Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Multicriteria Decision Making (MCDM 2007)

Prediction of Stock Price Movements Based on Concept Map Information

Ankit Soni * Nees Jan van Eck† Uzay Kaymak†

*Department of Computer Science, Indian Institute of Technology Kanpur, Kanpur, India
Email: ankit@iitk.ac.in

†Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
Email: {nvaneck, kaymak}@few.eur.nl

Abstract— **Visualization of textual data may reveal interesting properties regarding the information conveyed in a group of documents. In this paper, we study whether the structure revealed by a visualization method can be used as inputs for improved classifiers. In particular, we study whether the locations of news items on a concept map could be used as inputs for improving the prediction of stock price movements from the news. We propose a method based on information visualization and text classification for achieving this. We apply the proposed approach to the prediction of the stock price movements of companies within the oil and natural gas sector. In a case study, we show that our proposed approach performs better than a naive approach and a bag-of-words approach.**

I. Introduction

Data visualization is an important activity when studying and analyzing large amounts of data. Using data visualization techniques, decision makers can familiarize themselves with various aspects concerning their data sets and use this improved insight for making complex decisions. For this purpose, advanced visualization platforms are developed that can deal with data of different modalities from different sources [1]. Nowadays, mining of textual information is an important research area, since such sources are readily available and possess rich information content that can be used for improved decision making and advanced data modeling. Consequently, visualization of these data is also important.

One of the methods used for visualizing textual data is the concept map. Concept maps are often used in knowledge domain visualization, which is concerned with the construction of maps that visualize the structure and the evolution of a field of science [2]. In knowledge domain visualization, concept maps visualize the associations between concepts in a scientific field. In general, concept maps are used for visualizing the associations between concepts in a collection of documents. The associations between concepts can be measured in various ways, but it is often quantified based on the number of times concepts co-occur in the same documents. In concept maps, concepts that are associated more strongly with one another are placed closer to one another in the map than the concepts that are not strongly associated [3]. As such, these maps can be considered to group related concepts.

Because concept maps group related concepts, they also discover some structure within the collection of documents considered. An interesting question is then whether this discovered structure could be used in additional ways to improve decision making. In other words, what additional (implicit) information do such maps convey? Concept maps can be expected to convey implicit information, because they can be thought of summarizing similarity-related information that could entail multiple criteria. A concept map is a simple visualization of this information in a metric space with a small number of variables. In this paper, we investigate whether the locations of news items in a concept map could be used as features for a predictive algorithm. We conduct our study within the field of stock market analysis. The assumption behind our approach is that related concepts may have similar effects on the decision making. There are applications where this assumption might hold true. For example, related concepts in news items might reflect the sentiment of the market, making it possible to couple this information to the actionable knowledge.

In this paper, we choose stock market analysis as the application field to illustrate our ideas. We propose a method to classify news items by using features derived from a special concept map and use them to predict stock price movements. In a case study, we show that our proposed approach performs better than a vector-based approach.

The outline of the paper is as follows. In Section II, we give background information about predicting stock price movements. In Section III, we describe our approach to the classification of news for stock price movement prediction. In Section IV, we describe the application of our approach to the prediction of stock price movements of a group of eleven companies from the oil and natural gas sector. Some conclusions are drawn in Section V.

II. News and Stock Price Movements

The dynamics of stock prices are determined by the trading behavior of financial actors in the market. Traders form an important class of actors in the financial markets. The behavior of a trader, e.g. the buy and sell decisions that the trader makes, depends to a large extent on the information that is available to the trader. Public information and news are known to be important information resources for a trader [4]. Hence, researchers argue that there is a relation between news and the prices of the stocks.

With the advent of information revolution, we have a large amount of data and information about companies. It can be expected that a careful analysis of the news reveals much implicit knowledge about the underlying value of a financial asset, which could be used by humans to make more effective buying and selling decisions regarding that asset. However, the volume of information and its frequency of arrival has grown in such large numbers that it is virtually impossible for humans to analyze and make use of all the information. Hence, automatic processing and decision support methods are required.

The relation between news and its effect on the stock market has been the subject of a number of studies. In one study, the users have been presented with a set of "important" stories after which the user could decide on their effect [5]. A multi-agent automated system based on this manual labeling has shown a positive correlation between the news reports on a company's financial outlook and its attractiveness as an investment. Another study describes an automated system which correlates news items to trends in the financial time series [6]. The news analysis in most works is at a shallow level, concentrating on the analysis of the news headlines only [7], [8]. Even when the analysis covers the body text of the news items, most of the past work uses standard bag-of-words based models, which fail to cover the semantics of the news [6], [9]. Semantics of the news could be important, however. This paper contributes to the literature by focusing on the semantic information captured by single-word and multi-word terms in a concept map. Headline information only might not be sufficient to achieve acceptable level of performance [8], [10], and so our analysis is based on the concepts that occur in the whole news text.

Stated briefly, we analyze, in this paper, the effect of news on stock price movements. More specifically, we address the problem of predicting whether the stock prices increase or decrease as a result of arriving news items. We do this by considering the problem as a text classification problem. Text classification (also called text categorization) (e.g., [11]) is the process of classifying text documents into predefined categories based on their content. There are four main approaches to text classification, namely the manual approach, the rule-based approach, the supervised learning approach, and the unsupervised learning approach. In the manual approach, documents are classified by hand. In the rule-based approach, documents are classified using rules that are predefined. In the supervised learning approach, documents are also classified using rules but now these rules are first learned on the basis of a set of documents for which the category is known. Popular methods that are used in the supervised approach are naive Bayes, decision trees, nearest neighbor, and support vector machines (e.g., [12]). In the unsupervised learning approach, classification rules are also learned but now on the basis of a set of documents for which the category is unknown. Clustering methods are typically used in the unsupervised approach.

In the following, we take a supervised learning approach to the classification of news for stock price movement prediction. In our analysis, we focus on a group of eleven companies in the oil and natural gas sector. All these eleven companies are included in the AMEX index. During a time period of about eleven years, the daily stock prices of the companies are collected. For the same time period, news items that are relevant for the companies are retrieved from the Financial Times.

III. Classifying News with Concept Map Features

Our approach to the classification of news for stock price movement prediction can be divided into four main steps:

1) collecting and labeling of news items,
2) feature extraction,
3) classifier induction, and
4) classifier evaluation.

We now discuss each step of our approach in more detail.

The first step of our approach is the collecting and labeling of news items. In our analysis, we focus on a group of companies that are included in a stock market index. During a certain time period, for each of these companies the daily stock prices are collected. For the same time period, news items that are relevant for the companies are retrieved from a news archive. Each news item that is relevant for a company is then labeled according to the movement of the company's stock price. We make a distinction between the categories positive and negative. When the company's stock price increases on the day on which the news item has been published, the news item is labeled as positive. When the stock price decreases or stays the same, the news item is labeled as negative. All the labeled news items are divided into a training collection and an independent test collection.

The second step of our approach is that of feature extraction. The goal of feature extraction is to make the classification task easier. Generally, feature extraction improves the classification accuracy and the computational efficiency. In our approach, we aim to add a semantic dimension to our features. We identify related concepts in news items and then classify the news items by using their content in terms of the concepts they contain. This is done as follows. First, concepts are identified in the news items. For this, we make use of the term extraction tool presented in [13]. Second, the association strengths between all pairs of identified concepts are calculated. The association strength of two concepts is based on the co-occurrence frequency of these two concepts, i.e., the number of news items in which both concepts occur [14]. Mathematically, the association strength a_{ij} of the concepts i and j is defined as

$$a_{ij} = \frac{mc_{ij}}{c_{ii}c_{jj}} \qquad \text{for } i \neq j, \tag{1}$$

where c_{ij} denotes the number of news items in which the concepts i and j both occur, c_{ii} denotes the number of news items in which concept i occurs, and m denotes the total number of news items. The association strength of two concepts can be interpreted as the estimated co-occurrence frequency of the concepts normalized for the expected co-occurrence frequency of the concepts obtained under the assumption that occurrences of the concepts are statistically independent [14]. Finally, the identified concepts are positioned in a so-called concept map based on their association strengths. A concept map is a low-dimensional space in which concepts are positioned in such a way that the distance between the concepts reflects their association strength. In this paper, the positioning of concepts is performed using a method that we call VOS [15]. A brief discussion of VOS is provided in Paragraph III-A. Fourth, the original news items are positioned in a so-called document map based on the coordinates of the concepts in the concept map. A document map is a low-dimensional space in which each news item is positioned on the weighted average of the coordinates of the concepts that occur in the news item. Finally, the coordinates of the news items in the document map are taken as features.

The third step of our approach is that of classifier induction. This step is concerned with the construction of a classifier on the basis of the labeled news items in the training collection. In this paper, this step is performed using support vector machines (SVMs) (e.g., [16]). SVMs are used to model the category of a news items on the basis of its coordinates in the low-dimensional space that was generated in the third step of our approach. A brief discussion of SVMs is provided in Paragraph III-B.

The fourth step of our approach is that of classifier evaluation. In this step, the performance of the constructed classifier is measured. This is done by applying the classifier to the news items in the independent test collection and by measuring the hit rate, i.e., the percentage of correctly classified news items from the test set.

A. VOS

In this paragraph, we briefly discuss the general idea of VOS. For a more elaborate discussion of VOS, including an analysis of the relationship between VOS and another commonly used visualization method, multidimensional scaling (MDS), we refer to [15].

VOS aims to provide a low-dimensional visualization in which objects are located in such a way that the distance between any pair of objects reflects their similarity as accurately as possible. In our case, objects correspond to concepts and similarities correspond to associations. The objective function minimized in VOS is given by

$$E(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{i<j} a_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm and $\mathbf{x}_i$ is the location of the concept i in the low dimensional space. The minimization of the objective function is performed subject to the constraint

$$\frac{1}{n(n-1)} \sum_{i<j} \|\mathbf{x}_i - \mathbf{x}_j\| = 1. \tag{3}$$

VOS has the following three properties. First, VOS has the tendency to locate objects close to what we call their ideal coordinates. The ideal coordinates of an object i are defined as a weighted average of the coordinates of all other objects, where the coordinates of objects more similar to object i are given higher weight in the calculation of the weighted average. Second, VOS seems to pay more attention to indirect similarities via third objects than MDS. For example, if two objects i and j have a similarity of 0, the distance between the objects in a visualization obtained using VOS seems to depend on the number of third objects with which the objects i and j both have a positive similarity. The higher the indirect similarity via third objects, the closer the objects i and j are located to each other. Third, although VOS and MDS may provide very different visualizations, VOS is, under certain conditions, equivalent to a special variant of MDS called Sammon mapping. Furthermore, if weights are used in MDS and these weights are chosen in an appropriate way, then there also exists an equivalence, under certain conditions, between VOS and standard MDS.

B. SVM

In this paragraph, we briefly discuss the general idea of support vector machines (SVMs). For a more elaborate discussion of SVMs, we refer to, e.g. [16].

SVMs are based on the concept of decision planes that define decision boundaries. In the original version, they belong to a family of generalized linear classifiers. The main idea of SVMs is finding a hyperplane that separates two classes with a maximum margin. Consequently, SVMs are also known as maximum margin classifiers. The training problem in SVMs can be represented as a quadratic programming optimization problem. SVMs can be extended to nonlinear classifiers by applying kernel mapping (the so-called kernel trick). As a result of applying the kernel mapping, the original classification problem is transformed into a higher dimensional (possibly infinite dimensional) space. SVMs which represent linear classifiers in this high-dimensional space may correspond to nonlinear classifiers in the original feature space [17]. The kernel function used may influence the performance of the classifier. In our analysis, we have used the standard linear kernel function.

IV. Analysis

In this section, we provide details of the analysis that we have performed with our proposed approach to the classification of news for stock price movement prediction. Since we expected that the oil and natural gas sector is influenced largely by news, we confined the analysis to that sector. We studied companies from the AMEX index, since it includes the larger companies in the sector, for which the largest number of news items is available. Still, there were very few news items for

TABLE I
NUMBER OF NEWS ITEMS FOR EACH STUDIED COMPANY IN THE DOCUMENT COLLECTION.

Company name	Number of news items
Anadarko Petroleum	28
BP Amoco PLC	948
Chevron	330
Conoco Philips	224
Exxon Mobil Corporation	500
Halliburton	207
Occidental Petroleum Corporation	107
Repsol YPF SA	204
Schlumberger	102
Shell	641
Total Fina SA	202
	3493

some of the companies in the index. These companies were removed from the analysis. Furthermore, one company for which we did not have financial time series data was removed from the analysis and the list was augmented with some other sector companies such as Schlumberger, Anadarko Petroleum, Chevron, and Halliburton. After these changes, we ended up with a group of eleven companies. The discussion in the rest of this section follows the four steps mentioned in Section III.

A. *Collecting and Labeling of News Items*

For each of the eleven companies studied, we collected news items by using the Financial Times online service FT Intelligence [18]. FT Intelligence offers instant access to a range of key sources from one of the world's leading providers of accurate and unbiased financial, economic, and political news. Using FT Intelligence, news items can be searched based on key words, company, industry, country, and news category. Each retrieved news item has a time stamp indicating the day on which the news item was released. We included news from January 1, 1995 until May 15, 2006 for our study. In total, we collected 3493 unique news items for the eleven companies in our study. The distribution of the news items across the companies is shown in Table I.

Time series data of the eleven companies were used for labeling the news items. We used the DataStream service [19] for collecting time series data. The DataStream service provides a lot of information on stocks including the opening price, daily high price, daily low price, market index, and market volume. Since DataStream does not provide daily closing prices of stocks, opening prices were used to label the news items. If the opening price on day $k+1$ is higher than the opening price on day k, then the news occurring on day k is assumed to have caused a positive effect on the market and is labeled as positive. If the opening price on day $k+1$ is lower or equal than the opening price on day k, the news occurring on day k is assumed to have caused a negative effect on the market and is labeled as negative. In this way, the news items were divided into two categories: positive and negative. The two categories were distributed roughly equally.

TABLE II
NUMBER OF POSITIVE AND NEGATIVE LABELED NEWS ITEMS IN THE TRAINING COLLECTION AND THE TEST COLLECTION.

Label	Training collection	Test collection	Total
Positive	1357	367	1724
Negative	1437	332	1769
	2794	699	3493

The 3493 news items that were collected were divided into a training collection and a test collection. The training collection contains 2794 news items which correspond with the oldest 80% of all the news items. The test collection contains 699 news items which correspond with the most recent 20% of all the news items. The distribution of the news items across the training and test collections, and the positive and negative categories is shown in Table II. We note that the distribution across the two categories in the test collection differs considerably from the distribution across the two categories in the training collection.

B. *Feature Extraction*

As discussed in Section III, the feature extraction step of our approach is subdivided into a number of steps. Next, we discuss these steps.

1) Identification of concepts and calculation of association strengths: Concepts in news items might reveal information about the market sentiments. To identify concepts in the news items that were collected, we made use of a thesaurus. We constructed the thesaurus ourselves using the term extraction tool presented in [13]. Using the tool, we extracted 4973 concepts from the news items. A lot of concepts that were extracted might not be very relevant in the classification task. For example, a very frequent concept encountered is the sector-specific concept *oil*, which provides little extra information in our classification task. Hence, a filtering approach was taken to reduce the number of concepts by removing infrequent and non-discriminatory concept. After some experimentation, we removed all concepts (except the company names) that occur in less than 8 news items or whose percentage of occurrence in both the positive and negative labeled news items is more than 25% of their total occurrence in all the news items. In this way, the number of concepts was reduced from 4973 to 367. The association strengths of the 367 remaining concepts were calculated and stored in an association strength matrix.

2) Construction of the concept map: Using VOS, a two-dimensional concept map was constructed based on the association strength matrix. The concept map is shown in Figure 1. Concepts are located such that the distance between two concepts reflects the strength of their associations in the news items. The importance of a concept is indicated by the size of its label, and the distribution of occurrence of a concept in the positive and negative labeled news items is indicated

TABLE III

CONTINGENCY TABLE SHOWING THE CLASSIFICATION RESULTS OF OUR APPROACH ON THE TEST COLLECTION.

Correct label	Predicted label		Total
	Positive	Negative	
Positive	157	210	367
Negative	96	236	332
	253	446	699

TABLE IV

CONTINGENCY TABLE SHOWING THE CLASSIFICATION RESULTS OF THE SVM BAG-OF-WORDS APPROACH ON THE TEST COLLECTION.

Correct label	Predicted label		Total
	Positive	Negative	
Positive	147	220	367
Negative	136	196	332
	283	416	699

by the color of the concept label. Green is used for positive labeled news items and red is used for negative labeled news items. In Figure 1, it can be seen that the concepts that occur more frequently in positive labeled news items (indicated by the green color) and the concepts that occur more frequently in negative labeled news items (indicated by the red color) are separated quite well.

A more detailed description of the construction and interpretation of concept maps has been published elsewhere and can be found in [14], [20].

3) Construction of the document map: The 3493 news items that had been collected were positioned in a two-dimensional document map. The location of a news item in this document map was calculated by taking the average of the locations in the concept map of the concepts that occur in the news item. The locations of the news items in the document map are shown in Figure 2. Green and red dots refer to positive and negative labeled news items from the training collection, while green and red plusses refer to positive and negative labeled news items from the test collection. It can be seen that positive and negative labeled news items are not separated well. Hence, we are dealing with a hard classification task.

C. Classifier Induction

Using the SVM implementation that is available in the LibSVM package [21], the relationship between the location in the document map of the 2794 news items from the training collection and the labels of these news items was modeled. The only two input feature vectors that were used correspond to the two-dimensional coordinates of the news items in the document map. Furthermore, we used the standard linear kernel function.

D. Classifier Evaluation

We evaluated the constructed classifier by applying it to the 699 news items from the test collection. The classification results that were obtained on the test collection are reported in Table III. The hit rate on the test collection, i.e. the percentage of correctly classified news items from the test collection, was $(157 + 236)/699 = 56.2\%$.

To get an indication of how good this result is, we compared the hit rate of our classifier with the hit rate of a naive classifier and an SVM bag-of-words classifier. The naive classifier classified all the news items from the test collection as negative, i.e. the category to which most news items in the training set belong (see Table II). The hit rate of the naive classifier on the test collection was $332/699 = 47.5\%$. We note that this bad performance is due to the difference in the proportion of positive and negative labeled news items between the training collection and test collection. The SVM bag-of-words classifier was constructed by following the approach described in [22]. First, TF-IDF bag-of-words vectors were generated for the news items in the training collection using the WVTool [23]. The TF-IDF is a statistical measure from the information retrieval field that is used to evaluate how important a word is to a document. An SVM was trained using a linear kernel function and the TF-IDF bag-of-words vectors as input features. The classification results of the SVM bag-of-words based classifier on the test collection are reported in Table IV. The hit rate of the SVM bag-of-words classifier on the test collection was $(147 + 196)/699 = 49.1\%$. We used a two-proportion z-test to test the significance of the difference in hit rate between our classifier and the two other classifiers. From the z-test it resulted that at the 0.01 level the hit rate of our classifier was significantly higher than the hit rate of the naive classifier and the hit rate of the SVM bag-of-words classifier.

V. DISCUSSION AND CONCLUSIONS

Visualization of textual data can reveal interesting information upon which decision making and prediction algorithms can be based. Concept maps visualize, in a low dimensional space, concepts that are associated with one another. We have used the locations of documents in a concept map to train classifiers for predicting the influence of news on stock prices in the oil and natural gas sector. Our results indicate that the use of features from visualization can lead to better classifiers than other text classification approaches such as the bag-of-words method combined with an SVM classifier. The proposed approach improved the classification accuracy from 49.1% to 56.2% compared to the SVM bag-of-words approach. Although 56.2% is not a high classification rate, given the difficulty of the classification task that we studied, it establishes the potential of our proposed approach.

The classification task is a difficult one, because the different news items located on the concept map are not well separated in the feature space of the concept map. The concept map showed a better separation of positive and negative effects for the concepts than for the news items. This indicates that many news items contain both positive and negative concepts, suggesting a potential for a more detailed semantic analysis of news items. In a future study, we intend to investigate

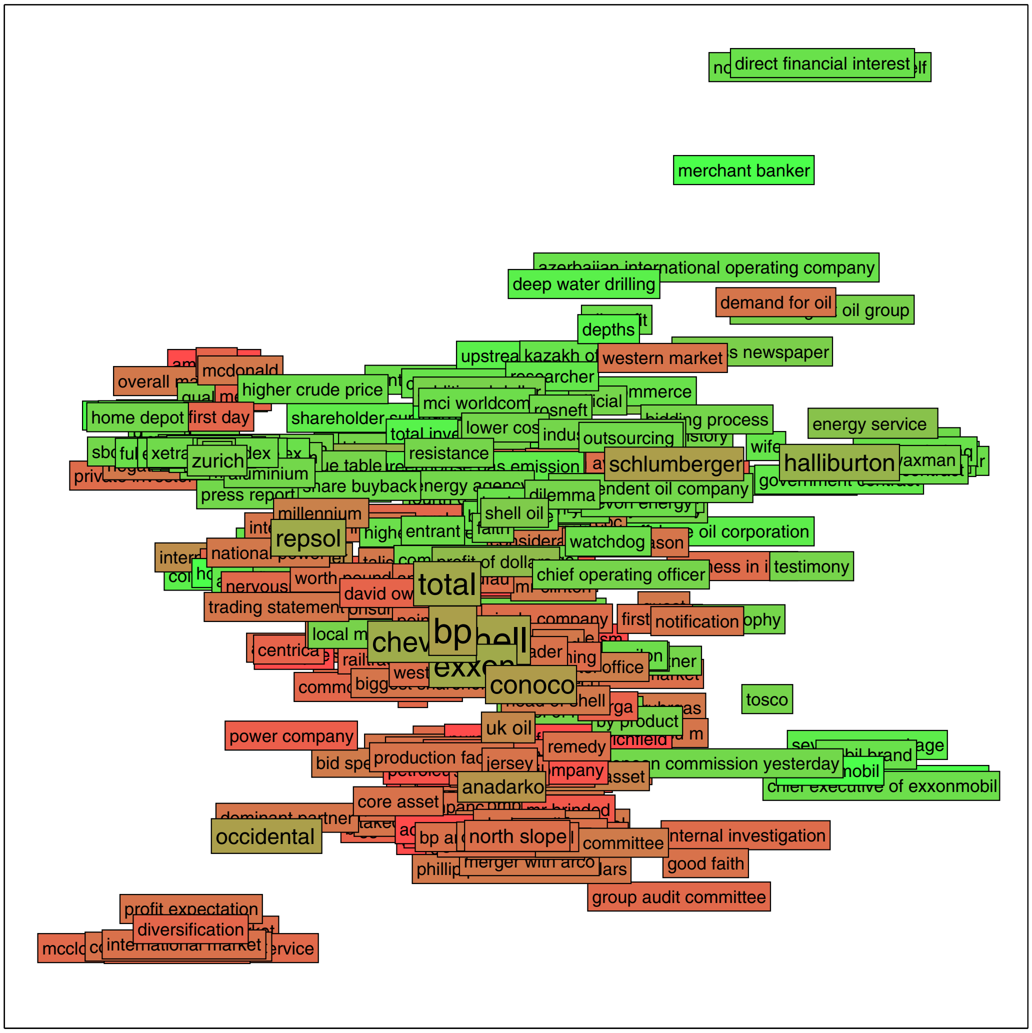

Fig. 1. Concept map.

how the concept groupings in the concept map relate to market sentiments and whether these sentiments can be traced dynamically as they change in time.

An important limitation of this work has to do with the information collected for the news items. In our study, only date information is available for each news item. We do not have data on the specific time at which a news item was released. Because of this, it is likely that some of the news items were published after (rather than before) a change in the price of a stock had taken place. Obviously, such news items cannot be used to predict stock price movements. In this study, we implicitly assumed that stock price movements took place after the publication of news items. This assumption may not always be correct. In future research, more detailed information on news items should be collected, in particular the precise time at which items were released as well as the stock price at that time.

ACKNOWLEDGEMENTS

Part of this work was supported by SENTER (Dutch Ministry of Economic Affairs) through a grant to the VICORE project (project number TSIT3024).

REFERENCES

[1] J. van den Berg, N. J. van Eck, L. Waltman, and U. Kaymak, "A VICORE architecture for intelligent knowledge management," in *Proc. KDNet Symposium on Knowledge-Based Services for the Public Sector*, 2004, pp. 63–74.

[2] K. Börner, C. Chen, and K. W. Boyack, "Visualizing knowledge domains," *Annu. Rev. Inform. Sci.*, vol. 37, pp. 179–255, 2003.

[3] N. J. van Eck, F. Frasincar, and J. van den Berg, "Visualizing concept associations using concept density maps," in *Proc. 10th Int. Conf. Information Visualisation*, 2006, pp. 270–275.

[4] M. L. Mitchell and J. H. Mulherin, "The impact of public information on the stock market," *J. Financ.*, vol. 49, no. 3, pp. 923–950, 1994.

[5] Y.-W. Seo, J. Giampapa, and K. Sycara, "Financial news analysis for intelligent portfolio management," Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-04-04, 2004.

[6] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Language models for financial news recommendation," in *Proc. 9th Int. Conf. Information and Knowledge Management*, 2000, pp. 389–396.

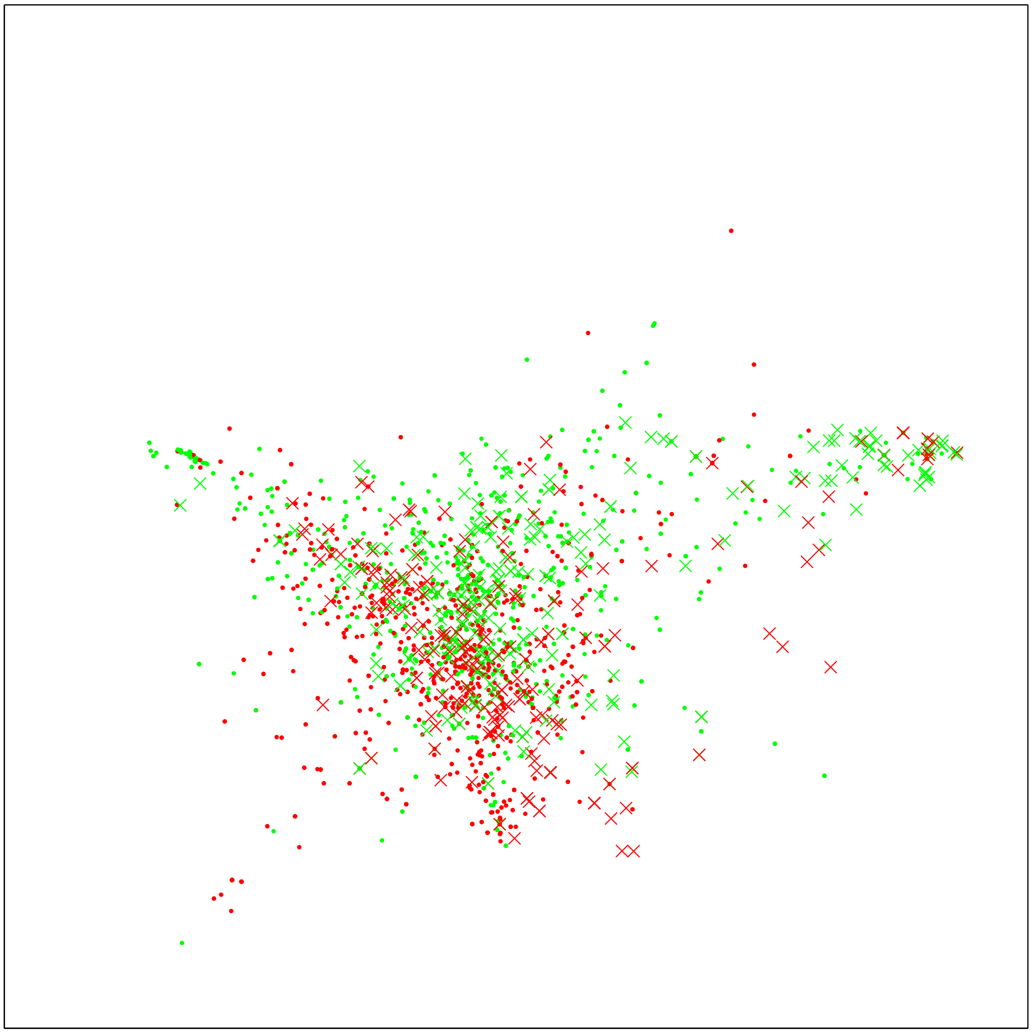

Fig. 2. Document map.

[7] V. Daskalopoulos, "Stock price prediction from natural language understanding of news headlines," in *Proc. 1st instructional Conf. Machine Learning*, 2003.

[8] J. D. Thomas, "News and trading rules," Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, 2003.

[9] S. A. Macskassy, F. Provost, H. Hirsh, R. Sankaranarayanan, and V. Dhar, "Intelligent information triage," in *Proc. 24th Ann. Int. Conf. Research and Development in Information Retrieval*, 2001, pp. 318–326.

[10] P. Kroha and R. Baeza-Yates, "Classification of stock exchange news," Fakultät für Informatik, Technische Universität Chemnitz, Tech. Rep. Chemnitzer Informatik-Berichte, CSR-04-02, 2004.

[11] P. Jackson and I. Moulinier, *Natural Language Processing for Online Applications: Text Retrieval, Extraction, and Categorization*. Amsterdam, The Netherlands: John Benjamins Publishing Company, 2002.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, USA: Springer-Verlag, 2001.

[13] N. J. van Eck, "Towards automatic knowledge discovery from scientific literature: Computer based tools for supporting scientific research," Master's thesis, Erasmus University Rotterdam, 2005.

[14] N. J. van Eck, L. Waltman, J. van den Berg, and U. Kaymak, "Visualizing the WCCI 2006 knowledge domain," in *Proc. 2006 IEEE Int. Conf. Fuzzy Systems*, 2006, pp. 7862–7869.

[15] N. J. van Eck and L. Waltman, "VOS: a new method for visualizing similarities between objects," in *Proc. 30th Ann. Conf. German Classification Society*, 2006, accepted for publication.

[16] V. Vapnik, *The Nature of Statistical Learning Theory*, ser. Statistics for Engineering and Information Science. Berlin, Germany: Springer Verlag, 2000.

[17] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.

[18] "FT intelligence," http://ft.chadwyck.co.uk.

[19] "Datastream service," http://www.datastream.com.

[20] N. J. van Eck, L. Waltman, J. van den Berg, and U. Kaymak, "Visualizing the computational intelligence field," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 6–10, 2006.

[21] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[22] T. Joachims, "Text categorization with support vector machines: learning with many relevant features," in *Proc. 10th European Conf. Machine Learning*, 1998, pp. 137–142.

[23] M. Wurst, "The word vector tool - a Java library for creating vector representations of text documents," Department of Computer Science, University of Dortmund, 2005, software available at http://sourceforge.net/projects/wvtool.