

# Event and Sentiment Detection in Financial Markets

Uta Hellinger

AIFB, Universität Karlsruhe, Germany  
hellinger@aifb.uni-karlsruhe.de

**Abstract.** Today, traders in financial markets are confronted with the problem that information is distributed over diverse sources and that there is too much information available. In our work we develop methods and tools to help traders to overcome this information overload by enabling the integrated view on news from various sources, by filtering relevant news and by providing decision support for traders. Another goal of our work is to propose a formal model of the impact of news on asset prices and thus enable better predictions of stock prices than possible with purely text mining based approaches.

## 1 Research Problem

Traders in financial markets are confronted with the problem that too much information is available from various, heterogeneous sources like newswires, forums, blogs and collaborative tools. In order to make accurate trading decisions, traders have to filter the relevant information efficiently so that they are able to react to new information in a timely manner.

The focus of our work is the development of methods and tools to support traders in this process. The goal is to provide an integrated view on news from different sources, to filter those news that have significant market impact and to help the users to decide how to react to newly published information. The development of these methods and tools raises the following questions:

- How can information from various, heterogeneous sources be integrated? News that are found in various sources differ in their content and in available annotations: news published by newswires are annotated with standardized metadata (which differ between newswires), blog posts are at the most tagged with some keywords. The information published on different web sites has to be collected and metadata has to be mapped to a single format such that all news can be processed using the same algorithms.
- How can the important news be filtered? Users can not monitor all relevant news services and process all the information that is published by these services. Therefore, methods are needed to filter news that have significant market impact. These methods have to detect important events, sentiments and expectations concerning the market.

- How can the users’ trading decisions be supported? Price changes are caused by changes in the expectations concerning a company. Therefore, expectations and precise information (like the amount of the quarterly result) should be used for the prediction of price changes. This requires mechanisms that extract necessary information from texts, formalize it and make predictions based on it.

Although our work focuses on a specific application domain, its results will be relevant for other applications, as we show how information from different sources can be integrated and used to provide decision support.

## 2 Related Work

Our work is related to research in the domains of text mining (especially event and sentiment detection), information extraction, semantic web and finance.

A variety of systems for the prediction of asset price developments based on recently published news have been developed (see [1] for an overview). These systems are based on text classification, where the target categories are derived from financial data. Although they are closely related to our intended application, they have two important weaknesses: (i) expectations, which heavily influence the development of asset prices, and (ii) quantified information (like the value of paid dividends or the amount of the annual profit), which enables the quantification of the expected price change, are not considered in these systems.

Online event detection methods have for example been developed by [2], and [3]. These methods only attempt to identify new events mostly using clustering techniques without trying to formalize them semantically, which is required for matching them against our expectation models.

Work on sentiment detection in a finance context include [4] and [5]. While Das and Chen [4] use linguistic features to classify messages in negative and positive ones and then examine the correlation with stock price changes, Koppel et al. [5] use stock price changes to identify positive and negative news from which then describing features can be extracted.

While to the best of our knowledge no method for modelling expectations exists, Halaschek-Wiener and Hendler [6] have proposed an OWL-based news syndication framework to match publications and information needs. The subscribers’ information needs which are described by conjunctive ABox queries are matched against publications which are formalized as ABox assertions.

Relation extraction is applied to populate ontologies from text - the problem we have to solve for extracting information on events and expectations from news. Bootstrapping usually is applied in these methods, where the web [7] or Wikipedia [8] are used as corpora to find patterns for describing relations. These methods have very low precision and recall in some applications which is problematic for our application.

Event studies study the impact of certain events on a firm’s value (see [9] for an overview of the methodology). These will be useful to find events and aspects that should be taken into account in our system.

### 3 News Analysis Tool

Our solution to the problems discussed above is a tool which offers support in filtering important information and in decision-making. The planned architecture can be seen in Fig. 1. The components of this framework and requirements regarding each of them will be discussed in the following.

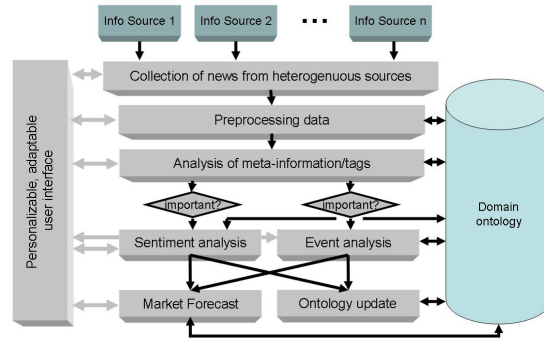


Fig. 1. Preliminary version of news analysis framework

The first component of our tool is responsible for the *collection of news from various, heterogenous sources*. This component will monitor a huge number of relevant sources for new information and will make it available to the *Preprocessing data* component. The latter maps the available metadata to a single representation such that all data is processable in the same way in later steps. We will define an ontology for each news source's metadata and map these ontologies to one general one which will be used for further processing. If no metadata is available, the extraction of some annotations of the news' content like named entities becomes necessary to enable filtering of these news. However, only very efficient methods can be applied here as information extraction is quite expensive in general.

The *Analysis of meta-information/tags* component will examine the metadata to filter relevant news. It will decide whether a news item contains expectations concerning future events and will thus be processed by the *Sentiment analysis* component and whether it contains information on an actual event and will thus be processed by the *Event analysis* component. It is possible that a news item is processed by both components or that a news item is not important and thus will not be processed further.

Both the *Event analysis* and the *Sentiment analysis* component will apply information extraction methods to extract formal descriptions of the news' content. These descriptions will be used by the *Market forecast* component to predict the impact of a news item on the market by quantifying the difference of the actually published information from the expectations and the current status of the

domain as described in the ontology. The *Ontology update* component is responsible for the integration of the changes that occur due to published expectations and events into the ontology.

The *Domain ontology* is the backbone of the three previously described components. It describes the current status of the market and expectations concerning future events. We currently try to identify the most predictive features of news items. We develop a linear regression model that predicts market responses based on text features. This is the technique of choice as the predicted impact can be quantified and as the influence of each feature on the result can easily be seen. The developed model will help us in identifying the information that should be modelled in the domain ontology. It will also provide some prediction facility that can serve as a base line for the evaluation of more elaborate methods.

The tool will be personalisable and adaptable in the sense that users can specify their preferences, e.g. companies that they are especially interested in. This will be possible through the *Personalizable and adaptable user interface*.

An important requirement for the whole process is that it has to be extremely fast as significant price changes (in the short-term trade that we consider) can only be observed within one minute after the publication of a news item by a newswire.

## 4 Evaluation

The framework presented in the previous section requires a set of different evaluations. Firstly, an evaluation of the methods employed in each component has to be done. This especially means that an evaluation of the classification and information extraction methods in terms of precision and recall is required.

As our goal is the prediction of price changes based on expectation changes, the quality of the predictions serve as an evaluation of the domain ontology.

Finally, a user study is necessary to see how well users are supported by this kind of system and whether it helps them to make better trading decisions.

## 5 Work plan

So far, we have acquired news published by Reuters and information on intraday trades and quotes for over 240 markets in 2003 and we have developed aggregation functions for the financial data such that it can be used as training data for the methods we will develop. Given the huge amount of data we focus our experiences on the German market.

As mentioned in section 3 we currently work on the identification of the most predictive features. The next steps will be to compare the features we find to results of event studies available in the finance literature. In parallel to these steps we will develop the metadata ontologies and mappings between them. We will then build a classifier that filters the relevant news based on metadata. Once these methods are available we will build our domain ontology that models events

and expectations, develop the necessary information extraction methods and define how discrepancies between expectations and newly published information can be quantified for predicting the associated asset price changes.

The last component that we develop will be the user interface before we finish our project with user experiments that will hopefully show the benefit of the proposed tool.

## 6 Conclusion

The goal of our work is the development of a news analysis tool that supports traders in financial markets by filtering news and making predictions on the impact of news on the market. The contributions of our work will be:

- refined information extraction methods for the analysis of financial news
- ontologies of the financial domain that allow the formalization of news and their annotations as well as of expectations and events
- a method to quantify the distance of an event from expectations

## 7 Acknowledgements

This work was funded by the German Research Foundation (DFG) in scope of Graduate School Information Management and Market Engineering.

## References

1. Mittermayer, M.A., Knolmayer, G.: Text mining systems for market response to news: A survey. Working paper (2006)
2. Zhang, J., Ghahramani, Z., Yang, Y.: A probabilistic model for online document clustering with application to novelty detection. In: Neural Information Processing Systems Conf. (2004)
3. Makkonen, J., Ahonen-Myka, H., Salmenkivi, M.: Applying semantic classes in event detection and tracking. In Sangal, R., Bendre, S.M., eds.: Proc. of Int. Conf. on Natural Language Process. (2002) 175–183
4. Das, S., Chen, M.: Yahoo! for amazon: Sentiment extraction from small talk on the web. Management Science **53**(9) (2007) 1375–1388
5. Koppel, M., Shtrimberg, I.: Good news or bad news? let the market decide. Computing Attitude and Affect in Text: Theory and Applications (2006) 297–301
6. Halaschek-Wiener, C., Hendler, J.: Toward expressive syndication on the web. In: WWW '07: Proc. of the 16th Int. Conf. on World Wide Web, ACM (2007) 727–736
7. Pantel, P., Pennacchiotti, M.: Espresso: leveraging generic patterns for automatically harvesting semantic relations. In: ACL '06: Proceedings of the 21st Int. Conf. on Computational Linguistics and the 44th annual meeting of the ACL. (2006) 113–120
8. Blohm, S., Cimiano, P.: Using the web to reduce data sparseness in pattern-based information extraction. In: Proc. of the 11th European Conf. on Principles and Practice of Knowledge Discovery in Databases, Springer (2007) pp. 18–29
9. MacKinlay, A.C.: Event studies in economics and finance. J. of Economic Literature **35**(1) (March 1997) pp. 13–39