Fine-grained analysis of explicit and implicit sentiment in financial news articles

Marjan Van de Kauter [a,*], Diane Breesch [b], Véronique Hoste [a]

[a] *LT³ Language and Translation Technology Team, Faculty of Arts and Philosophy, Ghent University, Groot-Brittanniëlaan 45, 9000 Ghent, Belgium*
[b] *Department of Business (BUSI) – Accounting and Auditing, Faculty of Economic Sciences and Solvay Business School, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium*

ARTICLE INFO

ABSTRACT

This paper focuses on topic-specific and more specifically company-specific sentiment analysis in financial newswire text. This application is of great use to researchers in the financial domain who study the impact of news (media) on the stock markets.

We investigate the viability of a new fine-grained sentiment annotation scheme. Most of the current approaches to sentiment analysis focus on the detection of explicit sentiment. As news text often contains implicit sentiment, i.e. factual information implying positive or negative sentiment, our approach aims to identify both explicit and implicit sentiment. Furthermore, this sentiment is analyzed on a fine-grained level by detecting the topic of the sentiment, as sentiment is not always expressed towards the topics one is interested in.

In order to test our approach, we assembled a corpus of company-specific news articles, which was manually labeled by four annotators to create a gold standard. We compare the results of our method to the performance of two coarse-grained baseline systems: a lexicon-based approach and a supervised machine learning approach that makes use of lexical features. Our fine-grained approach outperforms both baselines, and its output shows substantial to almost perfect agreement with the gold standard sentiment labels. Using our annotation scheme, we are able to filter out irrelevant sentiment expressions and detect explicit and implicit sentiment in a reliable way.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In the last 50 years, a significant amount of research has been dedicated to the impact of news on the stock markets. Experts in the financial field have diverging views on the manner in which markets respond to new information, and, consequently, on how news can be incorporated in trading and investment strategies.

In modern finance, which emerged in the 1950s, investors were all assumed to be rational decision makers. Starting from this rationality assumption, the idea of the Efficient Market Hypothesis was created (Fama, 1970). It was widely accepted that new information spreads very quickly and is incorporated into the stock prices without delay. As a consequence, stock prices at any point in time reflect all available information and are a good estimation of the intrinsic value of that stock ('the price is right' principle). If information is immediately reflected in stock prices, this also implies that future price changes will only reflect future news and will be independent of the past price changes. This logic is referred to as 'the logic of the random walk' (Alexander, 1961; Fama, 1965; Malkiel, 2003).

Consequently, it is impossible for investors to predict market prices, and so it is difficult for any investor to beat the market and earn above-average returns without accepting above-average risk.

The Efficient Market Hypothesis was however challenged in the 1980s when several academic time series studies had shown excess volatility of stock prices relative to what would be predicted by the efficient markets model. As too many market anomalies were noted, the field of behavioral finance was further developed in the 1990s (Shiller, 2003). Contrary to modern finance, behavioral finance uses concepts from psychology to describe investors' behavior, arguing that the decision-making process is not always driven by rational calculation, but is often influenced by emotion and other psychological factors (De Bondt & Thaler, 1994). Bernard Baruch, a multimillionaire through stock market operation, used the following quote: "But what actually registers in the stock market's fluctuations are not the events themselves, but the human reactions to these events". Also, Thaler (1993) stated that "sometimes in order

* Corresponding author. Tel.: +32 9 331 19 39.
  E-mail addresses: Marjan.VandeKauter@UGent.be (M. Van de Kauter), Diane.Breesch@vub.ac.be (D. Breesch), Veronique.Hoste@UGent.be (V. Hoste).

to find the solution to an empirical puzzle, it is necessary to entertain the possibility that some of the agents in the economy behave less than fully rationally some of the time".[1] One of the factors influencing the way a stock is perceived by investors is the news media. As stated by Nofsinger (2013), "by trying to appeal to our interests and emotions, the media naturally gravitate toward the active investment decisions of stock selection and market timing". All these findings have triggered the interest in sentiment analysis applied to financial news media (see Section 2.3).

Sentiment analysis, also known as opinion mining, is a Natural Language Processing task aimed at the automatic identification and analysis of "people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" in text (Liu, 2012). Note that the concept 'sentiment' is not used the same way in NLP as it is in the financial domain, where (investor) sentiment is defined as "the expectations of market participants relative to a norm: a bullish (bearish) investor expects returns to be above (below) average, whatever "average" may be" (Brown & Cliff, 2004).

In this paper, we investigate the use of sentiment analysis to detect positive and negative sentiment expressed about certain companies in financial news articles. The application of sentiment analysis to financial newspaper text enables researchers in the field of finance to identify positive and negative company news in an automatic way. Consequently, more data can be processed in less time, which could lead to new insights into the correlations between news (media) and the stock markets.

Most of the current approaches to sentiment analysis in the financial domain have focused on the coarse-grained analysis of sentiment expressed in text (e.g. Bollen, Mao, & Pepe, 2011; Devitt & Ahmad, 2007; Du, Xu, & Huang, 2014; Mostafa, 2013; Musat & Trausan-Matu, 2010; etc.). These methods detect the mood of a certain text fragment by taking into account all expressions of positive and negative sentiment in that fragment. However, for financial news articles, we hypothesize that not all expressions of sentiment are related to the company one is interested in. Even below sentence level, topic shifts may occur. We thus believe that coarse-grained methods are insufficient for the detection and polarity classification of sentiment expressed about companies in financial news text and investigate the viability of a new, fine-grained annotation scheme, which allows us to pinpoint the particular phrases in a text that express sentiment, and to analyze these sentiment expressions in a fine-grained manner. For each detected expression of sentiment, a target, i.e. a topic is identified, making it possible to filter out expressions irrelevant to a given company of interest. Moreover, as opposed to the majority of sentiment analysis methods (e.g. Dave, Lawrence, & Pennock, 2003; Ding, Liu, & Yu, 2008; Hu & Liu, 2004; Wilson, Wiebe, & Hoffmann, 2005; etc.), our goal is to detect explicit as well as implicit sentiment. We hypothesize that newswire text does not always explicit a positive or negative opinion of the author, but more often contains factual information implying positive or negative sentiment. To our knowledge, none of the existing sentiment annotation schemes combines the identification of all types of explicit and implicit sentiment expressions with the fine-grained analysis of these expressions.

In this paper, we evaluate the suitability of our fine-grained annotation scheme for the detection of positive and negative sentiment expressed about a given company in financial news articles. We applied the annotation scheme to a corpus of financial newswire text and evaluated its performance in comparison to that of two baseline coarse-grained approaches: a lexicon-based method and a machine learning approach. We will show that a coarse-grained method is insufficient for the detection of company-dependent sentiment in financial news text. Annotations made using our fine-grained annotation scheme, on the other hand, are shown to be a reliable resource for this purpose, despite the difficulty of the task. In future work, we will train an automatic sentiment analysis system on a corpus of financial news articles manually annotated using our annotation scheme. The results presented in this paper indicate the upper bound of the performance of such a supervised sentiment analysis system.

This paper is structured as follows. Section 2.1 gives a brief overview of studies on the impact of news (media) on the stock markets. Sections 2.2 and 2.3 discuss existing research on sentiment analysis in general and for the financial domain. In Section 3, we present the methods and data used for the evaluation experiments, of which the results are discussed in Section 4. Finally, Section 5 contains some conclusions and prospects for future research.

2. Theory

2.1. The impact of news (media) on the financial markets

The stock market is largely driven by news, which can refer to macroeconomic factors, geopolitics or company-specific factors. Good news tends to lift the market, whereas bad news has a tendency to dampen the market's growth. However, Engle and Ng (1993) assess that news influences the stock market in an asymmetric way, in that negative news exerts a greater impact on volatility than positive news. News can also have different effects on the market return depending on the overall state of the market. Investors in a 'bullish' market may react to positive news in a very different way than investors in a 'bearish' market. In line with the asymmetric impact noted by Engle and Ng (1993), Lee, Jiang, and Indro (2002) showed that volatility increases (decreases) when investors become more bearish (bullish).

Prior literature has clearly shown that the stock market is also driven by 'investor sentiment', i.e. beliefs about future cash flows and discount rates that are not supported by prevailing fundamentals like macroeconomic or company-specific factors, but related to (amongst others) emotional reasoning. The notable work of De Long, Shleifer, Summers, and Waldmann (1990) models the influence of this 'noise trading' on equilibrium prices. High sentiment causing excessively optimistic estimates of future expected cash flows leads to investors overvaluing stocks. The reverse is true for low sentiment periods, in which stocks are undervalued (Baker & Wurgler, 2006; Brown & Cliff, 2004).[2] Mian and Sankaraguruswamy (2012) concluded that the stock price sensitivity to good (bad) earnings news is higher (lower) when sentiment is high than when sentiment is low. The effect of sentiment on stock price sensitivity was also shown to be stronger for bad news than for good news.

Media stories play a strong role in this market noise and can even act as a prime source of instability in financial markets. The media enrich information with new elements such as 'emotion' or 'suspense', either in content or presentation style. In this process, however, the character of the information is altered, influencing the perception of the investor. The media are far from being

---

[1] Already in the 1970s, Simon (1972) proposed the idea of 'bounded rationality', asserting that decision makers intend to act rationally, but may be confronted with constraints in, for example, information processing capacities. As such, they are bound to make satisficing instead of optimal decisions.

[2] In modern finance supporting the Efficient Market Hypothesis, there is no scope for investor sentiment, since any mispricing in securities by irrational investors would immediately be arbitraged away. However, the limits on arbitrage constrain rational traders from forcing the prices back to their fundamental value (Baker & Wurgler, 2006).