



BizPro: Extracting and categorizing business intelligence factors from textual news articles



Wingyan Chung*

421 North Woodland Boulevard, DeLand, FL 32723, USA

ARTICLE INFO

Article history:

Available online 8 February 2014

Keywords:

Business intelligence
BI factor extraction
Categorization
Profiling
Online news

ABSTRACT

Company movements and market changes often are headlines of the news, providing managers with important business intelligence (BI). While existing corporate analyses are often based on numerical financial figures, relatively little work has been done to reveal from textual news articles factors that represent BI. In this research, we developed *BizPro*, an intelligent system for extracting and categorizing BI factors from news articles. *BizPro* consists of novel text mining procedures and BI factor modeling and categorization. Expert guidance and human knowledge (with high inter-rater reliability) were used to inform system development and profiling of BI factors. We conducted a case study of using the system to profile BI factors of four major IT companies based on 6859 sentences extracted from 231 news articles published in major news sources. The results show that the chosen techniques used in *BizPro* – Naïve Bayes (NB) and Logistic Regression (LR) – significantly outperformed a benchmark technique. NB was found to outperform LR in terms of precision, recall, *F*-measure, and area under ROC curve. This research contributes to developing a new system for profiling company BI factors from news articles, to providing new empirical findings to enhance understanding in BI factor extraction and categorization, and to addressing an important yet under-explored concern of BI analysis.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Company movements and market changes often are reported in news articles, providing managers with important business intelligence (BI). These articles report on financial performance, market shares, products and services, technological changes, legal issues, and competition. The information in these articles influences market reactions before financial statements are published. For example, the news story titled “Apple Adds 2 Publishers to its Store for E-Books” signals strategic risks to other e-book and tablet sellers, even though the financial implication is not yet known immediately. Unlike financial figures that focus only on past results, information in news articles is often up-to-date about companies’ current and projected future conditions. Investors, financial managers, and BI analysts use these articles to evaluate companies’ performance and to support their decisions.

According to the 2012 Society of Information Management (SIM) survey of IT executives who represent 195 organizations in 17 industries, business intelligence (BI) is ranked as the top most important application and technology development among all 51

items (Luftman & Derksen, 2013). Over the decade beginning 2003, the findings of the SIM annual survey show that BI has always been ranked among top three most important application and technology developments, rising steadily from the third to the first rank during the recent global financial crisis and maintaining its number-one spot between 2009 and 2012. In the aftermath of the crisis, corporate managers, BI analysts, and policy makers are searching for better ways to identify factors that indicate business intelligence (Tapscott & Tapscott, 2008). “Business intelligence” is defined as the result of “acquisition, interpretation, collation, analysis, and exploitation of information in business” (Chung, Chen, & Nunamaker, 2005; Davies, 2002). However, traditional management approaches such as risk analysis have become obsolete as quantitative models fail to describe today’s complex financial environment, according to former European Union parliament member Daniel Daianu (Daianu, 2008). While most corporate analyses are still based on financial ratios and earning figures, relatively little work has been done to reveal BI factors from textual news articles. We conceptualize a “BI factor” as a piece of qualitative evidence that can influence market reactions on a company. This evidence is often found in published reports such as textual news articles. For example, the sentence: “The company (Apple) has just signed a deal with two more independent publishers to sell electronic versions of their books on the new device” indicates a new BI factor that could affect

* Tel.: +1 386 822 7002.

E-mail addresses: wchung@stetson.edu, wchung9@gmail.com

corporate strategy. Prior works on analyzing textual media content try to predict asset prices (Engelberg, 2008), to detect fraudulent behaviors of corporate management (Churyk, Lee, & Clinton, 2008), to correlate between financial performance and chairman statements' readability and understandability (Smith & Taffler, 1992), to discover customer concerns from textual online reviews (Chung & Tseng, 2012), and to cluster quarterly reports to indicate performance changes (Kloptchenko et al., 2004; Magnusson et al., 2004).

Despite these works, extracting BI factors based on news articles is not widely studied. Findings from interviews conducted with over 30 executives from companies in various industries show that automatic pattern discovery from news articles would be particularly useful for future detection and mitigation of risks, thus providing BI in a timely manner (p. 4075, Blackhurst, Craighead, Elkins, & Handfield, 2005). Ideally, intelligent systems that can automatically identify key terms, extract BI factors from news articles, and group these factors into their respective categories should serve the needs of corporate managers, BI analysts, and policy makers. However, such systems are not widely available. Consequently, these tasks remain to be labor-intensive work done by business analysts.

In this research, we reviewed literature on mining textual documents for financial and BI applications, and developed and evaluated a system called *BizPro* that automatically extracts and categorizes BI factors from company-specific news articles. The system incorporates novel text mining techniques, BI modeling and categorization, and expert domain knowledge in BI to support extraction and profiling of company BI factors. Using input from a BI expert and two independent human raters, we conducted a case study of profiling BI factors of four companies in the information technology industry. We report the results that demonstrate the usability and effectiveness of the system. Implications and contributions are discussed.

The remainder of this paper is organized as follows. Section 2 surveys previous research in data and text mining of company textual reports. Section 3 presents a BI categorization scheme and the *BizPro* system. Section 4 describes the setup of a case study to demonstrate the system usability using company-specific news articles as data. Section 5 reports the findings and discusses the implications. Section 6 concludes the research and presents future directions.

2. Literature review

The recent financial turmoil that led to the greatest economic recession since the 1930s was widely attributed to poor management in lending, investment, and company debt management (Morgenson, 2010; Nasiripour, 2010). The industry practice of obtaining BI has focused on analyzing the quantitative data provided in financial statements and in published news articles. However, the current practice of relying primarily on quantitative data may not reveal sufficiently the BI factors. In contrast, qualitative information was found to be useful in improving understanding of numerical accounting information (McIntyre, 1975). As information technologies continue to advance, new techniques are developed to obtain BI using various types of data. Major applications of IT include financial text mining, fraud detection, and market prediction. In these applications, textual information from news and other documents is collected and analyzed to reveal trends and to provide valuable BI. Data mining and categorization techniques have been used in analyzing documents and in selecting the right features to represent the documents in order to achieve high performance in categorizing the BI factors. We therefore review below the two related areas of mining textual reports in finance and selecting features for text categorization.

2.1. Mining textual reports in business

Financial text has become more readily available due to the proliferation of the Internet and the openness of the U.S. Securities and Exchange Commission and other financial information services. This trend gives rise to financial text mining, the systematic analysis of textual information from financial reports, news articles, and financial market messages. Among the many applications of financial text mining, stock market prediction and fraud detection have attracted the most attention.

2.1.1. Market prediction and competitive intelligence

Various textual clues and features have been included in predicting market changes in stock prices. Fung et al. combine text information extracted from financial news article with multiple time series stock data to support stock price prediction (Fung, Yu, & Lam, 2003). They used the Support Vector Machine algorithm to build two classifiers for predicting upward and downward trend types. Their experiment showed a significant increase in profit using their system compared with benchmarks.

Apart from time-series data, linguistic data present another important source of information for market prediction. Companies can use a variety of linguistic and sentiment analysis algorithms to extract and to model human opinion from news and textual articles (Wright, 2009). Using a linguistic measure called Gunning fog index, Li found a significant correlation between annual report readability and firms' performance (Li, 2008). Word polarity, use of present and future tenses, word count in each sentence, and number of syllables in each word were considered in Li's study. The independent variables include the size of the firm, market to book value, firm age, volatility of business, complexity of operations, and incorporation state. Balakrishnan, Qiub, and Srinivasanc (2010) applied text categorization techniques to predicting stock market movement and found the portfolio based on the model predictions earns significantly positive size-adjusted returns. Also, the classification model captures information not contained in document-level features of clarity, tone and risk sentiment considered in prior research. The results highlight the potential use of automatic categorization techniques in extracting BI factors from other textual documents.

Apart from company reports, researchers have used financial news articles to predict market trends. Chen et al. quantified news coverage about S&P 500 companies in the *Wall Street Journal* (WSJ) and modeled the interaction between earnings announcements and the return-earnings relation (Chen et al., 2011). Their results show that news coverage has a significantly negative effect on unexpected earning and that higher news coverage decreases the information content of earnings and reduces market responses to unexpected earnings. Two other related studies also try to discover predictive indicators from financial news. Tetlock (2007) quantitatively measured the interactions between WSJ's *Abreast of the Market* column's content and stock market movement and found that high media pessimism predicts downward pressure on market prices, and unusually high or low pessimism predicts high market trading volume. In addition, Tetlock, Saar-Tsechansky, and MacSkassy (2008) found that the fraction of negative words in firm-specific news stories forecasts low firm earnings, and that earnings and return predictability from negative words is largest for news stories that focus on fundamentals. Other research on relating stock market movement to news articles covers the areas of revenue relations (Ma, Sheng, & Pant, 2009), news impact on stock markets (Mahajan, Sk, & Haque, 2008), agent-based intelligent news analysis (Chen, 2010), forecasting intra-day stock price movement using recently released news information (Groth & Muntermann, 2010), predicting changes in stock prices immediately following the publishing of quarterly reports (Mironenko & Durfee, 2007), and predict

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات