

Event Recognition Strategies applied in the Mercurio Project

Davide Azzalini¹, Fabio Azzalini¹ Davide Greco², Mirjana Mazuran¹, and
Letizia Tanca¹

¹ Politecnico di Milano

{davide.azzalini|fabio.azzalini}@mail.polimi.it

{mirjana.mazuran|letizia.tanca}@polimi.it

² info@davidegreco.it

Abstract. Mercurio is a project currently investigated at Politecnico di Milano whose aim is to support the decision-making process of financial investors. Mercurio identifies relevant events both from financial news articles and financial indexes and uses sequential pattern mining to predict exceptional events given their past occurrences and relationships with other events. The process of event recognition, both from textual and numerical data sources, is crucial to successfully reach the goals.

Investors constantly read financial news and analyze financial indexes, using their knowledge and experience to predict market events and make profitable investments. Mercurio [1, 2] aims at supporting this process by automatically extracting, from data freely available on the Web, *events* that influence and shake the market. Event recognition strategies are applied to both textual and numerical financial information.

Textual data sources. Mercurio monitors Italian financial data sources such as Sole 24 Ore ³, Corriere della Sera ⁴, Radiocor ⁵, etc. These data are processed according to different strategies:

Semantic recognition. Events are recognized through semantic rules that formalize the knowledge of our domain expert. Rules define relationships between sentence structures and events; they are designed to capture meanings that go beyond the sole natural language processing since they recognize “hidden” information inside the news, e.g. financial newspapers, usually, publish interviews when requested by a company: why would a company want to be interviewed? It seems that interviews are often published for reassuring investors in times of crisis.

Classification. Often, data information sources specify, for each article, one or more categories, possibly hierarchical, it belongs to, e.g. articles about balance, merge& acquisition, etc. These categories are mostly general-purpose

³ <http://www.ilsole24ore.com/>

⁴ <http://www.corriere.it/economia/>

⁵ <http://www.radiocor.ilsole24ore.com/>

and hard to use in our context in a significant way. Mercurio employs a domain-specific classification of articles, manually performed by our domain expert on a training set of articles, then automatically derived by the system for the remaining articles.

Communication style. Over time, companies develop a certain communication style characterized by the amount of published news, the different sources dealing with it, the diversity of reported topics, etc.. It is interesting to discover events where a company breaks its expected communication trend, e.g. an out-of-the-blue article breaking a “long” communication silence.

Article summarization. Long articles are hardly read by investors who, instead of reading the whole news, usually skim through its content or trust little more than title and introduction. Mercurio applies summarization techniques to provide investors with only the most relevant information. It ranks each sentence in the article according to its position in the text (at the beginning, end, etc.) and to its content (domain-dependent stop-words as well as significant expressions are kept into account) and constructs a summary that contains only the most informative sentences.

Numerical data sources. Mercurio gathers stock prices from Yahoo! Finance ⁶ and employs technical analysis [3] techniques to determine significant events.

Stock events. Mercurio identifies changes in stock using simple moving averages (sma) of different day lengths. For up/down trends and congestions, sma10, sma20 and sma40 are used, e.g., if the price is above sma10; sma10 is above sma20 and sma20 is above sma40 there is an ongoing uptrend. To find up/down jumps, Mercurio analyzes how sma3 and sma5 change in a time-window of a week. All stock changes are found a posteriori and then used for training.

Candlestick patterns. Japanese candlestick charts ⁷ are used to represent stock prices at possibly different aggregation levels; specific candlestick patterns predict particular market movements. Mercurio uses these patterns in combination with stock events to increment the precision of the event recognition. Moreover, we are investigating the use of patterns at different levels of aggregation, e.g. a candlesticks representing days or weeks.

References

1. S. Marchesini, M. Morlacchi Bonfanti, M. Mazuran, D. Greco, L. Tanca: Mercurio: predicting financial events from expert knowledge and financial news. SEBD 2016: 174-181.
2. M. Brambilla, D. Greco, S. Marchesini, L. Marconi, M. Mazuran, M. Morlacchi Bonfanti, A. Negrini, L. Tanca: Forecasting Out-of-the-Ordinary Financial Events. FINREC 2016: 11-15.
3. J. J. Murphy: Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance, 1999.

⁶ <https://finance.yahoo.com/>

⁷ https://en.wikipedia.org/wiki/Candlestick_pattern