analysis thus was a great fit for everyone on the teams.

Stock Market Illusions

The stock market is part of our daily lives through headline news feeds. Even if one isn't a direct participant, it is nearly impossible to miss the wild step function like market collapses; Black Monday in October 1987 when the DJIA lost over 22%, the Dot Com bubble burst of March 2000 when NASDAQ fell 25%, and of course the Financial crisis of 2008 when all indices fell over 50%. Individual stocks are not immune to these collapses: Dell, HP, AK Steel, AMD, JC Penney, Enron, Blackberry, Kodak, and even Apple to name just a few.

The price of a company's share of stock across all of its outstanding shares is what is known as the Market Capitalization for the company. It implies that of this point in time, under these circumstances, including all publicly known information, the "market" believes this is what this company is worth. It has always been a "head-scratcher" to many as to how the market can change its mind about the value of a company so quickly. Something must change this opinion or the market's "sentiment" about company value. We seek to understand this dynamic and this is our collective team's motivation.

Presentation

Slide 1:
We introduce the Team, Sentiment, and Stock Market Cycles

Slide 2:
We reinforce the sudden Market Collapse in 2008 with Bear-Stearns and Lehman Brothers stock prices.

Slide 3:
We suggest an appropriate analogy between the Stock Market "Illusion" of Market Capitalization via a short video: What the Bleep Do We Know, a look at life and the many illusions the Human mind injects upon itself. This just seem too fitting and it puts the Market collapses in perspective.

Slide 4:
We provide a definition of Sentiment according to Wikipedia and more importantly its measure, polarity from -1 to +1. We also introduce the two main resources we used; TextBlob sentiment analysis library which we apply to Seeking Alpha's Quarterly Earnings call transcripts. We choose TextBlob as it is user friendly, its results were similar to other libraries we tried including Google's NLP API and Microsoft's Azure, while both the latter charge a fee for service. We neglected to mention, Yahoo stock price history, well who could blame us for that?

Slide 5:
We introduce Domino's as a way to think about Sentiment and ask, who pushes the 1st domino in any market or stock price collapse? Who does the Market listen to? It is certainly not EF Hutton anymore, who is John Galt?

Slide 6:
What does a Stock Price reflect? The sum of all sentiments including Macro-Economic, Industry/Sector Outlooks, Company Fundamental and Technical Analysis, and of course the Sentiment, which we hypothesize is the result of the dialogue between Corporate Executives and Financial Analysts from the leading Financial Services companies. Our theory is that between the Corporate Officers and expert Financial Analysts, we can extract sentiment which will be reverberated throughout the market. A key principle here and probably not emphasized strongly enough is that these Financial Analysts guide Institutional Money Managers whose large volume trading influence market prices. That is the belief at least and for now we will accept this, although acknowledge it may be indeed not be.

Slide 7:
We introduce our approach: we look at Bull and Bear stocks to see if any correlation between the stock price and the sentiments we extract from the Earnings Conference call transcripts. Our program looked through seventy-five earnings call transcripts of eight companies, four bulls, four bears, over a five-year cycle; 2013-2018, producing 150 sentiment polarity scores.

Slide 8-13:
We introduce the code and some of the things we learned about TexBlob(TB), including its difficulty in handling negation. For example, TB would give a positive polarity to the sentence, “I don’t love it.” while indeed it is quite negative. Also, TB is not tailored for Financial dialect: it would not understand that Bull should have a positive polarity. We showed how we would replace some instances we found that seemed appropriate after reading through a number of transcripts.

Slide 14-19:
We provide charts showing the price history along with the scaled polarity sentiment scores to visually depict correlation or lack of correlation. We introduce Lehman Brother chart of 2008 to show that even under those severe negative market conditions, sentiment remained in the same range as all other scores we interpreted. Thus we conclude that we were unable to show any correlation between our interpretation of sentiment extracted from our source to price history.

Slide 20:
We provide some Insights from what we have learned so far: 1) No correlation of stock price with our implementation of sentiment analysis. 2) Analysts tended to be more positive than the Corporate Executives, except in the case of Dish Networks. 3) Amazon Corporate Sentiment trends in a very narrow channel, suggesting they are very careful with their statements within earnings calls. 4) The listener’s comments were insightful. These comments are posted at the end of the transcript by those that are unable to ask questions, only Analysts have that privilege. They have very strong sentiment which we are now extracting in the next version of our model. 5) We continue to look for the Domino pusher, the needle that pops the balloon. Where is Michael Barry?

Slide 21:
We provide some Project insights: 1) ETL is a pain especially if you want to scale your project into a real product and provide automatic or real time updates 2) Free Data is getting harder to extract. The trends we see are that API’s and Historical data is a premium pay service, indeed we had to sign up for Seeking-Alpha Pro Premium at $599 annually to get our transcripts. We were precluded from Google and Microsoft’s as they charge usage fees which are not transparent at all 3) ML is a next step although we need to feed it with a source that can assist us in finding sentiment. 4) Finally, we can create our own sentiment library for finance and provide our own weight factors while reducing noise. 5) Macro-sentiment is another idea to look across the totality of the transcript as opposed to words, sentences, and paragraphs. As we looked through the transcripts individually we can see some significant differences in the way Corporate Executive handle calls. Indeed, there is a lot that can be done here.

Final Thoughts

Market participants may claim that price is the ultimate sentiment, the interpretation being that the actual price is the best predictor of market sentiment, and this is the best one can

do in terms of predicting sentiment. We neither agree nor disagree with this view while acknowledging that automated trading models adopted by Institutional Money managers now account for over 60% of all market trading. Intuition suggests that even balloons and tires can hold more air than they are rated for, and expand well beyond their limit, before bursting. This is the inflection point we continue to seek and realize we indeed are hunting for the needle in the haystack. Wish us luck.