Design and Evaluation of *SentiEcon*: a fine-grained Economic/Financial Sentiment Lexicon from a Corpus of Business News

Antonio Moreno-Ortiz, Javier Fernández-Cruz, Chantal Pérez-Hernández

Universidad de Málaga (Spain)

Pontificia Universidad Católica del Ecuador Sede Esmeraldas (Ecuador)

{amo,fernandezcruz,mph}@uma.es

Abstract

In this paper we present, describe, and evaluate *SentiEcon*, a large, comprehensive, domain-specific computational lexicon designed for sentiment analysis applications, for which we compiled our own corpus of online business news. *SentiEcon* was created as a plug-in lexicon for the sentiment analysis tool Lingmotif, and thus it follows its data structure requirements and presupposes the availability of a general-language core sentiment lexicon that covers non-specific sentiment-carrying terms and phrases. It contains 6,470 entries, both single and multi-word expressions, each with tags denoting their semantic orientation and intensity. We evaluate *SentiEcon's* performance by comparing results in a sentence classification task using exclusively sentiment words as features. This sentence dataset was extracted from business news texts, and included certain key words known to recurrently convey strong semantic orientation, such as "debt", "inflation" or "markets". The results show that performance is significantly improved when adding *SentiEcon* to the general-language sentiment lexicon.

Keywords: sentiment analysis, opinion mining, financial texts, lexicons, language resources

1. Introduction

The analysis of public opinion on globally relevant entities has consistently been a matter of great import, especially since the advent of market research, starting in the second half of the 20th century. As the number and variety of information resources has grown exponentially due to the consolidation of the Web 2.0 in the last decade, we face the challenge of transforming the richness of this vast amount of information into computationally tractable data. From an economic point of view, there has always been a demand for information on opinions related to financial institutions such as central banks, or the appraisal of the economic state-of-affairs, such as the *Monthly Michigan Index of Consumer Sentiment* in the USA. Generally funded by either public or private institutions, these studies were carried out employing traditional surveying techniques, with data analysis methods that are not always transparent and, in many cases, without the possibility to access the original data, which would allow replication of results or find alternative interpretations.

A newer trend, however, has been used in recent years which employs data and text mining techniques. Sentiment Analysis can no doubt be seen as an alternative to traditional surveying methods, and has been considered as a more dynamic and robust solution to the analysis of online text, such as news, user-generated content (UGC) or financial reports (Batra and Rao, 2010; Yu et al., 2013).

Despite the formidable body of research generated on the topic of sentiment analysis (see Liu (2015) for a concise state-of-the-art), SA strategies are still far from perfect as systems fall short of accounting for the vast complexity of human speech. Traditionally, approaches to sentiment analysis employ machine learning (ML) techniques, sentiment lexicons or, more commonly, a combination of both. The success of machine learning techniques draws on the well-known fact that the expression of sentiment depends is heavily dependent on domain

specificity, that is, words and phrases tend to have the same sentiment within the same genre, but may vary across domains (Aue and Gamon, 2005; Ding and Liu, 2007; Pang and Lee, 2008; Choi et al., 2009). ML algorithms excel at capturing systematicity when trained on a large corpus of similar documents, where the same lexical items are used repeatedly to express the same sentiment. Conversely, lexicon-based approaches rely on the existence of sentiment lexica. Such resources can be automatically created (Lu et al., 2011; Tai and Kao, 2013) or crafted manually. As Perrie et al. (2013) discuss, the advantage of the former is a faster creation process, whereas the latter tend to be more reliable. Either way, the availability of high quality sentiment lexical resources is of utmost importance for high performance.

In this paper we describe the semiautomatic process of building a sentiment lexicon in the financial/economic domain, which assumes the existence of a general-language sentiment lexicon. Thus, we exploit the concept of a "plug-in" lexicon, compiled to be used in combination with general domain sentiment lexicons. To create this resource, we developed a corpus-driven method in order to extract candidate terms from a corpus of business news designed ad-hoc. These candidates were then matched against our existing general-language polarity lexicon, *Lingmotif-lex* (Moreno-Ortiz and Perez-Hernandez, 2018), to compile the final product. Finally, we evaluate its performance by running a sentiment classification task using several general-purpose sentiment dictionaries versus using them together with *SentiEcon*.

2. The Economic-financial Domain in Sentiment Analysis

The language of economics and finance is highly specialized, as has been recognized repeatedly by researchers in the field (Sinha et al., 2019; Loughran and McDonald, 2011; Loughran and McDonald, 2014;

Loughran and McDonald, 2016; Krishnamoorthy, 2018). The notion of domain dependence is particularly relevant in SA (Costa et al., 2016), as it impacts every level of analysis and often requires expert judgment to determine the polarity of a given text. At the lexical level, in particular, the polarity of one word may vary across domains. For example, Loughran and McDonald (2011) observed that in the economic-financial domain, the word "liability" is considered neutral, whereas it usually conveys a negative polarity in general language. This makes it difficult to reuse SA lexicons across domains and accentuates the necessity to develop domain-specific resources.

On the other hand, opinions recurrently shake market dynamics: good news may lift markets and bad news may sink them. Behavioral economics, e.g., Thaler (2015), considers that emotions trespass the boundaries of rationality and govern, together with other factors, the decision-making of investors. Even before the advent of automatic sentiment analysis techniques, applying textual analysis to financial markets was a well-established practice. For instance, Niederhoffer's pioneering research (Niederhoffer, 1971) proposed the development of computational text analysis techniques to obtain greater objectivity in the results. He analyzed manually the semantics of two decades of headlines in the business section of *The New York Times*. Similarly, Klein and Prestbo (1974) supported the intrinsic connection between markets and news after observing how bad news influenced stocks. It is therefore not surprising that sentiment analysis, as understood today, was regarded, since its advent, as a powerful decision-making toolkit for analyzing market behaviour (Shiller, 2000).

Financial sentiment analysis emerged in the mid-2000s and focused mainly on the prediction of stock market movements employing external information such as press headlines, e.g., Meyer et al. (2017), corporate reports, e.g., Hajek et al. (2014), social media (mainly *Twitter* and *Stocktwits*), e.g., Gaillat et al. (2018a), or expert sentiment towards different firms and their movements (O'Hare et al., 2009; Bar-Haim et al., 2011). This sentiment-finance correlation has been consistent through the ages; Garcia (2013) observed the effect of sentiment on asset prices during a century-long period (1905 to 2005) from two daily financial columns from the New York Times. His lexicon-based SA analysis evidenced that economic recessions correspond with times of heightened expression of sentiment in the news. In a very recent study, Shapiro and Wilson (2019) analyzed transcripts of meetings of the Federal Open Market Committee in order to estimate the institution's objective function; they concluded that text-based sentiment analysis of news returns similar results to survey-based consumer sentiment measures.

Within Sentiment Analysis, as understood today, it is common to distinguish corpus-based approaches from lexicon-based approaches. Although a combination of both methods can be found in the literature (Riloff et al., 2006), lexicon-based approaches are usually preferred for sentence-level classification (Andreevskaia and Bergler, 2007), whereas corpus-based, statistical approaches are preferred for document-level classification. Lexicon-based approaches to sentiment classification rely mostly on a dictionary that contains sentiment words with their associated polarity. But sentiment dictionaries have also been shown to improve performance in corpus-based approaches, and many researchers will use them as available or expand them to adapt them to their data sets. Although the application of general-domain lexicons tend to misclassify financial texts with a negative impact on performance, some influential financial research projects have used this type of lexicons, especially the Harvard General Inquirer (Stone and Hunt, 1963). Tetlock (2007) used a sentiment dictionary to observe the influence of negative words in financial media in order to predict the earnings of companies and stock returns.

The application of general-domain lexicons, such as the HGI, to the financial domain has been widely criticized. In this particular case, Loughran and McDonald (2011) point out that as much as as 73.8 percent of HGI's negative words does not have a negative sense in financial documents. They went on to re-examine this lexicon and produced a new one from documents published by the U.S. Securities and Exchange Commission. Entries in their dictionary may belong to one of the following categories: *positive, negative, litigious, uncertain, strong modal,* and *weak modal*. Our approach is similar, although it is based on the idea of a *plug-in* domain-specific lexicon, where words that show a different polarity in the general sentiment dictionary are given the correct polarity in the plug-in dictionary, and words and phrases that do no exist are included.

Other similar resources include Malo et al. (2014), who trained classifiers to conduct sentence-level SA of financial news. Their Financial Phrase Bank provides a set of 5,000 sentences, manually annotated by 16 subject experts. It features specific domain concepts, an extensive list of verbs and expressions that help to detect the direction of financial indicators and a classification of the potentiality of some directionality-dependent terms. The resulting compilation is sensitive to the directionality of entities through the parsing of sentences. This resource was updated by Sinha et al. (2019), who also released an entity-annotated news dataset containing over 12,000 headlines and their related financial sentiment.

Financial microblogs are another relevant field of research. Oliveira et al. (2016) produced a stock market sentiment lexicon, which includes 20,551 items extracted automatically from microblogs (StockTwits and Twitter).

In other languages, several resources have also been described. Van de Kauter et al. (2015) adapted a general-domain lexicon in Dutch language featuring 3,187 items. Mao et al. (2011; Mao et al. (2014) built a financial lexicon in Chinese from a large news corpus trained on stock returns. Gaillat et al. (2018b) compiled the multilingual SSIX Corpora that classified 2,886 microblogging messages and its different entities in Spanish, German and English.

3. Corpus compilation and lexicon extraction

As mentioned in section 1, *SentiEcon* is designed to be used in combination with a general domain lexicon. Our plug-in approach (see Moreno-Ortiz and Fernández-Cruz (2015)

for more details) aims to compile only those particular entries for which their domain-specific polarity differs or is not recorded in the general domain lexicon. It is designed, thus, to override the general polarity of lexical items with the one assigned in the plug-in lexicon, when they are used in the economic-financial domain and to add those sentiment words and phrases specific of the domain.

3.1. Data Sources

To extract term candidates, terminology extraction software requires a domain-specific corpus as the main source to extract specialized knowledge and a general-domain data collection to be used as a reference corpus. For our purposes, the *Esmeraldas Great Recession News Corpus* (GRNC) served as the main data source. This corpus consists of 41,000 news articles from the Business section of two major English-language daily newspapers: *The Guardian* and *The New York Times*. In total it contains 24 million words. It covers most releases published between January 2007 and December 2015. The general language reference corpus used was the English Web 2015 (EnTenTen15) (Jakubíček et al., 2013), of approximately 15.7 billion words.

3.2. Lexicon Creation Process

The compilation of *SentiEcon* followed a 3-step process: (1) extracting candidate terms automatically, (2) matching candidates against our general-language polarity database, and (3) refining results through corpus work and manual assigning sentiment scores by domain experts.

We used The Sketch Engine's KeyWords utility (Kilgarriff et al., 2014) to extract a set of some 30,000 single-word and multi-word term candidates. All term candidates were matched against our general-language polarity database (*Lingmotif-lex*). At this point it is worth remarking that our intention is not to extract a complete set of financial-economic terms, but rather to obtain a list of words and phrases that convey positivity or negativity when their orientation diverges from the sentiment they carry in other domains or in the general language or when the semantic orientation is not present in it. For example, terms like "sale" or "bank" are irrelevant here, as they are neutral in both the specific and the general domain. However, some single word terms such as "asset" or "fix" are included in the general lexicon as positive, but need to be recategorised in the plug-in lexicon as neutral, as their polarity differs: "The currency fixes are based on actual trades" vs. "There is no easy fix: removing the wires can be dangerous". In addition, all non-terms, neutral terms and terms that did not convey different semantic orientation in the financial domain were discarded. Once a complete list of candidates was compiled, we refined these results by matching candidates against the corpus and assigned sentiment scores with the support of a domain expert. The process was carried out as follows:

1. Initial speculative term selection and sentiment score assignment for clearly identified polarity words.

2. Checking intuitive assignment against corpus data by searching for alternative uses and local grammar patterns.

3. Enhancement of term set by adding lexical entries manually, in case, for instance, of obvious misses identified in the previous step. For example, a non-apparent term candidate found in the list was "haircut", which was rather frequent, especially related to the European debt crisis: "The Troika want haircuts of over 50% for wealthy savers (100k+ euro) at Bank of Cyprus."

4. Assignment of final sentiment and intensity scores according to the lexicon's format, which is compiled in a tab-separated UTF-8 file (e.g., money_<launder> VB neg 3).

3.3. Description of *SentiEcon*

Each lexicon entry consists of four data fields, as described in Table 1. A more extensive description of the lexicon's format can be found in Moreno-Ortiz and Perez-Hernandez (2018).

Data field	Example/List
Word form	launder, haircut, european_central_bank
PoS	[ALL, NN, JJ, VB, RB, UH, IN]
Polarity	[POS, NEG, NEU]
Intensity	[0, 1, 2, 3]

Table 1: *SentiEcon*'s data fields

SentiEcon includes both single words and multiword expressions, as they convey an important part of the semantic orientation of a text. Certain financial terms that reflect the results of a company's activity are technically considered neutral when isolated (e.g., "unemployment"), but undoubtedly convey a strong semantic orientation, which is revealed when combined with directional lexical elements. These elements are marked as neutral with intensity in the lexicon, and their semantic orientation when combined with directionality elements is expressed as multiword templates.

Such cases are very common in the economic and financial domain, where variance figures are more important than absolute values. These economic indicators are characterized by not having inherent sentiment but will be positive or negative as they increase or decrease. The polarity of these terms will depend on the expected direction of events (e.g., the results are positive when they are expected to increase, but neutral or negative when they decline). Directionality elements, mainly verbs or nouns (e.g., "rise", "rocket", "fall", "collapse") mark both the direction and intensity of their respective performance indicators. For example, "debt is rising" conveys a negative polarity. This is easily implemented in our lexicon through a set of nouns and verbs (e.g. "drop", "dip", "double", "rise") that are redundantly combined with different positive or negative indicators, within a span of three words. Table 2 shows several examples.

The final version of our economic/financial sentiment plug-in lexicon contains 6,843 entries, distributed as shown in Table 3. As is obvious from the single-word vs. multiword

MWE pattern	Example
<debt> all neu 2	"Holders of senior debt would be repaid in full."
<debt>_3_<rise> nn neg 2	"Carinthia's total debt will rise to euro 3.7 billion in 2014 (...)"
<debt>_3_<fall> nn pos 2	"Household debt has fallen from its peak."
unemployment_3_<dip> vb pos 2	"But the unemployment rate dipped to 5.1% and hourly wages were steady (...)"
unemployment_3_<grow> vb neg 2	"(...) with 24% unemployment also set to grow."

Table 2: Examples of MWE patterns with directionality elements

proportions, the latter play a crucial role in determining the sentiment conveyed in the texts in this domain, which means that lexicon-based approaches to sentiment classification should take into consideration n-grams in order to be successful.

Polarity	Words	MWE	Total
POS	343	1022	1365
NEG	736	1708	2444
NEU	309	2352	2661
Total	**1388**	**5082**	**6470**

Table 3: Count and distribution of entries in *SentiEcon*

4. Evaluation Methods

In order to evaluate the performance of *SentiEcon*, we developed a manually annotated gold standard dataset consisting of 1,000 sentences. We decided to use sentences rather than paragraphs or full documents because it has been shown in the literature that document-level SA in the financial domain does not generally account for the relevance of text segments, as individual sentences in financial news typically focus on different aspects which may express different sentiments. For example, after analyzing 1,000 random sentences from financial announcements, Lutz et al. (2019) concluded that an accurate classification of sentences would perform more fine-grained explanatory analyses on financial texts and also improve pre-existent prediction systems. Sentence-level classification was also used in other relevant works, such as Malo et al. (2014) and Sinha et al. (2019). The length of the news articles was also a deciding factor, since such texts are rather long articles where different concepts, events and entities are described and contrasted, which introduces a number of extraneous variables and unnecessarily complicates things for our purposes.

Next, the 1,000 sentences sample was extracted randomly from our corpus. Then two domain experts annotated the dataset by classifying each sentence as belonging to one of three categories: POSITIVE, NEGATIVE, and NONE. They were instructed to take into account only the information available in the sentences and to annotate sentences. Annotation was carried out independently and then they were asked to reach a consensus in differing cases. Similarly to Malo et al. (2014), our annotators were instructed to consider the following main guidelines while annotating the phrases:

- There are no fixed rules about how particular words should be annotated.

- Avoid bias based on prior knowledge about the company or institution. Thus, each sample sentence should be annotated by using the information that is explicitly available.

- Be as consistent as possible with respect to your own annotations.

This dataset was split into training and test sets in 80/20 percentages. The only features used to train the classifiers were the number of sentiment items (positive and negative) found in the sentences, the hypothesis being that the addition of a high-quality domain-specific lexicon such as *SentiEcon* should have a strong impact on the performance of the classifiers. Even though some of the lexicons include neutral lexical items, we decided not use this as a training feature for any. We carried out 24 experiments in total, using three different lexicons with and without the addition of our plug-in lexicon and, in order to take into account the impact of various classification algorithms, we trained and tested four different ML classifiers: logistic regression, Naive Bayes, Random Forest, and SVM. We tested the performance of *SentiEcon* with the following sentiment lexicons:

- The General Inquirer (Stone and Hunt, 1963). This is one of the oldest sentiment lexicons publicly available. It is based on work in cognitive psychology and content analysis. This resource offers syntactic, semantic, and pragmatic information to part-of-speech tagged words, with 1915 positive and 2291 negative words. Lexical items for "yes" and "no" (in the sense of refusal) are grouped in separate categories and further semantic dimensions, such as strength or active/passive orientation, are also included[1].

- MPQA (Multi-Perspective Question Answering) Subjectivity Lexicon (Wilson et al., 2005). This resource contains 2,718 positive and 4,912 negative words drawn from a combination of sources, including the General Inquirer lists, the output of the system created by Hatzivassiloglou and McKeown (1997) and a bootstrapped list of subjective clues (Riloff and Wiebe, 2003), hand-labeled for sentiment. The lexicon also includes labels for reliability (strongly subjective or weakly subjective) and four polarity tags: positive, negative, both and neutral. The majority of words are marked as having either positive (33.1 percent) or negative (59.7 percent) polarity, whereas only a small number of clues (0.3 percent) are marked as having both positive and negative, and

---

[1] Available at http://www.wjh.harvard.edu/~inquirer/.

6.9 percent of the clues in the lexicon are marked as neutral[2].

- Lingmotif-lex (Moreno-Ortiz and Perez-Hernandez, 2018) A manually-curated, wide-coverage, domain-neutral lexicon for sentiment analysis in English, developed for the sentiment analysis tool Lingmotif (Moreno-Ortiz, 2017a; Moreno-Ortiz, 2017b). It contains over 28,000 single-word forms and over 38,000 multi-word expressions.

All lexicons were adapted to the minimum shared features, but their performance was ultimately determined by the coverage and quality of their contents, as well as their relevance to the financial domain. In the following section we describe and discuss the results we obtained.

5.   Results

Table 4 summarizes our results in terms of accuracy. For every lexicon we have used the best results from the chosen classifiers, which is mentioned in each case (LR: Logistic Regression, NB: Naive Bayes, RF: Random Forest, SVM: Support Vector Machines).

Lexicon	Accuracy (No plug-in)	Accuracy (SentiEcon)	Diff.
LM-Lex	0.710 (SVM)	0.787 (RF)	**0.077**
MPQA	0.594 (SVM)	0.682 (RF)	**0.088**
HGI	0.538 (LR)	0.612 (SVM)	**0.074**

Table 4: Summary of results (Accuracy)

Improvements are significant for every lexicon in terms of accuracy, across all categories. Accuracy is thus improved by nearly 8 percent on average by employing *SentiEcon*. Next we offer the performance results thrown by the different combinations of classifiers and lexicons in terms of precision and recall for each of the classes in the dataset (POS, NEG and NONE). Tables 5, 6, and 7 show the results of the experiments performed with each of the lexicons without and with the *SentiEcon* plug-in lexicon. In general, performance figures show that the plug-in improves precision and recall for all categories (positive, negative and none), although some differences in degree are apparent. The precision, recall, and F1 figures in these tables are given using a "no plugin/with plugin" notation. the "F1 Diff." column summarizes the difference in performance for each of the classes. The "Avg" row displays the macro-average values for each metric.

Class	Precision	Recall	F1	F1 Diff.
POS	0.69/0.68	0.70/0.81	0.69/0.74	0.04
NEG	0.76/0.80	0.81/0.87	0.78/0.83	0.05
NONE	0.63/0.98	0.53/0.63	0.58/0.77	0.19
Avg	**0.69/0.82**	**0.68/0.77**	**0.69/0.79**	**0.11**

Table 5: Evaluation results for Lingmotif-lex.

These data suggest that the negative polarity is more easily identified, followed by the positive class. The NONE class,

Class	Precision	Recall	F1	F1 Diff.
POS	0.66/0.69	0.56/0.71	0.61/0.70	0.09
NEG	0.61/0.67	0.73/0.81	0.66/0.73	0.07
NONE	0.48/0.70	0.41/0.45	0.44/0.55	0.11
Avg	**0.58/0.69**	**0.57/0.66**	**0.57/0.66**	**0.09**

Table 6: Evaluation results for MPQA.

Class	Precision	Recall	F1	F1 Diff.
POS	0.46/0.52	0.63/0.81	0.53/0.63	0.10
NEG	0.61/0.70	0.76/0.69	0.68/0.69	0.02
NONE	0.00/0.79	0.00/0.21	0.00/0.33	0.33
Avg	**0.36/0.67**	**0.46/0.57**	**0.40/0.62**	**0.22**

Table 7: Evaluation results for General Inquirer.

which refers to sentences where no polarity is present, is, by far, the most difficult to classify correctly. This is because the presence of polarity lexical items does not always determine the polarity of the whole sentence, of course. The addition of the plug-in lexicon, where specialized lexical entries are able to "neutralize" a significant number of otherwise polarized words and phrases proves to be essential for the correct classification of these "none" cases.

The data also show that scores are higher for recall of negative items in the MPQA and General Inquirer and, consequently, the addition of the plug-in lexicon greatly improves their metrics, although not as significantly as in the case of positive items.

Special mention deserves the fact that neutral ("none") sentences are misclassified by the General Inquirer lexicon on its own, which is why there is such a big difference when the plug-in lexicon is added. Of course it also improves results in the other categories. Examples of misclassified NONE sentences are the following:

- UBS said it would receive a direct injection of government money worth some $5.3 billion, while Credit Suisse said it had raised $8.8 billion from "a small group of major global investors" including the Qatari authorities.

- The increase is higher than Next predicted only two months ago when it said prices would rise between 5% and 8%.

- That is worth about $363.63 a share, or 16 percent higher than Allergan's closing price on Friday.

- But the movable feast comes at a cost: pollution – especially carbon dioxide, the main global warming gas – from transporting the food.

- Johnson had also met AstraZeneca and GSK, whose headquarters are in Britain, while O'Brien had been to the Paris headquarters of the French firm, Sanofi-Aventis.

In contrast to Lingmotif-lex and MPQA's accuracy in the detection of neutral sentences, HGI's poor performance seems to be motivated by much lower polarity prediction figures in the lexicon-based classification. For instance,

---

[2]Available at http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/.

Lingmotif-lex detected 133 neutral sentences accurately and failed to do so 112 times. By contrast, HGI only succeeded detecting polarity 44 times, while failing to do so 201 times. Table 8 summarizes results for the category NONE in terms of identified positive and negative items. Again, domain specificity in the sentiment lexicon appears to be a key determining factor to assign the correct polarity, as some neutral items (e.g., "authorities" or "feast") have been tagged as positive while they are neutral in the target domain.

Lexicon	P. items=0	P. items>0	Ratio
Lingmotif-lex	133	112	1.19
MPQA	112	133	0.84
General Inquirer	44	201	0,22

Table 8: Identification of polarity items in the NONE class.

6. Conclusions

In this paper we have presented *SentiEcon*, a wide coverage, domain-specific lexicon for the analysis of economic and financial texts in English. We have described our approach to the compilation and labeling process of this resource according to the requirements of the Lingmotif SA system, although the resource can be used in combination with any general language sentiment dictionary. Three main features characterize *SentiEcon*: its careful manual curation, the strong emphasis placed on multi-word expressions, and a simple implementation of a directionality system for economic indicators that accounts for sentiment assignation by context. Even if the system requires labour-intensive terminographical sessions with heavy reliance on domain experts, results already offer enough evidence as to the high quality and flexibility of our resource. Our evaluation results evidence that, in combination with a high-quality general-language sentiment lexicon, such as Lingmotif-lex, *SentiEcon* achieves very high performance levels that may encourage us to carry out further performance tests, different applications and the compilation of other domain-specific lexicons. Future research paths will also take into account the influence of specialized languages variation, as lexicons must adapt to different levels of expertise along with the observation of cross-domain interference.

7. Acknowledgements

8. Bibliographical References

Andreevskaia, A. and Bergler, S. (2007). CLaC and CLaC-NB: Knowledge-based and Corpus-based Approaches to Sentiment Tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 117–120, Stroudsburg, PA, USA. Association for Computational Linguistics.

Aue, A. and Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. In *Proceedings of recent advances in natural language processing (RANLP)*, volume 1, pages 2–1. Citeseer.

Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., and Goldstein, G. (2011). Identifying and Following Expert Investors in Stock Microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1310–1319, Stroudsburg, PA, USA. Association for Computational Linguistics. event-place: Edinburgh, United Kingdom.

Batra, S. and Rao, D. (2010). Entity based sentiment analysis on twitter. *Science*, 9(4):1–12.

Choi, Y., Kim, Y., and Myaeng, S.-H. (2009). Domain-specific sentiment analysis using contextual feature generation. In *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 37–44, Hong Kong, China. ACM.

Costa, H., Zaretskaya, A., Corpas Pastor, G., and Seghiri Domínguez, M. (2016). Nine terminology extraction Tools: Are they useful for translators?

Ding, X. and Liu, B. (2007). The utility of linguistic rules in opinion mining. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 811–812, Amsterdam, The Netherlands. ACM.

Gaillat, T., Sousa, A., Zarrouk, M., and Davis, B. (2018a). FinSentiA: Sentiment Analysis in English Financial Microblogs. In *Conférence CORIA TALN 2018*, Proceedings of the TALN - CORIA 2018, Rennes, France. INRIA.

Gaillat, T., Zarrouk, M., Freitas, A., and Davis, B. (2018b). The SSIX corpora: three gold standard corpora for sentiment analysis in English, Spanish and German financial microblogs. In *The Language Resources and Evaluation Conference*, pages 2671–2675, Miyazaki, Japan, May. European Language Resources Association.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300.

Hajek, P., Olej, V., and Myskova, R. (2014). Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making. *Technological and Economic Development of Economy*, 20(4):721–738.

Hatzivassiloglou, V. and McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*, pages 174–181. Association for Computational Linguistics.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., and Suchomel, V. (2013). The TenTen Corpus Family. In *7th International Corpus Linguistics Conference CL*, pages 125–127.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., and Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, pages 7–36.

Klein, F. C. and Prestbo, J. A. (1974). *News and the Market*. H. Regnery Co.

Krishnamoorthy, S. (2018). Sentiment analysis of financial news articles using performance indicators. *Knowledge and Information Systems*, 56(2):373–394, August.

Liu, B. (2015). *Sentiment analysis: mining opinions, sentiments, and emotions*. Cambridge University Press, New York, NY.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65.

Loughran, T. and McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, 69(4):1643–1671.

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.

Lu, Y., Castellanos, M., Dayal, U., and Zhai, C. (2011). Automatic Construction of a Context-aware Sentiment Lexicon: An Optimization Approach. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 347–356, New York, NY, USA. ACM.

Lutz, B., Pröllochs, N., and Neumann, D. (2019). Sentence-Level Sentiment Analysis of Financial News Using Distributed Text Representations and Multi-Instance Learning. In *Proceedings of the Hawaii International Conference on System Sciences*.

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., and Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, 65(4):782–796, April.

Mao, H., Counts, S., and Bollen, J. (2011). Computational economic and finance gauges: Polls, search, and twitter. In *Meeting of the National Bureau of Economic Research-Behavioral Finance Meeting, Stanford, CT*, volume 11, page 2011.

Mao, H., Gao, P., Wang, Y., and Bollen, J. (2014). Automatic construction of financial semantic orientation lexicon from large-scale Chinese news corpus. *Institut Louis Bachelier, 20 (2)*, pages 1–18.

Meyer, B., Bikdash, M., and Dai, X. (2017). Fine-grained financial news sentiment analysis. In *Southeast-Con 2017*, pages 1–8, Concord, NC, USA, March. IEEE.

Moreno-Ortiz, A. and Fernández-Cruz, J. (2015). Identifying Polarity in Financial Texts for Sentiment Analysis: A Corpus-based Approach. *Procedia - Social and Behavioral Sciences*, 198:330–338, July. 00002.

Moreno-Ortiz, A. and Perez-Hernandez, C. (2018). Lingmotif-lex: a Wide-coverage, State-of-the-art Lexicon for Sentiment Analysis. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 2653–2659, Miyazaki, Japan. European Language Resources Association (ELRA).

Moreno-Ortiz, A. (2017a). Lingmotif: A User-focused Sentiment Analysis Tool. *Procesamiento del Lenguaje Natural*, 58(0):133–140, March.

Moreno-Ortiz, A. (2017b). Lingmotif: Sentiment Analysis for the Digital Humanities. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 73–76, Valencia, Spain, April. Association for Computational Linguistics.

Niederhoffer, V. (1971). The analysis of world events and stock prices. *The Journal of Business*, 44(2):193–219.

O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., and Smeaton, A. F. (2009). Topic-dependent sentiment analysis of financial blogs. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 9–16. ACM.

Oliveira, N., Cortez, P., and Areal, N. (2016). Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decision Support Systems*, 85:62–73, May.

Pang, B. and Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.

Perrie, J., Islam, A., Milios, E., and Keselj, V. (2013). Using google n-Grams to expand word-emotion association lexicon. In Alexander Gelbukh, editor, *Computational linguistics and intelligent text processing*, pages 137–148, Berlin, Heidelberg. Springer Berlin Heidelberg.

Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 105–112, Stroudsburg, PA, USA. Association for Computational Linguistics. ACM ID: 1119369.

Riloff, E., Patwardhan, S., and Wiebe, J. (2006). Feature subsumption for opinion analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 440–448, Stroudsburg, PA, USA. Association for Computational Linguistics. ACM ID: 1610137.

Shapiro, A. H. and Wilson, D. (2019). Taking the Fed at its Word: Direct Estimation of Central Bank Objectives using Text Analytics. Federal Reserve Bank of San Francisco.

Shiller, R. C. (2000). Irrational exuberance. *Philosophy and Public Policy Quarterly*, 20(1):18–23.

Sinha, A., Kedas, S., Kumar, R., and Malo, P. (2019). Buy, Sell or Hold: Entity-Aware Classification of Business News. Technical report, Working Paper. Indian Institute of Management Ahmedabad, Research and Publication Department.

Stone, P. J. and Hunt, E. B. (1963). A Computer Approach to Content Analysis: Studies Using the General Inquirer System. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference*, AFIPS '63 (Spring), pages 241–256, New York, NY, USA. ACM.

Tai, Y.-J. and Kao, H.-Y. (2013). Automatic Domain-Specific Sentiment Lexicon Generation with Label Propagation. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services*, IIWAS '13, pages 53:53–53:62, New York, NY, USA. ACM.

Tetlock, P. C. (2007). Giving content to investor sentiment:

The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Thaler, R. H. (2015). *Misbehaving: the making of behavioral economics*. W.W. Norton & Company, New York, first edition edition.

Van de Kauter, M., Breesch, D., and Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with applications*, 42(11):4999–5010.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 347–354, Stroudsburg, PA, USA. Association for Computational Linguistics.

Yu, Y., Duan, W., and Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55(4):919–926, November.