Creating Sentiment Dictionaries: Process Model and Quantitative Study for Credit Risk

Aaron Mengelkamp[1], Kevin Koch[1] and Matthias Schumann[2]
[1]Hanover University of Applied Sciences, Hanover, Germany
[2]Georg-August-University Goettingen, Goettingen, Germany
aaron.mengelkamp@hs-hannover.de
kevin_koch@arcor.de
mschuma1@uni-goettingen.de

Abstract: Since textual user generated content from social media platforms contains valuable information for decision support and especially corporate credit risk analysis, automated approaches for text classification such as the application of sentiment dictionaries and machine learning algorithms have received great attention in recent user generated content based research endeavors. While machine learning algorithms require individual training data sets for varying sources, sentiment dictionaries can be applied to texts immediately, whereby domain specific dictionaries attain better results than domain independent word lists. We evaluate by means of a literature review how sentiment dictionaries can be constructed for specific domains and languages. Then, we construct nine versions of German sentiment dictionaries relying on a process model which we developed based on the literature review. We apply the dictionaries to a manually classified German language data set from Twitter in which hints for financial (in)stability of companies have been proven. Based on their classification accuracy, we rank the dictionaries and verify their ranking by utilizing Mc Nemar's test for significance. Our results indicate, that the significantly best dictionary is based on the German language dictionary SentiWortschatz and an extension approach by use of the lexical-semantic database GermaNet. It achieves a classification accuracy of 59,19 % in the underlying three-case-scenario, in which the Tweets are labelled as negative, neutral or positive. A random classification would attain an accuracy of 33,3 % in the same scenario and hence, automated coding by use of the sentiment dictionaries can lead to a reduction of manual efforts. Our process model can be adopted by other researchers when constructing sentiment dictionaries for various domains and languages. Furthermore, our established dictionaries can be used by practitioners especially in the domain of corporate credit risk analysis for automated text classification which has been conducted manually to a great extent up to today.

Keywords: sentiment dictionaries, credit risk, Twitter analysis, user generated content, text mining

1. Introduction

User Generated Content (UGC) from social media platforms, which are "a group of internet-based applications that allow the creation and exchange of user generated content" (Kaplan and Haenlein, 2010), contains valuable information for decision making especially when extracting its sentiment. Sentiment is defined as "people's opinions in terms of views, attitudes, appraisals and emotions" (Kearney and Liu, 2014; Nassirtoussi et al., 2014; Stieglitz et al., 2014). Since the amount of UGC continuously increases, manual classification of the textual content is not appropriate due to the high effort associated with human coders (Neuendorf, 2002). Hence, automated real time analysis of UGC is desirable, which can be realized by sentiment dictionaries. Sentiment dictionaries consist of word lists to which sentiment scores are assigned. Although machine learning algorithms usually achieve better classification results, dictionary-based approaches are more suitable especially when the data is of great heterogeneity and a training of machine learning algorithms is difficult. Since domain specific sentiment dictionaries attain better classification results, we examine the first research question (Kearney and Liu, 2014):

How can sentiment dictionaries be created and evaluated?

In order to verify the applicability of our answer, we create and evaluate sentiment dictionaries for corporate credit risk analysis since recent studies in this domain indicate, that German language UGC contains valuable information (Mengelkamp, Hobert and Schumann, 2015). In order to do so, we answer the second research question:

How good do German sentiment dictionaries perform on textual UGC in the domain of corporate credit risk analysis?

After conducting a literature review in order to identify sentiment dictionaries, we analyse, how the dictionaries were constructed and derive a process model for creation and evaluation of domain and language specific

sentiment dictionaries. Then, we construct nine versions of German sentiment dictionaries based on the established process model. We apply these to a data set from Twitter, which has been proven to contain evidence for financial instability of companies. The results of the automated coding are compared to a manual coding performed by (Mengelkamp, Hobert and Schumann, 2015). Furthermore, we rank the dictionaries according to their classification accuracy, test for significant differences and discuss the results.

2. Approaches for sentiment dictionary creation – a systematic literature review

We describe the research methodology applied for the literature review which we conducted in order to identify existing sentiment dictionaries and how they can be constructed, in the following chapter. Then, we present the analysis, discuss results and draw conclusions.

2.1 Research methodology

We follow the methodological approach introduced by (Webster and Watson, 2002) as well (Levy and Ellis, 2006) for the literature review. The process is summarized in Figure 1. 6577 articles were identified by use of the listed keywords in the specified databases. 46 articles were considered as relevant for the analysis by the researchers.

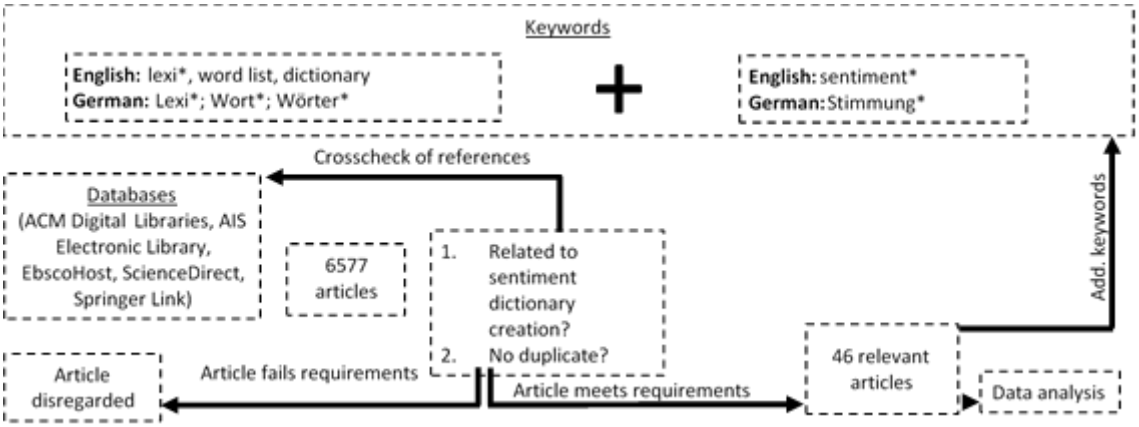

Figure 1: Methodology of literature review

The two researchers responsible for the analysis read the 46 selected articles independently from each other. They identified steps, that were carried out in order to construct sentiment dictionaries in each article. The list of steps was successively extended when new steps were identified. Then, results were discussed and names, which were assigned to the steps, were unified until the researchers agreed on each phase.

2.2 Analysis and results

In the majority of articles (31 respectively 67,4 %), English language dictionaries were constructed. Seven authors constructed dictionaries in Chinese, two for Arabic as well as German. One dictionary each was constructed for Czech, Danish, Spanish and multiple languages. The ration of domain specific to independent dictionaries is almost even since 26 (56,5 %) were created without reference to a specific domain, whereby 20 (43,5 %) were adjusted for a specific context.

The results show, that the process of dictionary creation can be divided into the superordinate phases initial construction and extension, which are described in the following chapters. The phase of initial construction was undertaken in every article, whereby a dictionary extension was conducted in only 17 articles (37 %).

2.2.1 Initial Construction

The steps of word *selection*, *polarization*, *evaluation* and *translation* are conducted in the initial construction phase. The selection and polarization phases are conducted in every article, whereby ten articles (21,7 %) translated words and 31 (67,4 %) evaluated their initial construction phase.

In the selection phase it is described how and from which sources words are extracted. Hereby, the authors of eight articles (17,8 %) relied on native speakers in order to construct sentiment word lists. In 26 of the articles (56,5 %) the words were selected from a combination of existing sentiment dictionaries. In 20 (76,9 %) of those, the word selection was conducted with software support whereas four authors (15,4 %) drafted their phrases manually and two (7,7 %) used a combined approach. Next to that, in 26 articles (56,5 %), a textual corpus was

the basis for word selection. Hereby, 20 authors (76,9 %) automatically and four (15,4 %) manually extracted words. Ghiassi et al. (2013) and Wilson et al. (2005) used a combined approach.

After the selection of words for the sentiment dictionaries, the scores and their numerical representation are defined during the polarization phase. In 10 articles (21,7 %) native speakers manually rated the selected words according to previously defined sentiment categories. In 30 of the articles (65,2 %) the polarization was based on existing sentiment dictionaries whose sentiment scores were usually transferred and adapted. A manual transfer from sentiment scores of previously established dictionaries was only performed in two cases (Bracewell, 2010; Mahyoub, Siddiqui and Dahab, 2014). 14 authors (30,4 %) utilized labels from textual corpora in order to automatically define the polarization of words.

In order to construct dictionaries in different languages, ten authors (21,7 %) translated existing dictionaries or words from labelled corpora. 7 of them relied on translation software whereas words were translated by native speakers in Mahyoub et al. (2014) only. Abdulla et al. (2014) as well as Molina-González et al. (2013) combined translation software and a manual verification.

For the evaluation of the constructed dictionaries, 26 of the authors (56,5 %) automatically applied their dictionaries to manually classified textual corpora. 5 (10,9 %) tested them against sentiment classifications of existing dictionaries and in 4 cases (8,7 %) native speakers manually evaluated the dictionaries. The authors of 15 articles (32,6 %) did not evaluate the performance of the initial construction phase. Still, all of these 15 authors conducted a dictionary extension whereupon 14 evaluated their dictionaries. Domain specific dictionaries usually attain higher rates in terms of accuracy (84,9 %) than domain independent dictionaries (61,5 %).

2.2.2 Extension

All of the authors who conducted dictionary extension algorithmically selected further words e. g. synonyms, antonyms or dialect words from lexical-semantic databases.

In order to polarize the words, which were incorporated in the dictionaries during the extension, authors rely on native speakers, sentiment dictionaries and textual corpora. 13 of the 17 authors (76,5 %), who extended their initial word lists, relied on automated extraction of polarity scores from existing sentiment dictionaries in this phase.

For the evaluation of the extended dictionaries, 15 authors (88,2 %) applied their dictionaries to already classified textual corpora.

2.2.3 Interim Conclusion

The final process for sentiment dictionary creation is summarized in Figure 2. The figure is divided into the initial construction, the extension phase and data sources. It is listed which data sources are used for the sub steps of the initial construction and the extension phase. The numbers next to the sub steps refer to the numbers assigned to the data sources at the bottom of the figure. Not all steps were undertaken by every author. Especially the evaluation and the extension phases are neglected in some articles.

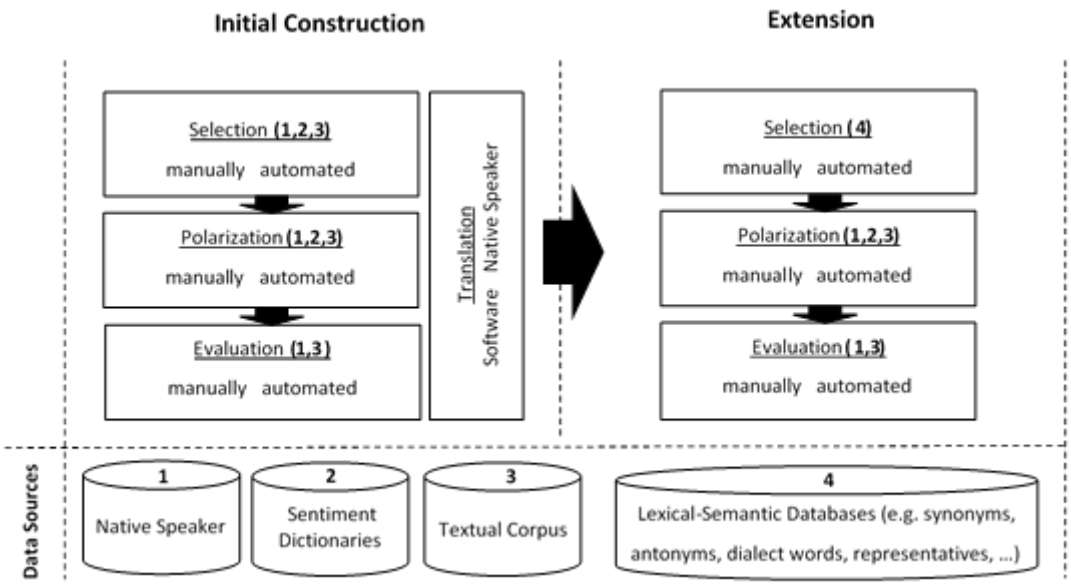

Figure 2: Process of Sentiment Dictionary Creation

Shortcomings of the identified literature include inconsistent and hence, hard to compare evaluation figures. Next to that, English is the predominant language for sentiment dictionary creation and dictionaries have not been constructed and evaluated in other languages to a great extent. Furthermore, studies about the applicability and performance of sentiment dictionaries on UGC from social media platforms in specific domains are still lacking.

3. Creating and evaluating German sentiment dictionaries for corporate credit risk analysis

In order to close the research gaps mentioned in the interim conclusion, we evaluate in how far sentiment dictionaries can be utilized to automate the process of German text classification in the domain of corporate credit risk analysis. The development of new dictionaries is based on the process model depicted in Figure 2. We evaluate nine dictionaries which we apply onto the manually classified textual corpus from Mengelkamp, Hobert and Schumann (2015).

3.1 Dictionary construction

Existing dictionaries have to fulfil the following criteria in order to be selected for automated text classification in our study:

- The sentiment categories of the dictionaries must be positive, neutral and negative since the textual corpus was classified utilizing these categories.
- The dictionaries must be publicly available.
- The language of the dictionaries must be German or English due to the language skills of the authors.

In accordance with the defined criteria, we select three existing sentiment dictionaries which we identified during the literature review to serve as a data base for selection and polarization in the initial construction phase These encompass the German language dictionary SentiWortschatz which was established by Remus et al. (2010) as well as the English language dictionary SentiWordNet 3.0 created by Baccianella et al. (2010). The second German language sentiment dictionary (GermanPolarityClues) which we identified during our literature review was created by Waltinger (2010). Since SentiWortschatz was synchronized with the GermanPolarityClues, we refrain from considering it as an independent dictionary (Waltinger, 2012). Tests on our textual corpus revealed, that the results were identical for SentiWortschatz and the GermanPolarityClues.

Since SentiwordNet 3.0 (197.036 words) encompasses more words than SentiWortschatz (31.277 words), it enables us to evaluate if more comprehensive word lists can be translated and achieve a better performance than existing German sentiment dictionaries. We use the Google Translator to translate the words into German.

Next to that, we combine SentiWortschatz and SentiWordNet 3.0 in order to test, if the classification accuracy increases, when a German sentiment dictionary is combined with a translated English dictionary. Since multiple English words are translated to the same German word, the outcome contains 128.565 entries only. If words appear multiple times, an average polarity score is calculated.

Next to that, we utilize the lexical-semantic database GermaNet in order to conduct extensions of the sentiment dictionaries according to the process model depicted in Figure 2 (Hamp and Feldweg, 1997; Henrich and Hinrichs, 2010). By doing so, we are able to evaluate in how far the performance of sentiment dictionaries can be increased, when further entries are identified by means of lexical-semantic dictionary extensions. For each of the three dictionaries (SentiWortschatz, SentiWordNet 3.0 and the combination of both), we conducted two approaches of lexical-semantic extension. In the first approach, synonyms and antonyms are extracted from GermaNet for each of the dictionaries. The polarity scores are adopted from the seed words for synonyms. Since antonyms represent words with contrary meaning, polarity scores from the seed words are multiplied with (-1). After the dictionary extension, average sentiment scores are calculated for duplicates.

The second approach for dictionary extension differs from the first in only one aspect. The sentiment polarities for synonyms and antonyms are not directly transferred, but scores are multiplied with 0,8 respectively -0,8. Hence, words identified through lexical-semantic relationships from GermaNet are included with less weighted polarity scores than the seed words. This was done because Blair-Goldensohn (2008) revealed, that sentiment dictionaries based on words identified within lexical-semantic dictionary extensions achieve best results when incorporated with 80 % of the seed word's polarity score.

The 9 dictionaries and their relations amongst each other are summarized in Figure 3. The initial versions of SentiWortschatz and SentiWordNet 3.0 are listed on the left side of Figure 3. Hereby, SentiWordNet 3.0 represents the already translated version into German. To each of the two initial dictionaries, the lexical-semantic extensions by use of GermaNet with different weighting of polarity scores are linked. Furthermore, the combination of both initial dictionaries and its lexical-semantic extensions with GermaNet are depicted in the center section.

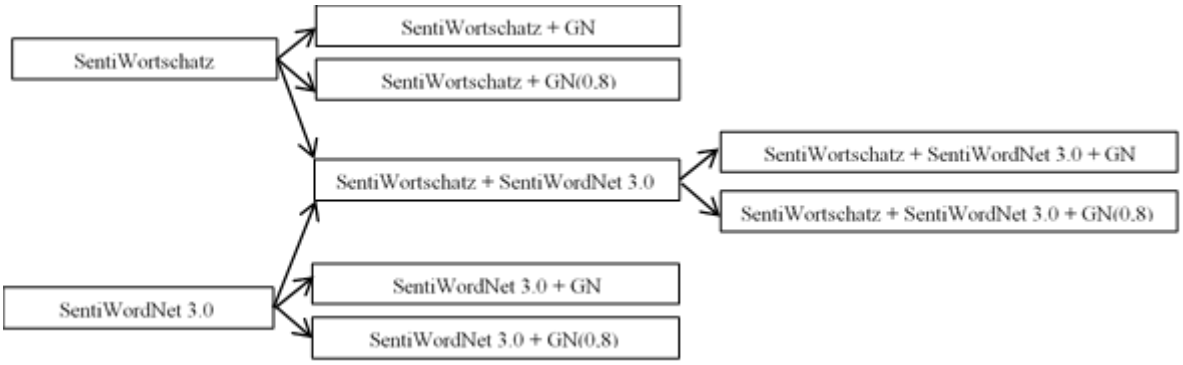

Figure 3:Sentiment Dictionaries and Relationships

3.2 Dictionary evaluation

We apply each of the dictionaries onto a textual corpus of 7071 company related German Tweets which were manually labelled by Mengelkamp, Hobert and Schumann (2015) based on content analysis guidelines established by Neuendorf (2002). The requirements for a score in order to be assigned to a Tweet are listed in the left column of Table 1. The associated scores can be seen in the right column. Since the dictionaries have only three sentiment categories (negative, neutral and positive), we summarized the scores. Hereby, we merged Tweets with scores of -2 and -1 into one category with a sentiment score of -1 and Tweets with scores of 2 and 1 into one category with a score of 1.

Table 1: Codebook

Requirement	Score
Tweet contains information regarding financial instability of the company	-2
Tweet does not contain information regarding financial instability but sentiment is negative	-1
Tweet does not contain sentiment or contains positive and negative sentiment regarding financial stability of the company	0

Requirement	Score
Tweet does not contain information regarding financial stability but sentiment is positive	1
Tweet contains information regarding financial stability of the company	2

While applying the dictionaries, the sentiment polarity of a Tweet (SPT) is calculated according to Equation (1). The sentiment polarities of words (SPW) which occur in the Tweet and the applied dictionary are summed up. Then, the sum is divided by the number of words which appear in the Tweet and the dictionary (n). By doing so, an average sentiment score considering all words in the Tweet, which are relevant for sentiment analysis, is calculated.

$$SPT = \frac{\sum_{i=1}^{n} SPW_i}{n} \mid (1)$$

Since the outcome of the dictionary-based coding results in numbers with fractional digits, a direct comparison with the manual coding is not possible due to the differing number format. Hence, Tweets for which a negative value is calculated are labelled with a score of -1 and Tweets to which a positive value is assigned are labelled with a score of 1. Only Tweets which do not contain words with positive or negative polarities are hence labelled with a score of 0 which represents neutral sentiment.

For a ranking of the dictionaries, we calculate the accuracy of each dictionary-based coding which represents the percentage of Tweets in which the manual and the automated approach are consistent. The accuracy is chosen because alternatives such as F-Measure, precision or recall, which are also used in existing literature, are usually considered for binary classification or focus on the evaluation of one sentiment category only (Kincl, Novák and Pribil, 2013; Lu et al., 2011).

After the ranking of dictionaries based on their performance, we verify if the differences in accuracy are significant. For this purpose, we utilize Mc Nemar's test for significance. The null hypothesis states that two classifiers (in the underlying case dictionary A (DA) and dictionary B (DB)) achieve the same classification accuracy (Lan et al., 2009). Hence, we start with the dictionary which attains the lowest accuracy and test it against the dictionary with the second worse accuracy. Then, we test if the dictionary with the third worse accuracy performs significantly better than the dictionary with the second worse accuracy and so forth. In the last step, the two best performing dictionaries are tested against each other. For each test, a contingency table needs to be constructed which is shown in Table 2 (Dietterich, 1998; Lan et al., 2009).

Table 2: McNemar's Test Contingency Table

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

Based on the contingency table, the statistic χ is calculated which is approximately χ^2 distributed with 1 degree of freedom. The formula in order to calculate χ is depicted in Equation (2) (Dietterich, 1998; Lan et al., 2009). Hereby, the significance level of 0,01 corresponds to a threshold of $\chi = 6.64$ above which the null hypothesis can be disregarded. In this case, the dictionaries are significantly different from each other (Lan et al., 2009).

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

The results can be seen in Table 3. It is evident, that all sentiment dictionaries achieve better classification accuracies than a random distribution which would result in 33,3 % in our three-case scenario. The best dictionary attains an accuracy of 59,19 % which is in accordance with domain independent dictionaries identified in our literature review whereby accuracies up to 87,6 % are attained with domain specific dictionaries (Hogenboom et al., 2014; Molina-González et al., 2013; Thet, Na and Khoo, C. S. G., 2010; Zhang and Peng, 2012). Furthermore, all dictionaries except the two worst differ significantly from each other. Regarding both lexical-semantic extensions of the translated SentiWordNet 3.0, no significant difference could be identified. It is noticeable, that lexical-semantic extensions which are based on SentiWordNet 3.0 represent the four worst dictionaries whereby SentiWordNet 3.0 without lexical semantic-extensions performs even better than SentiWortschatz. This indicates, that lexical-semantic extensions based on translated word lists include

incongruous words whereas solely translated word lists might perform better than language specific dictionaries due to their extent. Still, the three best dictionaries are all based on the German language dictionary SentiWortschatz and extensions of it by SentiWordNet 3.0 or GermaNet. Regarding the best performing dictionary SentiWortschatz and a lexical-semantic extension with GermaNet, the weakened polarity scores with a weight of 0,8 regarding their seed words do not improve the performance. Instead, the extension with GermaNet and polarity scores with 0,8 times the weight of the seed words performs significantly worse in contrast to a dictionary in which lexical-semantic extensions are assigned with equal scores compared to seed words. The weakened weight with 0,8 times the polarity scores of seed words improves the dictionary in one of three cases only. These results challenge findings from Blair-Goldensohn et al. (2008) and highlight a demand for further research concerning the appropriate weighting of lexical-semantic sentiment dictionary extensions.

Table 3: McNemar's Test Results

Dictionary	Accuracy	Significance Levels								
SentiWortschatz + GN	59,19 %									<0,01(***)
SentiWortschatz + GN(0,8)	59,00 %								<0,01(***)	
SentiWortschatz + SentiWordNet 3.0	54,81 %							<0,01(***)		
SentiWordNet	52,55 %						<0,01(***)			
SentiWortschatz	49,97 %					<0,01(***)				
SentiWortschatz + SentiWordNet 3.0 + GN(0,8)	45,25 %				0,08(*)					
SentiWortschatz + SentiWordNet 3.0 + GN	45,12 %			<0,01(***)						
SentiWordNet 3.0 + GN	41,99 %		-							
SentiWordNet 3.0 + GN(0,8)	41,98 %									

4. Conclusion

In order to answer the first research question - *How can sentiment dictionaries be created and evaluated?* – we conducted a systematic literature review in which we identified steps undertaken for sentiment dictionary construction. The result is a process model based on which further sentiment dictionaries can be created. The process can be divided into an initial construction and an extension phase. The initial construction phase encompasses steps for word selection, word polarization and an evaluation of the created dictionaries which can be conducted manually or automatically. Next to that, software or native speaker-based translation tasks support the initial construction at some times. In the extension phase, which is conducted for 37 % of the dictionaries, selection, polarization and evaluation are performed a second time. While the selection of words is based on native speakers, existing sentiment dictionaries or textual corpora in the initial construction phase, lexical-semantic relationships such as synonyms, antonyms etc. are considered during the extension phase. For polarization of the words, native speakers, sentiment dictionaries or textual corpora are utilized in both phases whereas the evaluation is conducted by native speakers or by applying the dictionary to a textual corpus which has been classified beforehand. Then, the accordance of both classifications is assessed.

While domain specific dictionaries usually perform better than general dictionaries, statements regarding the suitability of data sources as well as detailed selection or polarization methods can hardly be made due to inconsistent evaluation figures. Furthermore, non-English dictionaries especially for UGC from social media platforms have not been considered to a great extent yet.

While working out an answer to the second research question - *How do German sentiment dictionaries perform on textual UGC in the domain of corporate credit risk analysis?* – we constructed nine sentiment dictionaries based on existing word lists, which we identified during our literature review. The best dictionary achieves a classification accuracy of 59,19 % which is significantly better than the other dictionaries and in accordance with classification accuracies of domain independent word lists identified in the review. Furthermore, all dictionaries outperform a random classification.

Our results enable researchers to draw on an overarching process model when creating sentiment dictionaries. Hereby, we recommend to unify the publication of evaluation figures and focus on figures such as accuracy

which is an indicator for the overall performance of text classification algorithms. In many approaches, differing figures which focus on only few sentiment categories are published. This hampers an assessment of effects which process characteristics have onto dictionary performance. Practitioners especially in the domain of corporate credit risk analysis can gain insights in how to automatically extract hints for financial stability from textual data by taking our dictionaries into account. Hence, manual coding effort can be reduced and processes for decision support can be improved.

Our results are restricted by the scope of our literature review. Although we aim at identifying all relevant articles, further sources might add steps to the process model. Especially articles which have been published after the begin of our analysis and are therefore not considered in our model, should be integrated in a next iteration. In addition to that, our selection of dictionaries is limited due to language skills and sentiment categories of the manual classified corpus. Dictionaries translated from other languages or a transformation of sentiment categories in order to be comparable with the corpus might attain different classification accuracies. Finally, procedures for authenticity verification of UGC need to be developed since users can anonymously create fake posts which can hamper results of the analysis (Mengelkamp, Hobert and Schumann, 2015).

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.

Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G. A., and Reynar, J. (2008) "Building a Sentiment Summarizer for Local Service Reviews", in *Proceedings of the 2008 Workshop of Natural Language Processing Challenges in the Information Explosion Era*, pp. 1–10.

Bracewell, D. B. (2010) "Semi-Automatic WordNet Based Emotion Dictionary Construction", in *Proceedings of the 9th International Conference on Machine Learning and Applications*, Institute of Electrical and Electronics Engineers (ed.), Washington, DC, pp. 629–634.

Dietterich, T. G. (1998) "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms", *Neural Computation*. Vol. 10, No. 7, pp. 1895–1923.

Ghiassi, M., Skinner, J., and Zimbra, D. (2013) "Twitter brand sentiment analysis - A hybrid sytem using n-gram analysis and dynamic artificial neural network", *Expert Systems with Applications*. Vol. 40, No. 16, pp. 6266–6282.

Hamp, B., and Feldweg, H. (1997) "GermaNet - a Lexical-Semantic Net for German", in *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, P. Vossen, G. Adriaens, N. Calzolari, A. Sanfilippo and Y. Wilks (eds.), Madrid.

Henrich, V., and Hinrichs, E. (2010) "GernEdiT - The GermaNet Editing Tool", in *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, N. Calzolari, K. Chourki, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner and D. Tapias (eds.), Valletta, pp. 2228–2235.

Hogenboom, A., Heerschop, B., Frasincar, F., Kaymak, U., and Jong, F. (2014) "Multi-lingual support for lexicon-based sentiment analysis guided by semantics", *Decision Support Systems*. Vol. 62, pp. 43–53.

Kaplan, A., and Haenlein, M. (2010) "Users of the world unite! The challenges and opportunities of Social Media", *Business Horizons*. Vol. 53, No. 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*. Vol. 33, 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*. Vol. 31, No. 4, pp. 721–735.

Levy, Y., and Ellis, T. (2006) "A Sytems Approach to Conduct an Effective Literature Review in Support of Information Systems Research", *Information Science Journal*. Vol. 9, No. 1.

Lu, Y., Castellanos, M., Dayal, U., and Zhai, C. (2011) "Automatic construction of a context-aware sentiment lexicon", in *Proceedings of the 20th International Conference on World wide web*, S. Sadagopan, K. Ramamritham, A. Kumar, M. P. Ravindra, E. Bertino and R. Kumar (eds.), Hyderabad, pp. 347–356.

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*. Vol. 26, No. 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*. Vol. 40, pp. 7250–7257.

Nassirtoussi, A. K., Wah, T. X., Aghabozorgi, S. R., and Lings, D. N. C. (2014) "Text mining for market prdiction: A systematic review", *Expert Systems with Applications*. Vol. 41, No. 16.

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

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*. Vol. 6, No. 2, pp. 89–96.

Thet, T. T., Na, J.-C., and Khoo, C. S. G. (2010) "Aspect-based sentiment analysis of movie reviews on discussion boards", *Journal of Information Science*. Vol. 36, No. 6, pp. 2544–2558.

Waltinger, U. (2010) "A Lexical Resource for German Sentiment Analysis", in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, N. Calzolari, K. Chourkri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner and D. Tapia (eds.), Valetta.

Waltinger, U. (2012). *GermanPolarityClues - A Lexical Resource for German Sentiment Analysis (Onlinedocumentation)*, http://www.ulliwaltinger.de/sentiment/. Accessed 2 December 2015.

Webster, J., and Watson, R. t. (2002) "Analyzing the Past to Prepare for the Future: Writing a Literature Review", *MIS Quarterly*. Vol. 26, No. 2, pp. xiii–xxiii.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005) "Recognizing contextual polarity in phrase-level sentiment analysis", in *HLT 05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, R. J. Mooney (ed.), Vancouver, pp. 347–354.

Zhang, J., and Peng, Q. (2012) "Constructing Chinese domain lexicon with improved entropy formula for sentiment analysis", in *Proceedings of the International Conference on Information and Automation*, Institute of Electrical and Electronics Engineers (ed.), Shenyang, pp. 850–855.