Using Structured Events to Predict Stock Price Movement: An Empirical Investigation

Xiao Ding[†]*, **Yue Zhang**[‡], **Ting Liu**[†], **Junwen Duan**[†]
[†]Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{xding, tliu, jwduan}@ir.hit.edu.cn
[‡]Singapore University of Technology and Design
yue_zhang@sutd.edu.sg

Abstract

It has been shown that news events influence the trends of stock price movements. However, previous work on news-driven stock market prediction rely on shallow features (such as bags-of-words, named entities and noun phrases), which do not capture structured entity-relation information, and hence cannot represent complete and exact events. Recent advances in Open Information Extraction (Open IE) techniques enable the extraction of structured events from web-scale data. We propose to adapt Open IE technology for event-based stock price movement prediction, extracting structured events from large-scale public news without manual efforts. Both linear and nonlinear models are employed to empirically investigate the hidden and complex relationships between events and the stock market. Large-scale experiments show that the accuracy of S&P 500 index prediction is 60%, and that of individual stock prediction can be over 70%. Our event-based system outperforms bags-of-words-based baselines, and previously reported systems trained on S&P 500 stock historical data.

1 Introduction

Predicting stock price movements is of clear interest to investors, public companies and governments. There has been a debate on whether the market can be predicted. The Random Walk Theory (Malkiel, 1973) hypothesizes that prices are determined randomly and hence it is impossible to outperform the market. However, with advances of AI, it has been shown empirically that stock price movement is predictable (Bondt and Thaler, 1985; Jegadeesh, 1990; Lo and MacKinlay, 1990; Jegadeesh and Titman, 1993). Recent work (Das and Chen, 2007; Tetlock, 2007; Tetlock et al., 2008; Si et al., 2013; Xie et al., 2013; Wang and Hua, 2014) has applied Natural Language Processing (NLP) techniques to help analyze the effect of web texts on stock market prediction, finding that events reported in financial news are important evidence to stock price movement prediction.

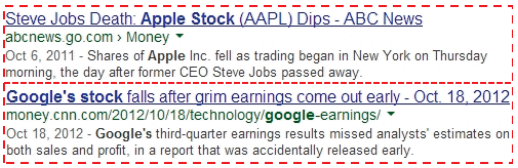

Figure 1: Example news for *Apple Inc.* and *Google Inc.*

As news events affect human decisions and the volatility of stock prices is influenced by human trading, it is reasonable to say that events can influence the stock market. Figure 1 shows two pieces of financial news about *Apple Inc.* and *Google Inc.*, respectively. Shares of *Apple Inc.* fell as trading began in New York on Thursday morning, the day after its former CEO Steve Jobs passed away. Google's stock fell after grim earnings came out. Accurate extraction of events from financial news may play an important role in stock market prediction. However, previous work represents news documents mainly using simple features, such as bags-of-words, noun phrases, and named entities (Lavrenko et al., 2000; Kogan et al., 2009; Luss and d'Aspremont, 2012; Schumaker and Chen, 2009). With these unstructured features, it is difficult to capture key events embedded in financial news, and even more difficult to model the impact of events on stock market prediction. For example, representing the event "Apple has sued Samsung Electronics for copying 'the look and feel'

*This work was done while the first author was visiting Singapore University of Technology and Design

Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1415–1425,
October 25-29, 2014, Doha, Qatar.