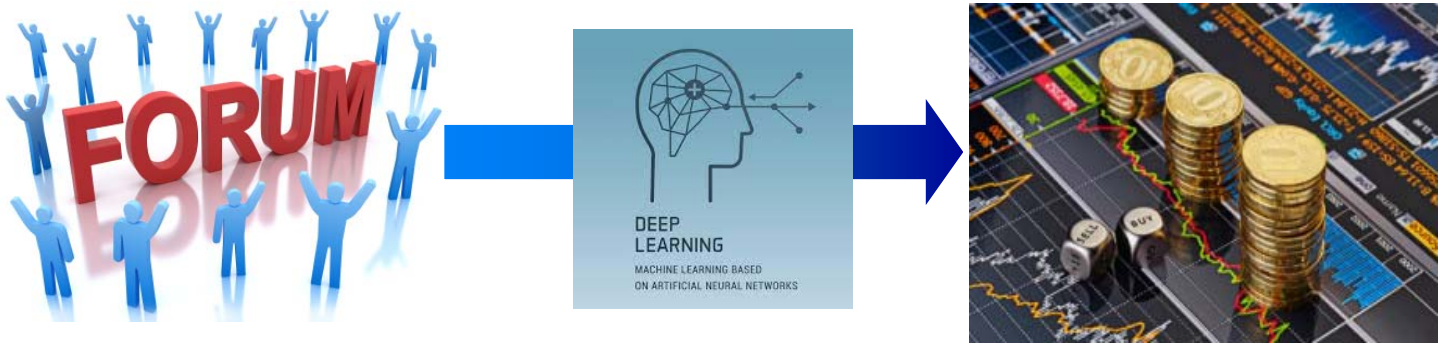
A Deep Learning Model for Stock Return Prediction Using Social Media



Michael Chau
HKU Business School
The University of Hong Kong

May 27, 2022

About Myself

- Professor (Innovation and Information Management), HKU Business School, The University of Hong Kong
- Doctoral degree in management information systems from the University of Arizona
- Bachelor degree in computer science and information systems from the University of Hong Kong
- Research in big data, data mining, artificial intelligence, search engines, and social media analytics
- Recipient of the HKU Outstanding Young Research Award (2014) and Knowledge Exchange Award (2013 & 2016), the IEEE ISI Leadership Award, and the INFORMS ISI Design Science Award.

Outline

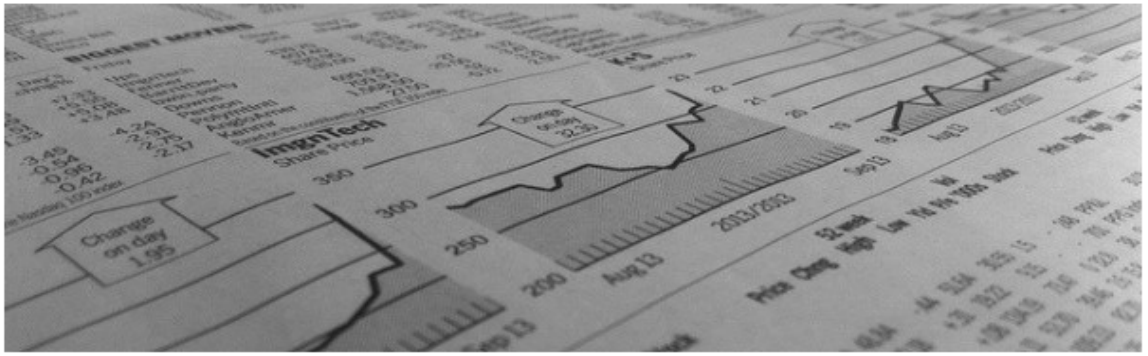
- Research Background
- Proposed Design
- Methodology and Data
- Main Results
- Contributions
- Ongoing Work

Research Background

- Prevalence of using social media for investment decisions

MEDIA TO INFORM INVESTMENT DECISIONS

26 February 2013 / Matt Rhodes / Financial services, Social business / No Comments



New research shows that 70% of affluent US investors have made an investment decision based on information they have learned from social media; 34% use social media specifically to help inform their personal finance and investment decisions. Even for High Net Worth individuals with more than $1m in investable assets, 25% seek investment advice from social media.

The report, from Cogent Research, is based on a survey of 4,000 US investors with more than $100,000 in investable assets and shows the growing importance of social media in the investment management industry.

Research Background

Auto Sales Boost Sirius Subscriber Numbers

Aug 12, 2009 5:59 AM | 20 comments | About: Sirius XM Holdings Inc. (SIRI)

By Brandon Matthews

The Cash For Clunkers program has received a lot of press recently, and there are some reporters and analysts that have made the correct assumption that the program will give a much needed boost to Sirius XM Radio's (NASDAQ:SIRI) subscriber rolls through promotional subscriptions. The first half of 2009 revealed that most of Sirius XM Radio's reported subscriber losses had come from the promotional subscription bucket, rather than those subscriptions which are self paid by individual consumers.

During last week's earnings conference call, Sirius XM Radio C.E.O. Mel Karmazin stated that the subscriber acquisition rate for July had turned positive. We can now look at the July 2009 auto sales statistics as a benchmark clue of the number of new cars that must be produced and/or sold in a given month/quarter to net Satellite Radio positive gains in its promotional subscription base. The reliability of the data, however, will need to be measured against future statistics.

The Cash For Clunkers program, combined with the bankruptcies of General Motors and Chrysler, can skew the data from various perspectives. At last check, General Motors subscriptions were counted at the time of sale, while Chrysler vehicles are counted at the time of production. The July auto sales stats for General Motors were less than spectacular. Similarly, Chrysler plants had been shut down until August.

This is good news for Satellite Radio investors, considering that the company was able to achieve positive subscriber numbers in July despite Chrysler and General Motors playing a limited role in that success. Reports are now coming in that Chrysler plants have not only restarted, they are running on overtime. The boost from Chrysler alone in the current quarter will certainly lead to a continuation of positive subscriber growth for Sirius XM. General Motors sales improvements under the Cash For Clunkers program are sure to boost these numbers even further, which goes for all of Sirius XM's top partners including Ford (NYSE:F).

The Cash For Clunkers program will not only boost subscriber rolls in Q3, it promises great results for Q4 as well. In that quarter, many vehicles sold under the CFC program will see their promotional free trials end, most notably Toyota (NYSE:TM) which is thus far the largest beneficiary of new car sales under the program, as well as others including Honda (NYSE:HMC).

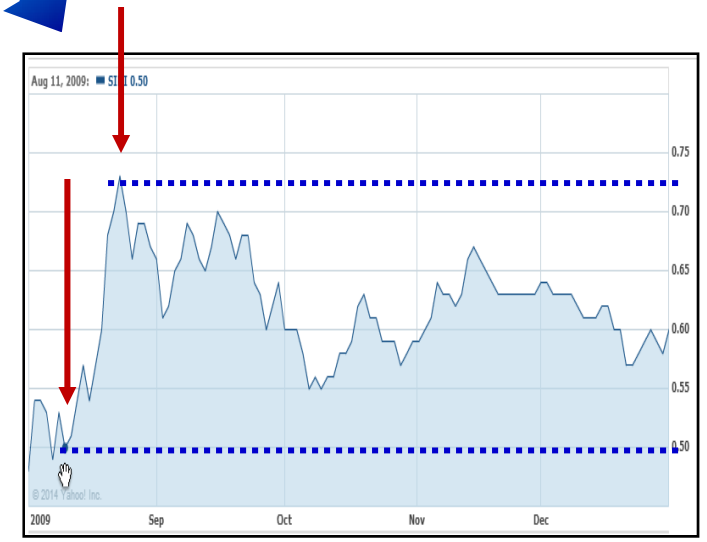
Position: Long SIRI, F....no other securities mentioned.

Tagged: Services, Broadcasting - Radio

Positive Words

Negative Words

Positive effects on stock returns?



Research Background

- Deep learning has achieved excellent performance in various text-mining tasks in the past few years.
- Many researchers have used deep learning methods on text (news or social media articles) to predict stock returns and achieved good results.
- In this study, we propose a deep learning model for stock price movement prediction.
- We design a novel way to incorporate social media engagement data as sample weights into the deep learning model.

Research Background

- Reflection of information in stock prices
  - Efficient Market Hypothesis (Fama 1970)

    Weak form – reflects all past available info

    Semi-strong form – reflects all publicly available info

    Strong form – reflects all info (including private)
  - As information is costly, price cannot perfectly reflect the information when it is available (Grossman and Stiglitz 1980)

Research Background

- Influence of social media on stock prices
  - Sentiment from discussion forums can have a larger impact on stock returns than views in news articles (Chen et al. 2011; Chen et al. 2014)
  - Sentiment in 24 tech-sector stocks can predict returns in the aggregate stock level, but the effect is weak for individual stock level (Das and Chen 2007)
  - Volume of messages has a significant positive effect on volatility on 45 large-cap stocks. Its effect on stock returns is negatively significant but economically small (Antweiler and Frank 2004)

Research Background

- Deep learning
  - Deep learning models (e.g., LSTM, CNN, BERT) have been shown to perform very well in text mining tasks.
  - Applied on text (news or social media articles) to predict stock returns and achieved good results (e.g., Akita et al. 2016; Kraus and Feuerriegel 2017; Fischer and Kraus 2018; Chen et al. 2020).

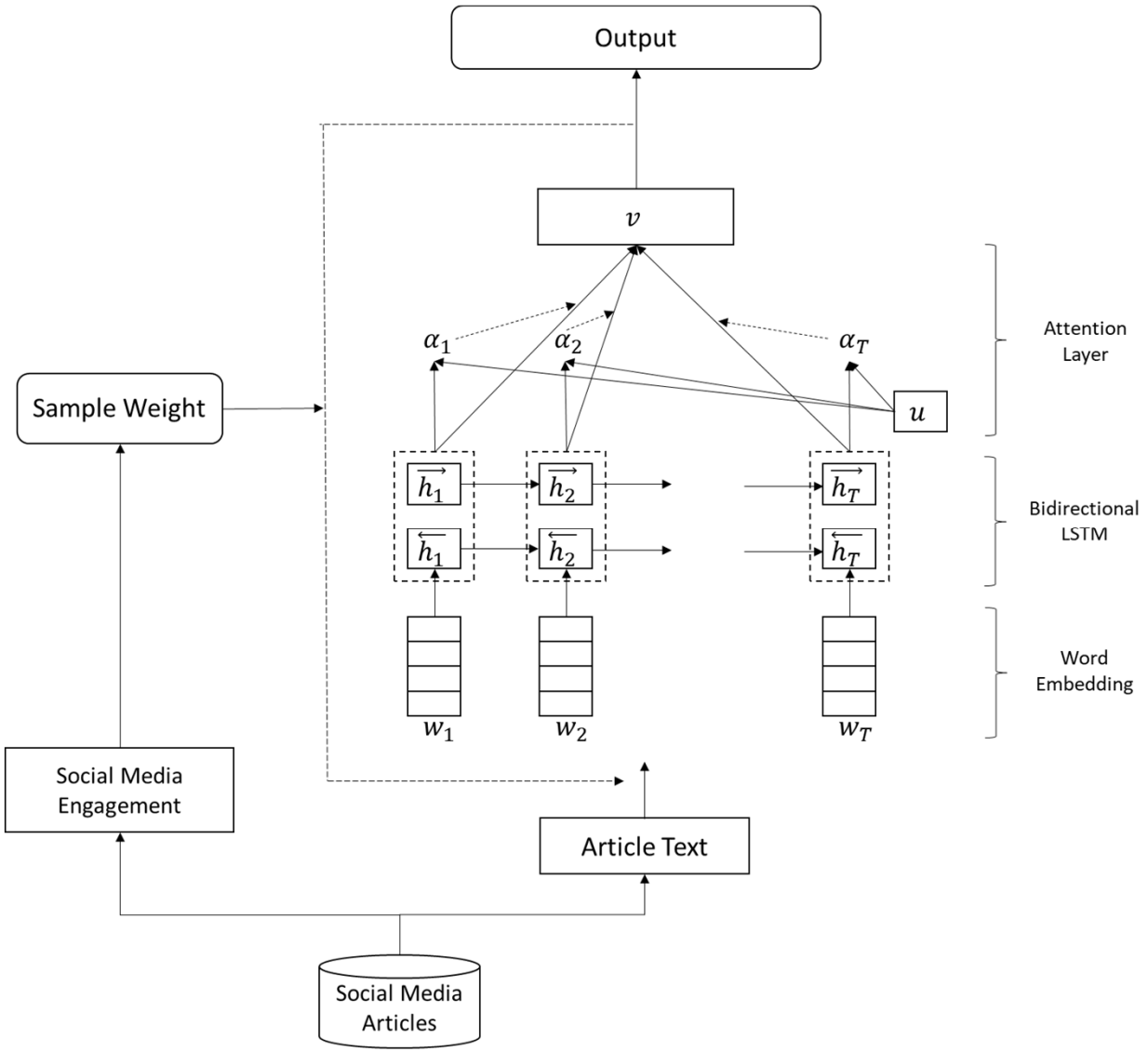
Research Background

- Social media engagement data
  - On social media, readers interact with the contents.
  - More interactions may indicate that the contents are considered more important.
  - User comments may also reflect the perceived quality of the contents.

Research Background

- Main Research Question
  - How to incorporate social media engagement data into a deep learning model for stock return prediction?

Proposed Deep Learning Model



Social Media Engagement

- We consider four measures of social media engagement:
  - Whether the editor picks the article (EditorPick)
  - The number of followers of the author (FollowerCount)
  - The number of comments (CommentCount)
  - The sentiment of the comments (CommentSentiment)
    - As positive comments can be seen as a support to an article, we assign a higher score to articles with more positive comments. Articles with more negative comments than positive comments will be assigned a score of 0.

Social Media Engagement

- We incorporate a combination of these four measures as sample weights in the training process of the model.
- The role of sample weights is to scale the error delivered to previous layers and further influence the weight update for each sample.
- Because there is great variability in the information content of articles, these measures on social media engagement by other users can be a good indicator of the quality importance of articles.
- We build a weight vector of samples, where the weight of a sample s_i is w_i, which is calculated as the average of the four social media engagement measures.

Data

- Social media data
  - Collected from Seeking Alpha (SA) (e.g. contents of the articles, stock tickers, dates of publication, stock sectors, etc.)
  - Selected SA articles and comments posted from Dec 2004 to Jun 2014
  - Proprietary data obtained directly from SA (author's and article's information)
- Financial data
  - stock price related data (e.g. opening/closing prices, S&P 500 indices prices) from Compustat
  - analyst's related data (e.g. analysts' upgrade/downgrades, ratings) from IBES

Data

- Why choose SA?
  - It has broad coverage of stocks
    - includes more than 3,000 small and mid cap stocks covered in the past year
  - Daily email alerts and SA apps are available for Android, iPhone and iPad
  - Higher credibility as SA contributors can earn
    - US$10 per 1,000 page views
    - a guaranteed minimum of US$500 when the article is classified as "Top Idea"
- To ensure that stock opinions are related to a particular company
  - Only single-ticker articles are chosen

Data

GE: Too Little Earnings, Too Much Dividend, Share Price Will See $10 Before $20

Jul. 8, 2018 1:54 PM ET | 10 likes | About: General Electric (GE), Includes: GEC

Robert Honeywill
Long only

Follow

(3,150 followers)

Summary

- GE continues to suffer losses on a GAAP basis of financial reporting.
- Non-GAAP reporting has its uses, but non-GAAP results are not a valid basis for paying dividends.
- GE's financial reporting is complex, partly because GE is a conglomerate, and partly because GE's reporting is not as helpful as it could be.
- A summary approach to balance sheet presentation, source and application of funds, and dividend affordability, reduces complexity of review of GE's restructuring efforts.

A Bird's Eye View Of GE's Restructuring Efforts

General Electric's (GE) 2017 form 10-K filed with the SEC is a voluminous document of 206 pages of information, designed to meet statutory disclosure requirements. Much of the information for disclosure is necessarily repetitive, and creates a need for a "sorting the wheat from the chaff", for the reader. While that is an issue for all reporting entities, I would not be alone in considering GE's approach to reporting adds to complexities. GE is undergoing

Data

- There are in total 138,756 articles involving 8,202 individual stocks.
- We split the data set into 80% training set and 20% testing set based on chronological order.

Data

	Training	Testing
Total # Articles	111,004	27,752
Total # Articles for "positive direction" (Class 1)	49,588	12,369
Total # Articles for "negative direction" (Class 2)	61,416	15,383
Total # Ticker Symbols	6,786	5,167
Avg. # Words per article	519	671
Avg. # Comments per article	5	9
Avg. % Negative words	1.98%	1.77%

Comparison Benchmarks

- Standard neural network with three layers (NN)
- Bidirectional LSTM
- Bidirectional LSTM with attention layer
- Proposed Model: Bidirectional LSTM with attention layer + sample weight adjusted by social media engagement data

Main Results

Day	Model	F-score	Recall	Precision
t+7	NN	0.504	0.503	0.505
	Bi-LSTM	0.514	0.563	0.473
	Bi-LSTM + attention	0.602	0.623	0.582
	Proposed Model	**0.662**	**0.691**	**0.635**
t+1	NN	0.507	0.511	0.503
	Bi-LSTM	0.605	0.670	0.551
	Bi-LSTM + attention	0.621	0.633	0.609
	Proposed Model	**0.691**	**0.712**	**0.671**

Short-term Prediction

- The power of using the information on social media to predict stock price movement is most visible at short horizons (e.g. 1 day in this study).
- The performance difference in the time dimension have two possible explanations:
  - Predictive ability of models
  - Timeliness of social media information

Trading Simulation

- We calculate the average profits over the 8,202 individual stocks that are presented in our dataset.
- Assume an investment of $10,000 for every stock
  - Long when the prediction is positive
  - No action when the prediction is negative

Trading Simulation

Model	Average Profit on Day *t+1*	Average Profit on Day *t+7*
NN	-$10.40	-$20.32
Bi-LSTM	$49.11	-$20.46
Bi-LSTM + attention	$100.85	$73.52
Proposed Model	**$175.69**	**$143.91**

Contributions

- We are among the first to investigate the how social media engagement data can be incorporated in deep learning models
- Propose a method that uses social media engagement data as sample weights to improve prediction results
- Trading strategies can be developed based on the findings

Ongoing Work

- Improve the proposed model using other advanced text mining methods (e.g., BERT)
- Evaluate the impact of the different social media engagement data

Acknowledgments

- The research is supported in part by the HKU-SCF FinTech Academy.
- We thank Seeking Alpha for providing part of the data used in this study.

Questions?

- Contact info:

Michael Chau
HKU Business School
The University of Hong Kong

Email: mchau@business.hku.hk
URL: http://www.business.hku.hk/~mchau/