Financial News Analytics

Wing Lon Ng
CCFEA, University of Essex, UK

Liutauras Petrucionis
CCFEA, University of Essex, UK

INTRODUCTION

In financial markets, trading is traditionally influenced by information, and, especially in computerised markets, trade execution is nowadays algorithmically triggered by certain signals or events. Market movements can be very sensitive to new information. Therefore, gathering and understanding news announcements of public, company-specific, macroeconomic or political sources is essential for successful investments in general, be it long term or short term. As the sources and volumes of financial news are growing every day, new technologies that automatically collect, extract and categorise news feeds become indispensable for the success of trading in electronic markets, particularly in the high-frequency domain, where orders are expected to be submitted and executed within seconds, or fractions thereof.

The arrival of news influences the market's expectations of future price movements and has a big impact on investor sentiment and behaviour (see also Aldridge, 2010). Innovative messaging models that include textual news data are in demand. Most traders and market participants have access to some kind of news feeds. They have to assimilate news very fast, and, if necessary, revise their trading strategy immediately according to their updated expectations of future market movements. Investment banks and brokerage firms that do not have the means or the expertise to analyse news data can buy news streams and ready-made sentiment indices as provided by business intelligence firms. These news vendors apply text mining techniques to process financial information on a large scale.

BACKGROUND

A human trader's capability to digest financial news "on the fly" is limited in speed and in quantity. However, with increasing computational capacities and faster text mining methods nowadays, new automated news reading and interpreting technologies are available to efficiently extract, aggregate and categorise large volumes of information streams in an instant.

Text mining refers to the application of artificial intelligence to automatically extract information from machine-readable sources (XML, HTML, JSON, and most other electronic documents) by distinguishing and detecting linguistic patterns in written text. Using the accuracy and efficiency of a computer, text mining tools aim to mimic a human's ability to comprehend contextual meaning by looking for repeated patterns, key terms, named entities, or similar subsets thereof in a vast collection of texts within seconds.

In recent years, more and more market participants consider the addition of financial news analytics into their algorithmic trading engine to better predict the direction or volatility of market movements before making an investment decision. Particularly in high-frequency trading, these decisions need to be made almost immediately to reduce latency. Financial news analytics, however, is not just limited to textual data mining exercises. It is a new interdisciplinary research area requiring knowledge and expertise from computer science, finance, and economics. The sole implementation of advanced machine-learning algorithms without the consideration of market microstructure effects and economics of financial markets will have only little value for investors in real-world applications.

Financial news analytics combines methods from information retrieval, statistical learning, natural language processing, and financial econometrics to collect, categorise, interpret unstructured textual input data and convert this into metric output data, such as a financial sentiment score. In the following, we first give an overview of the different common information protocols used in financial applications. We then discuss the implementation of news analytics in an investment strategy in five steps (see also Johnson, 2010): news filtering (what is economically or financially relevant?), news association (what is interesting for which investor?), news interpretation (what does the news mean?), econometric modelling (how can the return and its volatility be statistically modelled?), and strategy testing and implementation (how can the trader capitalise the information?).

Information Protocols in Financial Applications

The battle for nanoseconds in high frequency trading is not only driven by the ultra-low latency technology but also by efficient use of information transmission protocols. The most popular two protocols for the information delivery over electronic networks are TCP and UDP. TCP (Transmission Control Protocol) allows a stable and controlled delivery of data packets over the network. UDP (User Datagram Protocol) does not require a confirmation of the packet delivery and can achieve a lower latency. The use of TCP is preferred for lower frequency messages as delivery is always ensured. The trade-off relation of latency over features is present at all higher level information transmission protocols.

The rise of computers has changed the way how financial news agencies provide real-time information. Financial decision makers can now receive news data in compressed XML (eXtensible Markup Language) format. More explicit content of secondary importance is often delivered via alternative routes, e.g. published on the websites. The popularity of the XML format is due to cross-platform, cross-application and cross-reader accessibility. The cross-reader accessibility refers to the ease of transforming a machine readable message to a human eye appealing presentation and the other two refer to the ability to access the material from different software platforms. The core idea of the XML format is to send a plain text message encoded with a Unicode format and structured with tags. As messages are longer in XML format, such standard is mainly used for a lower frequency data. Many different extensions and variations of this format are also considered in the industry. In financial markets, for example, high frequency data from stock exchanges is often delivered in the FIX (Financial Information eXchange) format. FIX uses a similar idea to structuring messages but the message length is limited and the structure of the packet is fixed. As a result messages are more compact and achieve lower latency of delivery. A more recent extension of FIX is the FAST (FIX Adapted for Streaming) format that allows to transmit a higher quantity of market data in one packet. However, the latency is affected due to additional encoding requirements for each data packet.

In general, news analytics information can be provided in any different machine-readable formats. Sophisticated tagging concepts allow news agencies to tailor their news feeds to customers' individual needs. For example, news data with textual content from financial information providers can also arrive in RSS (Rich Site Summary), ATOM (Atom Syndication Format), a more general XML format, or even in a custom structured raw packets to a socket connection for lower latency of delivery. RSS and ATOM formats are the two most popular methods for information delivery over web feeds and both are based on the XML format making them applicable for algorithmic processing. As a result, processing of such data is a straightforward task.

News Filtering

Today's communication technology allows senders to disseminate their message almost instantly to a broad audience around the globe. Financial news agencies, as well as government institutions, stock exchanges and private corporations are supplying a plethora of financial information such as real-time stock prices, online news wires, accounting reports, or even online conferences through different networks, most prominently the Internet (see also Ettredge et al., 2002). However, not every single word or piece of information is relevant. To filter the vast amount of information coming through the chosen protocols, fast data processing techniques from computational linguistics are required. Sentence

Related Content

Tradeoffs Between Forensics and Anti-Forensics of Digital Images
Priya Makarand Shelkeand Rajesh Shardanand Prasad (2017). *International Journal of Rough Sets and Data Analysis (pp. 92-105).*
www.irma-international.org/article/tradeoffs-between-forensics-and-anti-forensics-of-digital-images/178165

Do We Mean Information Systems or Systems of Information?
Frank Stowell (2008). *International Journal of Information Technologies and Systems Approach (pp. 25-36).*
www.irma-international.org/article/mean-information-systems-systems-information/2531

Intellectual Capital Measurement
Lukasz Bryl (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 5056-5066).*
www.irma-international.org/chapter/intellectual-capital-measurement/184208

Positioning Methods and Technologies in Mobile and Pervasive Computing
Dragan Stojanovic, Billur Barshan, Apostolos Papadopoulos, Nico Van de Wegheand Christophe Claramunt (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 5713-5722).*
www.irma-international.org/chapter/positioning-methods-and-technologies-in-mobile-and-pervasive-computing/113026

From Digital Natives to Student Experiences With Technology
Sue Bennettand Linda Corrin (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 2512-2520).*
www.irma-international.org/chapter/from-digital-natives-to-student-experiences-with-technology/183963