“Money Can’t Buy Love?” Creating a Historical Sentiment Index for the Berlin Stock Exchange, 1872–1930

Borst, Janos

janos.borst@uni-leipzig.de
Computational Humanities Group, Leipzig University, Germany

Wehrheim, Lino

lino.wehrheim@ur.de
Economic and Social History, Regensburg University, Germany

Burghardt, Manuel

burghardt@informatik.uni-leipzig.de
Computational Humanities Group, Leipzig University, Germany

Motivation

Financial economists and psychologists agree that collective sentiment plays a crucial role in financial markets (Akerlof and Shiller 2009, Tuckett 2011). In this paper, we present our workflow for creating a daily aspect-based index that captures the sentiment at the Berlin Stock Exchange from 1872 to 1930, a time when Berlin was the key financial market in Germany. This index is based on market reports published every trading day in the Berliner Börsen-Zeitung, which give a verbal description of the sentiment among market participants. Due to daily publication and the long observation period, our corpus consists of about 18,000 market reports. We apply a combination of expert annotation and machine learning. With this newly-created data, we will be able to gain a better understanding of the historical German stock market and, more generally, the nature of financial sentiment itself. How has sentiment developed over time and how is it related to historical events? Did sentiment influence prices and/or trading volume, or vice versa? To answer such questions, we need data on financial sentiment which, to our best knowledge, is not available for the aforementioned case yet. From a DH perspective, this paper covers two novel aspects: First, we focus on both historical and highly domain-specific language, a dual challenge that thus far has rarely been addressed. As there are many similar historical sources, such as reports by chambers of commerce, our solutions will be helpful for future research. Second, we address a characteristic but neglected feature of financial texts that might be relevant also in a broader sentiment analysis context. Particularly, we focus on aspect- and entity-based sentiment analysis.

Method and Corpus

There is an abundance of financial texts, such as newspapers, financial reports, or tweets. Accordingly, many scholars have applied some form of natural language processing to quantify sentiment. Beginning with the work by Tetlock (2007), scholars first relied on dictionaries to perform sentiment analysis (e.g., García 2013, Hanna et al. 2020). As machine learning approaches display better performance (Mishev et al. 2020, van Atteveldt 2021), scholars have started to use more advanced methods, especially language models such as BERT and FinBERT (Araci 2019, Malo et al. 2014). One reason is that financial texts pose several distinct problems for sentiment analysis (Njølstad et al. 2014, Loughran and McDonald 2011, Xing et al. 2020). One of them regards the role of entities which can complicate sentiment analysis considerably (Sinha et al. 2022). In our case, a major challenge is that texts include statements referring to entities at three different levels: individual assets (category 1), industries or sectors (category 2), and the overall stock market (category 3). Typically, sentences contain multiple entities from different levels, often with different sentiment valuations. As we are primarily interested in overall market sentiment, we must account for these differences to avoid misleading sentiment scores.

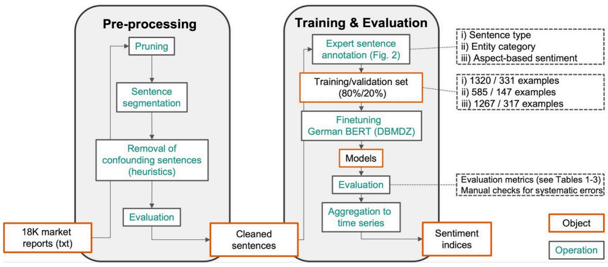

Figure 1: Workflow to create aspect-based sentiment indices.

Our corpus and workflow are outlined in Figure 1, several aspects of which shall be mentioned briefly. The removal of confounding sentences includes those containing numerical expressions stating pure information and neutral sentences. The first type of sentence was removed using heuristics, the latter a classification step (see below). We performed this pre-processing because Wilson et al. 2009 have shown that removing neutral sentences can increase classification performance. The training set was created by a financial historian, making use of the standardized financial language and primary sources on historical stock markets (e.g., Krupke 1904, Kautsch 1912). In a first run, the expert annotated 1,500 random sentences in a three-step procedure (see Figure 2):

i) Sentence type: does the sentence contain a sentiment-related statement? (If not, the sentence was discarded from further analysis.)

ii) Entity category: which categories do the entities mentioned in the sentence belong to?

iii) Aspect-based sentiment: which sentiment is connected to a given entity?

In a second run, the expert annotated another 750 sentences to address systematic errors. We included entity-level classification during this stage because we lack information on named entities due to various issues (spelling, abbreviations, OCR errors, etc.). Also, NER often has trouble recognizing industry-type entities even for modern financial texts (Sinha et al. 2022).

Sentence type and entity level classification were tackled employing the standard procedure of finetuning a language model using a classification head (Devlin et al. 2018). Our base model was a German BERT trained by the MDZ. We train for 15 epochs and choose the best model based on the macro averaged F1-score. For macro average the Precision/Recall/F1 values are calculated per class and then averaged, weighing every class equally. We chose this as a target metric to avoid a bias toward any specific class overall.

Aspect-based sentiment was tackled using a sentence-pair classification similar to NLI tasks (Sun et al. 2019): A sentence (premise) was combined with a formulation of the aspect related sentiment (hypothesis) and classified if the sentences match. We use the hypothesis formulation: "Die Stimmung bezüglich [aspect] ist [label]". After some experiments, the best model was achieved by finetuning using Compacter Adapter (Mahabadi et al. 2021), increasing macro F1 significantly over standard finetuning. The resulting performance of these models is reported in the next section.

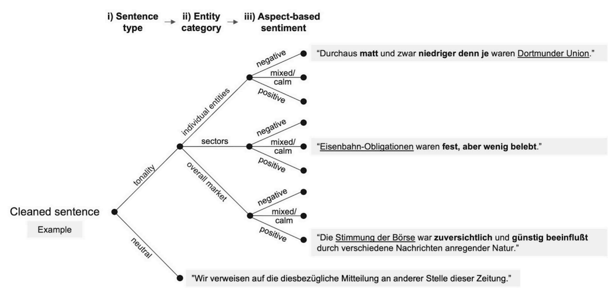

Figure 2: Sentence annotation process.

Evaluation and Preliminary Results

Tables 1-3 show the respective evaluation of each machine learning step. The sentence-type classification shows high performance, especially in recall of sentences containing tonality (96,5%), which is vital for the following steps. The aspect-based sentiment classification shows an F1 macro average of 85.9%. While there is still room for improvement, we argue that since there is no significant difference among classes, the systematic error will not change the overall trend of the chronological analysis.

Table 1: Evaluation metrics of sentence type classification.

	Neutral	Tonality	Micro	Macro
Precision	94,3%	92,7%	93,4%	93,5%
Recall	88,5%	96,5%	93,4%	92,5%
F1-Score	91,3%	94.6%	93,4%	93,0%
Examples	120	211	331	331

Table 2: Evaluation metrics of category detection.

	Individual Entities	Sectors	Overall Market	Micro	Macro
Precision	92,5%	76,2%	77,1%	82,5%	81,9%
Recall	89,9%	84,2%	90,2%	88,0%	88,1%
F1	91,1%	80,0%	83,1%	85,2%	84,8%
Examples	69	57	41	167	167

Table 3:>Evaluation metrics of aspect-based sentiment classification.

	Negative	Mixed / Calm	Positive	Micro	Macro
Precision	81.6%	90.5%	85.7%	86.1%	85.9%
Recall	88.6%	85.7%	83.5%	86.1%	85.9%
F1-Score	84.9%	88.0%	84.6%	86.1%	85.9%
Examples	105	133	79	317	317

Figure 3 shows aggregated monthly sentiment scores for every category. To identify trends, we applied twelve-month moving averages, which reveals clear periods of positive/negative sentiment. Some of them appear straightforward, such as the decline of sentiment after the stock market crash in 1873, others are more surprising. For example, there seems to be no trace of the financial turmoil that took place around 1900.

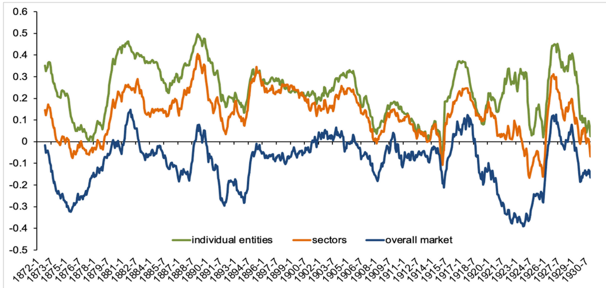

Figure 3: Monthly aspect-based sentiment scores, 1872–1930, rolling 12 month averages.

Outlook

In this paper, we describe the process of annotating a corpus of historic German stock market reports with aspect-level sentiment. We focused on the methodological part, making our procedure as transparent as possible, enabling readers to comprehend the data if they intend to use it, once it is published.

Obviously, our results demand further in-depth analysis, both technical and historiographical. A more comprehensive historiographical interpretation of the results is part of plannedfuture work.A natural starting point will be to determine the correlation of the three indices with macro-historical data, such as stock market indices or GDP. Also, the training data could be extended to include concrete entities, which would allow us to identify individual companies. If we were able to provide company-level sentiment data, this could be combined with share price data that is created by other financial historians, opening the research on financial sentiment also to a business history perspective.

Bibliography

Akerlof, George A., Robert J. Shiller. Animal spirits. Princeton, 2009.

Araci, Dogu. „FinBERT: Financial Sentiment Analysis with Pre-trained Language Models", 2019.

Hanna, Alan J., John D. Turner, Clive B. Walker. „News Media and Investor Sentiment during Bull and Bear Markets". The European Journal of Finance 26, 14 (2020): 1377–95.

García, Diego. „Sentiment during Recessions". The Journal of Finance 68, 3 (2013): 1267–1300.

Kautsch, Jacob. Handbuch des Bank- und Börsenwesens für Kaufleute, Industrielle, Kapitalisten, Bankiers und Bankbeamte. Mit besonderer Berücksichtigung Deutscher, Österreichischer und Schweizerischer Verhältnisse und den in Deutschland, Österreich und der Schweiz Geltenden Bank- und Börsengesetzen. Berlin, 1912.

Krupke, Franz. Krupkes Konversations-Lexikon der Börse und des Handels und praktischer Führer für Kapitalisten. Berlin, 1904.

Liebl, Bernhard Manuel Burghardt. "From Historical Newspapers to Machine-Readable Data: The Origami OCR Pipeline." Proceedings of the 1st Workshop on Computational Humanities Research (CHR) (2020).

Loughran, Tim, Bill McDonald. „When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". The Journal of Finance 66, 1 (2011): 35–65.

Malo, Pekka, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, Pyry Takala. „Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts". Journal of the Association for Information Science and Technology 65, 4 (2014): 782–96.

Mishev, Kostadin, Ana Gjorgjevikj, Irena Vodenska, Lubomir T. Chitkushev, Dimitar Trajanov. „Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers". IEEE Access 8 (2020): 131662–82.

Njølstad, Pål Christian S., Lars S. Høysæter, Wei Wei, Jon Atle Gulla. „Evaluating Feature Sets and Classifiers for Sentiment Analysis of Financial News". In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 71–78. Warsaw, Poland: IEEE, 2014.

Karimi Mahabadi, Rabeeh, Henderson, James, & Ruder, Sebastian (2021). Compacter: Efficient low-rank hypercomplex adapter layers. Advances in Neural Information Processing Systems, 34, 1022-1035.

Sinha, Ankur, Satishwar Kedas, Rishu Kumar, Pekka Malo. „SEntFiN 1.0: Entity-Aware Sentiment Analysis for Financial News". Journal of the Association for Information Science and Technology 73, 9 (2022): 1314–35.

Sun, Chi, Luyao Huang, and Xipeng Qiu. "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence". Proceedings of NAACL-HLT, 2019, 380–385.

Tetlock, Paul C. „Giving Content to Investor Sentiment: The Role of Media in the Stock Market". The Journal of Finance 62, 3 (2007): 1139–68.

Tuckett, David. Minding the markets: an emotional finance view of financial instability. Basingstoke, 2011.

van Atteveldt, Wouter, Mariken A. C. G. van der Velden, Mark Boukes. „The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms". Communication Methods and Measures 15, 2 (2021): 121–40.

Xing, Franz Z., Lorenzo Malandri, Yue Zhang, Erik Cambria. „Financial Sentiment Analysis: An Investigation into Common Mistakes and Silver Bullets". Proceedings of the 28th International Conference on Computational Linguistics, 2020, 978–87.

Wilson, Theresa, Janyce Wiebe, Paul Hoffmann. „Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis". Computational Linguistics 35, 3 (2009): 399–433.