

# Do Media Sentiments Reflect Economic Indices?

Paul Hofmarcher, Stefan Theußl, Kurt Hornik  
WU Wirtschaftsuniversität Wien, Vienna, Austria

Forecasting economic indices on the basis of information extracted from text documents, like newspaper articles is an attractive idea. With the help of text mining techniques, in particular sentiment analysis, we evaluate the tone of individual *New York Times* (NYT) articles and compare our results to the Chicago Fed National Activity Index (CFNAI). In this paper, we present a simple, intuitive framework to derive sentiment scores from text documents. In particular articles are tagged based on terms and their connotated sentiment. Subsequently, we forecast the CFNAI movements via support vector machines (SVM) trained on a subset of the observed sentiment scores. We apply our model into two different data sets, the whole NYT articles and the articles categorized as NYT business news. On both data sets, we applied a simple performance measure to evaluate forecasting accuracy of the CFNAI.

**Keywords:** text mining, sentiment analysis, support vector machines (SVM), forecasting

## Introduction

Many agents in an economy, like economists, investors and politicians avidly follow monthly economic ratios, like unemployment or inflation rates. Usually, such ratios are published by institutions with a lag of one or several months and are then considered in their daily decision making process.

Beside this, agents constantly update this information via extracting relevant facts from daily newspaper articles. Therefore, the coherence between sentiments derived from media coverage and economic “welfare” is an important question. Doms and Morin (2004) identified three channels, how media news affect the expectations about the economy: First, the media covers the economic data and the outlook of economic development. Second, through the tone of economic articles individuals get an estimate about the significance of the topic. Third, news coverage may influence people in their consumption behavior as well as in their expectations about economic development.

In this paper, we try to investigate the relationship between published newspaper articles from the *New York Times* (NYT) and economic indices, more specifically the Chicago Fed National Activity Index (CFNAI), for a time period of more than 20 years. To study the coherence between the sentiment derived from NYT articles and the observed CFNAI, we use text mining techniques and subsequent sentiment analysis. The usage

---

Paul Hofmarcher, M.Econ, M.Math, research assistant, Institute for Statistics and Mathematics, WU Wirtschaftsuniversität Wien.

Stefan Theußl, M.S., research assistant, Institute for Statistics and Mathematics, WU Wirtschaftsuniversität Wien.

Kurt Hornik, Ph.D., professor, Institute for Statistics and Mathematics, WU Wirtschaftsuniversität Wien.

Correspondence concerning this article should be addressed to Paul Hofmarcher, WU Wirtschaftsuniversität Wien, Augasse 2-6 1090 Wien, Austria. E-mail: paul.hofmarcher@wu.ac.at.

of such statistical and machine learning methods, has been widely used in the last decade, both in academics and business intelligence. To name but a few, e.g., Askitas and Zimmermann (2009) find a strong relationship between keyword searches and unemployment rates using monthly data of the German economy; Li and Wu (2010) use text mining techniques and sentiment analysis to detect hotspots of online forums; Scharl and Weichselbraun (2008) present an automated approach to investigate the online media coverage of US Presidential Elections, and Das and Chen (2007) develop a methodology for extracting small investor sentiments from stock message boards.

With the help of support vector machines (SVM) trained on derived sentiment scores of the NYT articles, we try to find evidence how the sentiments reflected by those articles can be used to forecast the development of such an index. In doing so, we consider several cases. The first case uses the whole NYT articles (politics, sports, business-economics, lifestyle, etc.). We refer to this approach as the “uninformative” approach, since a priori we do not differ between the informational content of the article categories. Therefore, life style articles may also mirror the economic welfare like business articles. In the second approach, we only consider business news articles, as descriptors for the economic development. Finally, the last two approaches use the monthly differences in the estimated sentiments to describe the development of the welfare index. These approaches are based on the hypothesis that media report more about changes of the underlying theme than always reporting the current state, e.g., a very low unemployment rate will be lauded for a few times, and media will disregard this theme until unemployment starts to rise. In the context of forecasting (financial) time series via SVM, we refer to Cao and Tay (2001); Huang, Nakamori and Wang (2005).

This paper is organized as follows: The next section briefly reviews the used methodology. Then we describe the data employed. Finally, we present the results before concluding this paper with a discussion for further research.

## Methodology

### Sentiment Classification

The first part of the sentiment analysis deals with the extraction of opinions in order to derive a sentiment score for each article in the sample. Those scores may be aggregated on a daily, weekly or monthly basis in order to make them comparable with the used economic index. In our analyses, we chose to work with monthly sentiment scores since we compared them to monthly economic CFNAI data. Sentiments reflected by a given article are derived in a three-step approach: (1) calculating term frequencies (TF); (2) tagging terms according to its sentiment; and (3) aggregating TF based on tag and date. In detail, we retrieve TF from a so-called document term matrix (DTM). TF were calculated for each (preprocessed) document  $d$  denoted by  $tft$ ,  $d$  representing the number of occurrences of term  $t$  in  $d$ . This is also known as the bag of words model. Subsequently, terms have to be tagged according to their connotated sentiment, e.g., via general inquirer (GI) tag categories which are a collection of the following sources: the Harvard IV-4 dictionary, the Lasswell value dictionary, several categories recently constructed, and “marker” categories primarily developed as a resource for disambiguation. GI is said to be appropriate for developing scoring procedures that enlist multiple categories and analyze the tag patterns assigned to each sentence. This tag assignment pattern across sentences could then be used to generate an overall

assessment of a document (retrieved from <http://www.wjh.harvard.edu/~inquirer/>)<sup>1</sup>.

Finally, based on the DTM and the tagged words, we derive sentiment scores, which are referred to as the polarity. These scores depend on the counted positive  $p$  and negative  $n$  connotations within the considered month.

Formally, at any time  $t$ , the polarity score is given as:

$$score_{pol}(t) = \frac{p(t) - n(t)}{p(t) + n(t)} - \sum_{t=1}^T \frac{p(t) - n(t)}{p(t) + n(t)}$$

i.e., the ratio of the difference between the counted positive and negative words over the sum of these words minus the long term mean of this ratio. For  $p = n = 0$ , we define  $score_{pol} = 0$ . By its definition, the polarity score is appropriate to distinguish between periods including an above average number of positive connotated words, and periods accompanied by more negative connotated terms. So, under the hypothesis that newspaper articles reflect in a rough sense the current state of an “economy”, crisis should come along with a concentration of negative connotated words, and boom conditions should be identifiable by bulks of positive connotated words.

### Forecasting via SVM

Support Vector Machines (SVM) perform classification by constructing hyper planes that optimally separate the data. An implicit mapping function  $\Phi$  maps input data into a (higher dimensional) feature space—where learning takes place—defined by a kernel function, i.e., a function returning the inner product between two data points in the feature space<sup>2</sup>. If we use the projection  $\Phi: X \rightarrow H$ , the inner product  $\langle \Phi(x), \Phi(y) \rangle$  is represented by the kernel  $k$  via:

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle$$

In our experiment we rely on  $R$ , a software package for statistical computing and graphics (version 2.11.1), and its extension package kernlab with respect to SVM classification as this package provides several kernel implementations, e.g., linear kernel, Gaussian radial basis function or polynomial kernels.

As the linear kernel and the hyperbolic tangent kernel are widely used for text categorization and neural networks (Karatzoglou, Meyer, & Hornik, 2006), we use these kernels to forecast the direction of the CFNAI. For both kernels, we evaluate the hit-ratio, i.e., the success rate of forecasting the direction of the index development of each kernel.

## Data

### New York Times

In our application, we investigated how the sentiment reflected by documents in the *New York Times* Annotated Corpus can be used to forecast the CFNAI. The corpus contains over 1.8 million short-to-medium-length articles (on average 3,303.68 characters per document) published by NYT between 1987-01-01 and 2007-06-19 (Sandhaus, 2008). Documents and corresponding metadata are provided in an XML like format: News Industry Text Format (NITF). We used  $R$  and the extension packages *tm* (Feinerer, Hornik, & Meyer, 2008) to extract the relevant data from the corpus. This includes meta data, like date-timestamps, as well

<sup>1</sup> However, it has to be noted that tagging with the GI simply maps text according to categories and does not necessarily identify meaning. Nevertheless, these mappings may provide useful information about the general sentiment reflected by given terms.

<sup>2</sup> Detailed descriptions about SVMs can be found in Huang, Nakamori and Wang (2005); Hastie, Tibshirani and Friedman (2009), Karatzoglou, Meyer and Hornik (2006).

as the actual content of the articles. As the size of this corpus (approximately 16 GB) is beyond what would fit into the main memory (RAM) of a usual desktop system we needed to use the distributed processing environment Hadoop (The Apache Software Foundation, 2010) in order to efficiently access and preprocess the documents, i.e., doing stopword and punctuation removal, as well as to construct the DTM.

### CFNAI

The CFNAI is a weighted average of 85 monthly US economy indicators, like production series, employment, consumption and housing. CFNAI provides a single number as a measure of economic activity in these national series. A zero value for the index indicates that the national economy is expanding at its historical trend rate of growth; negative values indicate below-average growth; and positive values indicate above-average growth (retrieved from <http://www.chicagofed.org/webpages/publications/cfnai/index.cfm>).

### Results

In order to evaluate the appropriateness of the used score methodology, we tested our polarity score on the pre-classified movie review data provided by Pang and Lee (2008). This data set contains 2,000 manually classified movie reviews, 1,000 movies are classified as positive (1), and 1,000 as negative (-1). To test the plausibility of our sentiment score we checked our algorithm on these data. For the positive, we achieved a hit-ratio of 0.67 and for the negative reviews 0.676<sup>3</sup>.

Aggregating the NYT data to monthly sentiment scores as described above enabled us to compare the retrieved scores to any given time series observed in the same time frame. Figure 1 shows the time series of the CFNAI and the estimated (scaled) sentiment scores. In this plot, we observe high agreement between the movements of the two series (see, e.g., the peaks on February 1991).

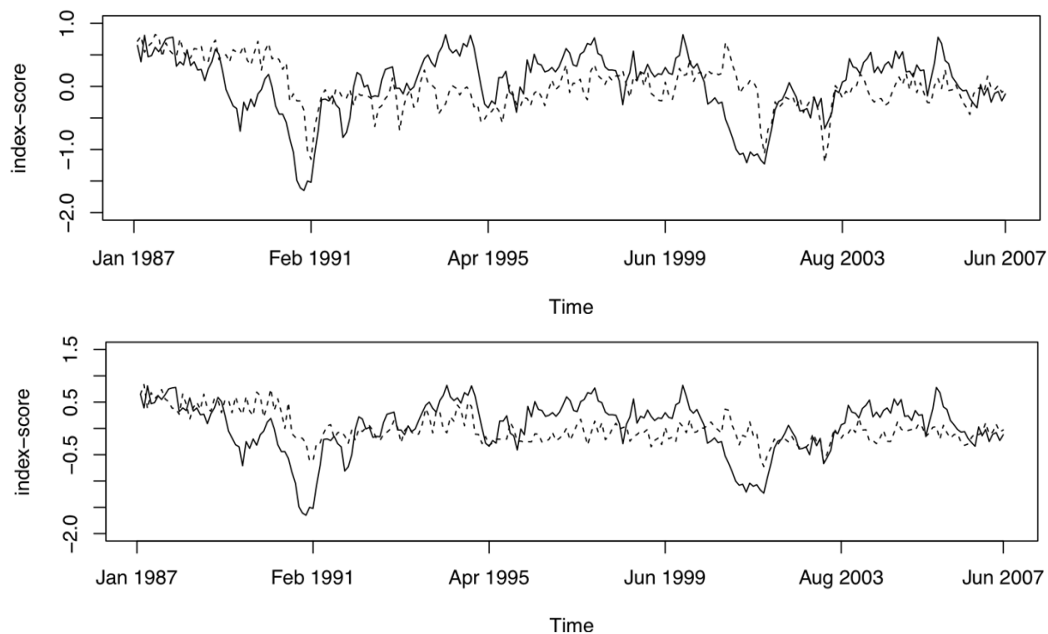


Figure 1. The CFNAI index and the estimated (scaled) sentiments of the NYT data (dotted line, top) and business data (dotted, bottom) from 1987 to 2007.

<sup>3</sup> Misclassification was observed primary on the reviews on action movies, containing words like bad guy, fire, enemy...

By looking at Table 1 showing the forecasting performance of different kernels, we infer that both the linear as well as the polynomial kernel achieve the best hit-ratio.

Table 1

*Forecasting Performance of Different SVM Kernels*

Kernel	NYT	NYT-Business	NYT-diff	NYT- Business-diff
Linear	0.596	0.692	0.6	0.6
Tanhdot	0.377	0.616	0.497	0.497

Figure 2 visualizes the forecasted CFNAI and the original CFNAI data. As a training period we assumed the first 100 months.

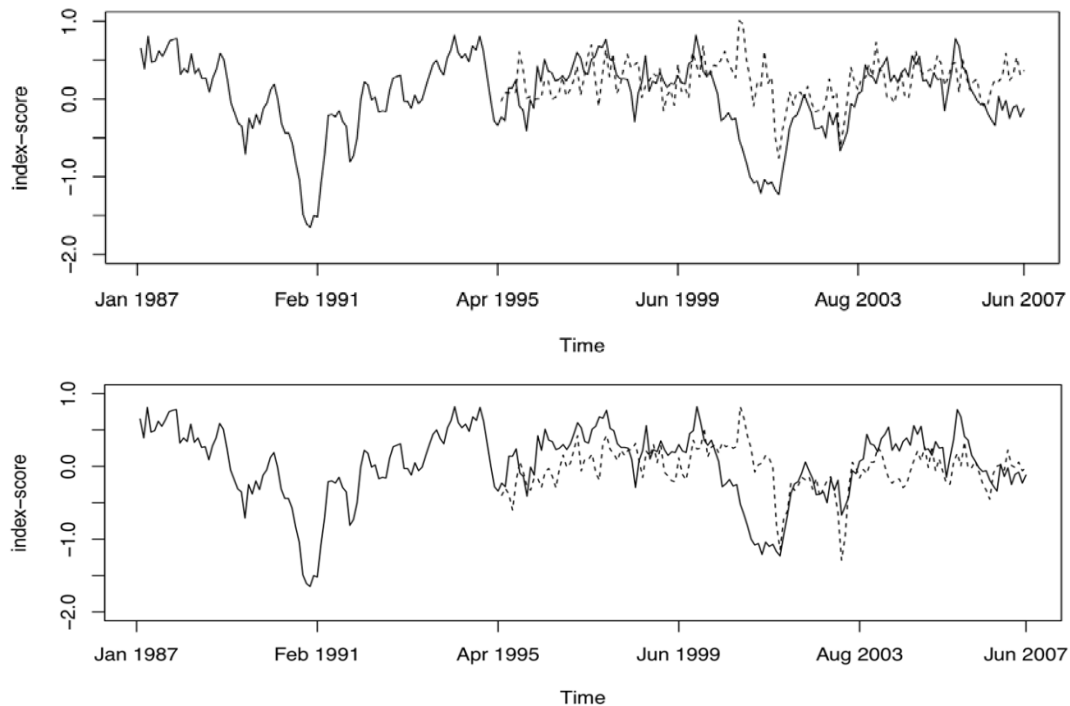


Figure 2. The forecasted CFNAI index (dotted line) and the raw index, using a training period of 100 month and a linear kernel in the SVM framework for the business articles (top) and the whole NYT Data (down).

For all four considered cases (all texts, only business texts, differences all, differences business) the linear kernel (which is also recommended for text mining) achieves the best hit-ratio. Taking the business news articles as a basis for estimating the sentiment score, we even get a hit-ratio near to 70%. Taking a look at Figure 2 nourishes the hypothesis, that our sentiment score seems to work well for calm periods (1995-June 1999), but fundamental mistakes are done during stormy periods, like the early 2000s recession.

### Conclusion and Discussion

In this paper, we studied the use of SVM to forecast directions of the CFNAI based on sentiment scores derived from NYT articles. Several kernels in the SVM model have been employed and the one with the best hit-ratio was chosen to forecast index direction movements. We found that our proposed forecasting approach works well for periods of moderate economic growth. However, further research has to be done in order to

improve the forecasting ability for periods where heavy fluctuations in economic growth are observed.

Sentiment scores were estimated with a naive classifier, which sets the scene for further research, e.g., the improvement of sentiment scores via vector distance classifiers, or Bayesian classifiers (Das & Chen, 2007).

### References

- Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55(2), 107-120.
- Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. *Neural Computing and Application*, 10, 184-192.
- Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375-1388.
- Doms, M., & Morin, N. (2004). Consumer sentiment, the economy, and the news media. FRBSF Working paper 2004-2009. Retrieved from <http://ideas.repec.org/p/fip/fedfap/2004-09.html>
- Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1-54.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed). Springer.
- Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32, 2513-2522.
- Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support vector machines in R. *Journal of Statistical Software*, 15(9), 1-28.
- Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48, 354-368.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Sandhaus, E. (2008). *The New York Times annotated corpus*. Retrieved from <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2008T19>
- Scharl, A., & Weichselbraun, A. (2008). An automated approach to investigating the online media coverage of us presidential elections. *Journal of Information Technology & Politics*, 5, 121-132.
- The Apache Software Foundation. (2010). *Hadoop*. Retrieved from <http://hadoop.apache.org/core/>