Data Driven Creation of Sentiment Dictionaries for Corporate Credit Risk Analysis

Full paper

Aaron Mengelkamp
University of Goettingen, Germany
amengel@uni-goettingen.de

Sebastian Wolf
University of Goettingen, Germany
sebastian.wolf@stud.uni-goettingen.de

Matthias Schumann
University of Goettingen, Germany
mschuma1@uni-goettingen.de

Abstract

It has been shown, that German-language user generated content can improve corporate credit risk assessment, when sentiment analysis is applied. However, the approaches have only been conducted by human coders. In order to automate the analysis, we construct 20 domain-dependent sentiment dictionaries based on parts of a manually classified corpus from Twitter. Then, we apply the dictionaries to the remaining part of the corpus and rank the dictionaries based on their accuracy. Results from McNemar's tests indicate, that the three best dictionaries do not differ significantly, but significant difference can be assured regarding the first and the fourth dictionary in the ranking. In addition to that, a general German-language dictionary is inferior compared to the constructed dictionaries. The results emphasize the importance of domain-dependent dictionaries in German-language sentiment analysis for future research. Furthermore, practitioners can utilize the dictionaries in order to create an additional indicator for corporate credit risk assessment.

Keywords

Credit risk analysis, sentiment analysis, social media, user generated content.

Introduction

Many researchers have shown, that textual data and particularly user generated content (UGC) from social media platforms contains valuable information for financial risk management (Bollen et al. 2011; Kearney and Liu 2014). Especially automated sentiment analysis of English-language textual data for financial market prediction has received great attention from researchers in the past years, whereby foreign-language data has only received minor attention (Kearney and Liu 2014). Next to a transfer of new approaches in risk management to foreign-language countries, analysis of this data will enable English speaking risk managers to get information from people whose language they do not speak. In addition to scientific results, startups such as Iwoca, Kabbage and onDeck offer corporate credits based on the analysis of company related web-content. Still, most of the startups are dependent on venture capitalists and have to take default rates into account, which are substantially higher, than default rates of traditional creditors, e.g. banks (Buhse 2016). This indicates, that the algorithms they use, need to be improved in order to enable profitable business models. Next to that, the companies do not reveal their scoring algorithms, because these are part of their competitive advantage. Hence, the methods need to be evaluated from a scientific perspective. For this purpose, Mengelkamp et al. (2015) have demonstrated, that German-language UGC can enrich the data base for corporate credit risk analysis, which indicates, that corporate credit risk assessment, as performed by the startups, is possible. However, the approach has only been conducted by human coders up to today. Automated approaches could reduce the costs associated with the analysis. To this effect, dictionary based coding of textual data is a widely established and easy to apply method for automated sentiment analysis (Neuendorf 2002). Furthermore, it has been shown, that domain specific sentiment dictionaries attain better performance measures than general sentiment dictionaries (Loughran and McDonald 2011). Hence, the research question we aim to answer in this article is stated as:

In how far can the analysis of German textual UGC in the domain of corporate credit risk analysis be automated by utilizing domain and language specific sentiment dictionaries?

In order to answer the research question, we outline the theoretical background and key concepts in the next chapter. Then, we present an overview regarding related literature. Subsequently, we follow the methodological approach of Oliveira et al. (2014) who start with an introduction of the data set and data preparation approaches, whereupon the sentiment dictionaries are created. In the end, the performance of the sentiment dictionaries is evaluated and discussed. We finish the paper by highlighting implications, limitations and future research opportunities.

Theoretical Background

The key concepts, which are corporate credit risk analysis, social media and UGC, as well as sentiment analysis, are delineated in the following.

Corporate Credit Risk Analysis

Lenders in corporate credit relationships are exposed to the risk of payment default. Hence, they attempt to only give credit to borrowers who are likely to pay their debts (Stiglitz and Weiss, 1981). Nevertheless, the borrower's actions, which can influence their solvency, cannot be monitored perfectly (Stiglitz, 1988). These are characteristics of a principal-agent problem in which "the principal does not know whether the agent undertook the action the principal would himself have undertaken, in the given circumstances" (Stiglitz, 1988, pp. 1 f.). Thus, a credit relationship can be classified as a principal agent problem (Stiglitz, 1988). Lenders in the role of the principal use several screening devices and sources of information in order to assess the credit worthiness of borrowers. The asymmetry of information is usually greater for trade credits than for bank loans, because banks have access to more information (e. g. checking accounts), if the borrower is a regular customer (Beck, 2014). The asymmetry of information between the principal and the agent is reduced during the process of credit assessment (Stiglitz and Weiss, 1981). For this purpose, internal information, which is available to the lender due to previous credit agreements (e.g. payment behavior), and external information (e.g. annual reports or ratings from credit agencies) is used. External information can be acquired from third parties or by voluntary disclosure, directly from the debtor (Graham and Coyle, 2000). Then, typical characteristics of good and bad debtors are identified with statistical methods such as discriminant/regression analysis or the k-Nearest-Neighbor algorithm (Helfrich 2010). If a potential debtor reveals poor characteristics, then the credit is likely to be denied or more rigorous interest rates or payment terms are imposed (Keßler, 2010).

Social Media and UGC

Kaplan and Haenlein (2010) define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content". Hereby, UGC encompasses all data, which is created by the use of social media by people, e.g. posts on Facebook or Twitter as well as network connections between user accounts on the platforms. In the following, only textual UGC is of relevance as described in the introduction. Although not explicitly stated by Kaplan and Haenlein (2010), we also consider customer ratings and comments published on e-commerce platforms such as Amazon and Ebay as UGC, because the mentioned startups partially rely on this data in order to establish their scoring models (Kabbage 2015).

Sentiment Analysis

In order to extract information from textual UGC, the texts are usually transferred into a numerical format by use of sentiment analysis techniques (e.g. negative numerical values represent negative sentiment whereas positive values indicate positive sentiment and neutral sentiment refers to a value of 0) (Stieglitz et al. 2014). Hereby, the texts are pre-processed according to opinion or sentiment related attributes, which represent people's "views, attitudes, appraisals and emotions towards entities" (Stieglitz et al. 2014). The entities in the context of this study are corporates whose risk of payment default shall be evaluated.

According to a review conducted by Kearney and Liu (2014), two approaches (dictionary-based and machine learning) are prevalent in sentiment analysis. Utilizing the dictionary-based approach, words or phrases in the texts are classified into predefined dictionary categories (e.g. negative, neutral, positive). Hereby, the word lists for each sentiment category in the dictionary and the weighting of words are important. In order to apply machine learning approaches, training sets of textual data need to be manually classified at first (Kearney and Liu 2014). These training sets are then passed on to an algorithm (e.g. Naïve Bayesian, regression algorithms, or support vector machines), which derives rules for text classification with statistical inference from the training data. The rules are then deployed in order to automatically classify new textual data.

Kearney and Liu (2014) identified a majority of dictionary-based approaches, which are easier to apply. We limit our study to simple dictionary based coding as well, because automated sentiment analysis of UGC for corporate credit risk assessment has not been conducted yet for German-language data.

Related Literature

Sentiment dictionaries have been created by researchers for various domains. The most used word lists were established by Stone et al. (1968) and Hart (2000). These include many categories such as certainty, optimism or realism. The dictionaries are applicable for the general English language rather than a specific domain (Kearney and Liu 2014). Baccianella et al. (2010) presented the dictionary SentiWordNet, which is limited to three categories (positivity, neutrality and negativity). These categories are predominately used today (Kearney and Liu 2014). In addition to that, it has been ascertained, that domain specific sentiment dictionaries attain better results than general sentiment dictionaries (Henry 2006; Loughran and McDonald 2011). Loughran and McDonald (2011) state, that almost three-fourth of the negative words in general dictionaries are not associated with negative sentiment in the financial domain. Oliveira et al. (2014) created a sentiment dictionary especially for stock market prediction using data from the social media platform StockTwits. They are able to significantly improve the sentiment classification in the stock market domain compared to studies which utilize general dictionaries.

Next to English-language sentiment dictionaries, Kaji and Kitsuregawa (2007) as well as Kanayama and Nasukawa (2006) have created sentiment dictionaries for the Chinese language. Abdulla et al. (2014) and Mahyoub et al. (2014) have done the same for Arabic. Furthermore, dictionaries for Czech, Danish and Spanish have been created (Jijkoun and Hofmann 2009; Kincl et al. 2013; Molina-González et al. 2013).

Remus et al. (2010) developed the first German-language dictionary (SentiWortSchatz) for sentiment analysis. It is based on the translated word list of Stone et al. (1968) and was extended for a study on the effects of newspaper articles and blog posts on the 30 companies of the German stock index. They report competitive evaluation figures and conclude, that SentiWortSchatz is "a useful resource for sentiment analysis related tasks" (Remus et al. 2010). Nevertheless, a comparison to other sentiment analysis based work in German was not possible due to the non-existence.

The summary shows, that sentiment analysis has made progress especially in the English language. Researchers can draw on many established dictionaries and apply or adapt these to their textual data. Sentiment dictionaries in other languages have not been evaluated to a great extent and partially rely on translations of available English-language resources revealing substantial research opportunities (Kearney and Liu 2014; Remus et al. 2010).

Data Set

We rely on the manually classified data set created by Mengelkamp et al. (2015) in order to construct German-language sentiment dictionaries for the domain of corporate credit risk analysis, because it has been proven, that hints for financial (in)stability of companies are present in the data. Thus, positive and negative words, which indicate financial strength or weakness, should be extractable. Furthermore, it is the only data set known to us, which is available in the considered domain. The data set is composed of 7071 posts from the microblogging platform Twitter, which are called Tweets. The Tweets are related to the ten biggest German companies which filed for bankruptcy in 2013. A reference group was build, which encompasses a solvent counterpart for each of the bankrupt companies. For each pair of companies, Tweets in the year prior to the date on which the bankrupt company filed for insolvency, were extracted. Then, the Tweets were manually labelled by two human coders independently. They assigned negative (-2 and -1), neutral (0) and positive (1 and 2) sentiment scores to the Tweets. The coders achieved an intercoder reliability of 93,81 %, which is above the threshold of 80 % generally accepted by researchers (Neuendorf 2002). The final coding was determined during a joint verification in which the differently labelled Tweets were discussed. The number of Tweets in each sentiment category is pictured in Table 1.

	Sentiment Scores				
	-2	-1	0	1	2
Number of Tweets	2787 (39,4 %)	393 (5,6 %)	3139 (44,4 %)	558 (7,9 %)	194 (2,7 %)

Table 1: Number of Tweets in Sentiment Categories

Data Preparation

Before we constructed sentiment dictionaries, we transformed the sentiment scores in order to narrow down the categories because classification results usually improve, when fewer categories are used (Nassirtoussi et al. 2014). Thus, Tweets with scores of -1 and -2 were merged into one category with a score of -1 and Tweets with scores of 1 and 2 were united into a category with a score of 1. Then, we executed operations for data cleansing, because these usually improve the results (Kearney and Liu 2014). Hereby, we deleted the following elements according to Oliveira et al. (2014):

- Hyperlinks
- Special characters including @, €, $ and Hashtags
- Numbers
- Punctuation marks

In addition to that, we created a list with stop words, which are neglected during dictionary creation, because they are not meaningful for sentiment classification. These stop words encompass:

- Company names
- Terms which represent legal forms of companies
- Names of news agencies and other information providers
- Locations (e.g. city names)
- Articles, prepositions and pronouns

Then, we filtered out duplicates and divided the data set into two samples, one used for creation of the dictionaries and the other used for evaluation. Hence, 2/3 of the Tweets were used for dictionary creation and 1/3 for evaluation. The proportioning is in accordance with similar research endeavors. Nassirtoussi et al. (2014) for example state, that approximately 70 % of the data is used for the derivation of classification rules and 30 % for evaluation.

Sentiment Dictionary Creation

For dictionary creation, we ranked the words from Tweets of each sentiment category (negative, neutral and positive) descending based on their frequency counts, because frequency based sentiment analysis has been applied widely in English literature and is the most basic technique (Kearney and Liu 2014). The number of words extracted for each sentiment category is listed in Table 2. Loughran and McDonald (2011) state, that "raw word counts are not the best measure of a word's information content." Instead, the importance of terms in a document as well as the whole corpus and the document length should be normalized in some form (Loughran and McDonald 2011). Tests, which we conducted, revealed, that almost all words appeared only once in a Tweet. Hence, the frequency count of words is equal to the number of documents in which a term appears. Furthermore, Tweets are about the same length. Thus, most lists, which we created based on different term weighting, did not differ from the lists based on frequency counts. Therefore, different term weighting is not pursued in the following.

Sentiment Score	Number of Words
Negative: -1	1700
Neutral: 0	2076
Positive: 1	673

Table 2: Number of Words in Sentiment Categories

Based on the word lists, 20 different sentiment dictionaries were created. The first dictionary includes all words in the 5 % quantile of the lists related to each sentiment score. Each word is then labelled with the score of the list in which it appears. Hence, the 5 % of words, which occurred most often in each sentiment category, are incorporated in this dictionary. For construction of the second dictionary, the 10 % quantile was considered and so forth, until the 20th dictionary was created, which is comprised of all words (100 % quantile). This enables us to evaluate, how many of the most frequent words from each sentiment category should be used for automated analysis of unclassified data in order to achieve the best classification accuracy. While word lists with few words are expected to have low hit ratios and hence, low classification accuracy, lists, which are to comprehensive, will include meaningless terms, which are expected to hamper the results.

For application of each dictionary, the sentiment polarity of a Tweet (SPT) is calculated as stated in Equation (1). Hereby, for each word (i) in a Tweet, the sentiment polarity of the word (SPW) in the applied dictionary is determined. If a word does not appear in the dictionary, it is neglected. The sentiment scores of all words are added up and divided by the total number of words which occur in the Tweet and the dictionary (n). If n is equal to zero, then the SPT is considered to be zero as well.

$$SPT = \frac{\sum_{i=1}^{n} SPW_i}{n} \tag{1}$$

Sentiment Dictionary Evaluation

In order to evaluate which of the 20 dictionaries achieves the best performance, we calculated the accuracy, which is the percentage of accordance from automated and manual coding, for all dictionaries (Kincl et al. 2013). Since the scores of automated coding can be fractional digits, a direct comparison with the manual coding is not be possible due to the differing number format. Thus, we analyzed, in how far a different mapping of automatically calculated sentiment scores to the established categories (-1, 0, 1) affects dictionary performance. For this purpose, we assigned automatically coded Tweets within the range of -0.1 and 0.1 to the neutral category and successively extended the threshold in steps of 0.1 in both directions. Hence, all Tweets within the range of -0.2 and 0.2 were labelled as neutral in the next iteration and so forth. The results indicate, that dictionary performance decreases, when the thresholds for Tweets to be considered as neutral are applied and extended. Thus, thresholds were not applied and Tweets with negative sentiment scores from automated coding were labeled with -1 and Tweets with positive sentiment scores were labeled with 1. Only Tweets, which do not contain words with positive or negative polarities, were labeled with a score of 0, which represents neutral sentiment. The results for this classification are depicted in Figure 1. Hereby, the different dictionaries are listed on the X-axis. The accuracy of each dictionary is presented on the Y-axis. It can be comprehended, that the best performance (accuracy of 67.9 %) is achieved, when 20 % of the most frequent words in Tweets of each sentiment category are incorporated in the dictionary. The second best dictionary (accuracy of 66.6 %) contains 30 % of the words. All dictionaries achieve better accuracies than a random classification which would be 33.3 % in the underlying scenario with three sentiment categories.

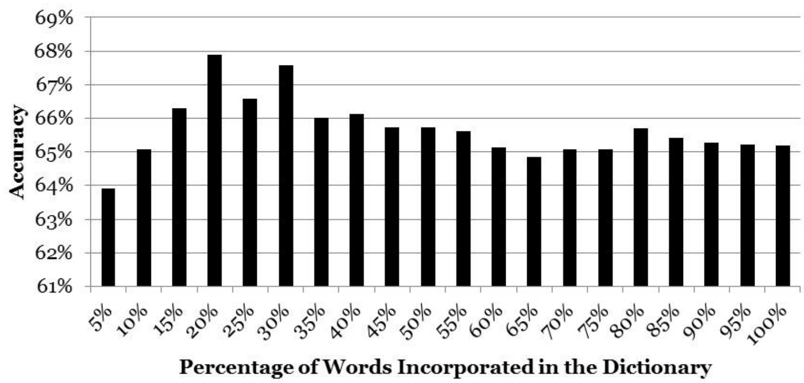

Figure 1: Dictionary Performance

Then, we verified if the differences in accuracy between the best dictionaries are significant. We applied McNemar's test for this purpose. The null hypothesis of the test states, that two classifiers achieve the same classification accuracy (Lan et al. 2009). In the present case, the classifiers are represented by two sentiment dictionaries (DA and DB). Based on the automated classification of the dictionaries, a contingency table as shown in Table 3 is constructed (Dietterich 1998; Lan et al. 2009).

N00: Number of Tweets misclassified by both classifiers DA and DB	N01: Number of Tweets misclassified by DA but not DB
N10: Number of Tweets misclassified by DB but not DA	N11: Number of Tweets misclassified by neither DA nor DB

Table 3: McNemar's Test Contingency Table

Then, χ is calculated based on the contingency table according to equation 2. χ is approximately χ^2 distributed with one degree of freedom. If χ is greater than 6.64, this indicates, that the null hypothesis can be disregarded with a statistical significance of 0.01. Thus, the dictionaries differ significantly from each other (Lan et al. 2009).

$$\chi = \frac{(|N01 - N10|) - 1)^2}{N01 + N10} \quad (2)$$

McNemar's test revealed, that the difference in accuracy between the best and the second as well as the third-best dictionary in the ranking is not significant. The significance levels are above the threshold of 0.1, which is the highest generally accepted by researchers (Field 2009). However, the fourth-best dictionary, which was constructed based on 15 % of the most frequent words in the sentiment categories, differs significantly from the best. Hence, the dictionaries, which contain 20 % - 30 % of the most frequent words in the sentiment categories, should be applied for dictionary-based sentiment analysis in the domain of corporate credit risk analysis. The statistical significant differences might not be of relevance to practitioners since the technique has shown to be robust due to the maximal difference of only 4 % between the best and the worst dictionary. The results of McNemar's tests from the comparison of the best dictionary with the second, third and fourth-best are summarized in Table 4. The second, third and fourth dictionaries are listed according to their classification accuracy. Then, the significance levels of McNemar's tests for the best compared with the three dictionaries are stated. Significant difference of the best with all other dictionaries could also be established.

Rank	Most Frequent Words	Accuracy	Significance Level
2	30 %	67.53 %	>0.1
3	25 %	66.54 %	>0.1
4	15 %	66.21 %	<0.01***

Table 4: McNemar's Tests of the 1st (20 %) Dictionary (Accuracy of 67,9 %)

In addition to the comparison of the created dictionaries, we applied the only sentiment dictionary available in German (SentiWortSchatz), which was created by Remus et al. (2010), to the evaluation corpus. The dictionary achieved an accuracy of 49.97 %, which outperforms a random classification but is inferior to all domain dependent dictionaries created within the present study. An analysis utilizing McNemar's test revealed, that the difference between SentiWortSchatz and the best dictionary created by us, can be verified on all established levels of significance, which are 0.1, 0.05 and 0.01 (Field 2009).

Implications, Limitations and Future Research

The aim of our study was to analyze in how far the analysis of German textual UGC in the domain of corporate credit risk analysis can be automated by utilizing domain and language specific sentiment dictionaries. We constructed 20 German-language dictionaries based on the manually classified corpus of Mengelkamp et al. (2015) for this purpose.

Results of the dictionary evaluation revealed, that the three best dictionaries achieve classification accuracies of 67.90 %, 67.53 % and 66.54 %. The differences in classification accuracy were found not to be significant by McNemar's tests. However, the next best dictionary, which achieves an accuracy of 66.21 %, significantly differs from the best dictionary. Furthermore, the only domain-independent sentiment dictionary in German achieves an accuracy of 49.97 %, when applied to our evaluation corpus. Thus, it is inferior to all dictionaries created within the present study.

The findings reinforce the need of domain-dependent sentiment dictionaries in order to improve sentiment analysis especially in the German-language, because these are superior to currently available resources for German-language sentiment analysis. Next to the propagation of sentiment analysis into non-English speaking countries, researchers will be enabled to gain insight into textual UGC, which is verbalized in languages they are not capable of.

Implications for the principal agent theory include additional risk indicators in the domain of corporate credit risk analysis, which can be created based on our results. When textual UGC related to corporates is analyzed with the developed sentiment dictionaries, decision support in this domain can be improved. Thus, the information asymmetry between the principal and the agent in credit relationships can be further reduced.

In addition to that, practitioners (e.g. credit decision makers) can utilize the dictionaries in order to automatically monitor textual UGC related to their debtors and be immediately informed about negative posts. Based on the instant notification, they can adjust e.g. credit limits and interest rates for high risk debtors. Thus, implications for practice encompass lower default rates and bad debt when the sentiment dictionaries are appropriately utilized in corporate credit risk assessment.

It has to be taken into account, that the results are limited by size and scope of the data set. In future research, it should be investigated, if the outcome can be confirmed, when the dictionaries are applied to heterogeneous data sets, e.g. from other social media platforms. In addition to that, we took 5 % steps when selecting words from the list with frequency counts. It is possible, that a better dictionary can be constructed, when other percentages from the word frequency count list are considered.

Future research should also focus on different term weighting methods, when heterogeneous texts are analyzed. In the data set from Mengelkamp et al. (2015), words occurred once in a Tweet only and hence, frequency counts and the quantity of documents, which contain a term, are similar. Thus, dictionaries created based on different term weighting have been neglected in our study. Still, more complex term weighting methods might improve dictionary performance. Next to that, recent research in English-language sentiment analysis has shown, that machine learning algorithms obtain superior results when compared to dictionary-based approaches (Kearney and Liu 2014). Thus, these should be evaluated in German-language sentiment analysis as well.

REFERENCES

Abdulla, N., Majdalawi, R., Mohammed, S., Al-Ayyoub, M., and Al-Kabi, M. 2014. "Automatic Lexicon Contruction for Arabic Sentiment Analysis," in *Proceedings of the 2nd International Conference on Future Internet of Things and Cloud*, Institute of Electrical and Electronic Engineers (ed.), Barcelona, pp. 547–552.

Baccianella, A., Esuli, A., and Sebastiani, F. 2010. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in *Proceedings of the 7th conference on International Language Resources and Evaluation*, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner and D. Tapias (eds.), Malta, pp. 2200–2204.

Bollen, J., Mao, H., and Zeng, X.-J. 2011. "Twitter mood predicts the stock market," *Journal of Computational Science* (2:1), pp. 1–8.

Buhse, M. 2016. "Kreditwürdig getwittert," *Technology Review* (2), pp. 78–79.

Dietterich, T. G. 1998. "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," *Neural Computation* (10:7), pp. 1895–1923.

Field, A. 2009. *Discovering Statistics using SPSS*, London: Sage Publications.

Hart, R. 2000. *Diction 5.0*. http://rhetorica.net/diction.htm. Accessed 10 February 2016.

Helfrich, M. 2010. *Kreditscoring und Scorewertbildung der Schufa*, Münster.

Henry, E. 2006. *Are Investors Influenced by How Earnings Press Releases are Written?* http://ssrn.com/abstract=933100. Accessed 7 April 2014.

Jijkoun, V., and Hofmann, K. 2009. "Generating a non-English Subjectivity Lexicon: Relations That Matter," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics (ed.), pp. 398–405.

Kabbage 2015. *How Kabbage Loans Work*. https://www.kabbage.com/how-it-works/. Accessed 19 January 2015.

Kaji, N., and Kitsuregawa, M. 2007. "Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents," in *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computing Machinery (ed.), Prague, pp. 1075–1083.

Kanayama, H., and Nasukawa, T. 2006. "Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics (ed.), Stroudsburg, pp. 355–363.

Kaplan, A. M., and Haenlein, M. 2010. "Users of the world unite! The challenges and opportunities of Social Media," *Business Horizons* (53:1), pp. 59–68.

Kearney, C., and Liu, S. 2014. "Textual sentiment in finance: A survey of methods and models," *International Review of Financial Analysis* (33:1), pp. 171–185.

Kincl, T., Novák, M., and Pribil, J. 2013. "Getting Inside the Minds of the Customers: Automated Sentiment Analysis," in *Proceedings of the 9th European Conference on Managment Leadership and Governance*, Academic Conferences and Publishing (ed.), Klagenfurt, pp. 122–128.

Lan, M., Tan, L. C., Su, J., and Lu, Y. 2009. “Supervised and Traditional Term Weighting Methods for Automatic Text Categorization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* (31:4), pp. 721–735.

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), pp. 35–65.

Mahyoub, F. H., Siddiqui, M. A., and Dahab, M. Y. 2014. “Building an Arabic Sentiment Lexicon Using Semi-supervised Learning,” *Journal of King Saud University - Computer and Information Sciences* (26:4), pp. 417–424.

Mengelkamp, A., Hobert, S., and Schumann, M. 2015. “Corporate Credit Risk Analysis Utilizing Textual User Generated Content - A Twitter Based Feasability Study,” in *Proceedings of the 19th Pacific Asia Conference on Information Systems*, Singapore.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M., and Perea-Ortega, J. M. 2013. “Semantic orientation for polarity classification in Spanish reviews,” *Expert Systems with Applications* (40), pp. 7250–7257.

Nassirtoussi, A. K., Wah, T. Y., Aghabozorgi, S. R., and Ling, D. N C. 2014. “Text mining for market prediction: A systematic review,” *Expert Systems with Applications* (41:16), pp. 7653–7670.

Neuendorf, K. A. 2002. *The Content Analysis Guidebook*, Chicago: Sage Publications.

Oliveira, N., Cortez, P., and Areal, N. 2014. “Automatic Creation of Stock Market Lexicons for Sentiment Analysis Using StockTwits Data,” in *Proceedings of the 18th International Database Engineering & Applications Symposium*: *IDEAS'14 : Instituto Superior de Engenharia do Porto, Porto, Portugal, July 7-9, 2014*, B. C. Desai, A. Almeida, J. Bernardino and E. F. Gomes (eds.), pp. 115–123.

Remus, R., Quasthoff, U., and Heyer, G. 2010. “SentiWS - A Publicly Available German-language Resource for Sentiment Analysis,” in *Proceedings of the International Conference on Language Resources and Evaluation*, N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis and M. Rosner (eds.), Valletta, pp. 1168–1171.

Stieglitz, S., Dang-Xuan, L., Bruns, A., and Neuberger, C. 2014. “Social Media Analytics,” *Business & Information Systems Engineering* (6:2), pp. 89–96.

Stone, P. J., Dunphy, D. C., Smith, M. S., and Olgilvie, D. M. 1968. “The General Inquirer: A Computer Approach to Content Analysis,” *Journal of Regional Science* (8:1), pp. 113–130.