Influence of neutral word removal on sentiment analysis

Daphne Rietvelt
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
d.c.j.c.rietvelt@student.utwente.nl

ABSTRACT

Nowadays, lots of information is shared on the Internet, including opinions. In order to deal with these opinions, computers perform sentiment analysis, which provides insight into the sentiment of a piece of text. Currently, not a lot of research is done for Dutch sentiment analysis. Preliminary experiments found that a larger lexicon enhances sentiment analysis, but this increase was too small compared to previous literature. A possible cause for this is the large amount of neutral words the larger lexicon contained after extension. This paper will investigate the influence of neutral word removal on the performance of sentiment analysis. Two experiments were conducted, one on an unbalanced dataset and one on a balanced dataset. Neutral words were gradually removed from the extended lexicon and the performance was measured. The two experiments both show that neutral word removal enhances sentiment analysis, but the differences are small. Furthermore, the experiment on the balanced dataset shows that a larger lexicon does not enhance sentiment analysis. Due to small enhancements and the opposite results compared to literature, no conclusion is drawn.

Keywords

Sentiment analysis, Lexicon-based, Neutral word removal

1. INTRODUCTION

Over the last decade the World Wide Web has grown enormously. Lots of information is shared among humans on the Internet. This also includes *opinions*: subjective expressions which contain sentiment, appraisals or feelings toward entities, events and their properties [11]. These opinions are used in the decision-making process of a person [18]. Before the rise of the World Wide Web, opinions were acquired by asking relatives and friends. For example, when buying a freezer, relatives and friends were asked for their advice. Based on this advice and ones knowledge about freezers a decision was made. Nowadays, these opinions (advice) can be found on the Internet in the form of reviews. Since these digital opinions (reviews) are so important in the decision-making process, they are a valuable source for companies to gain understanding of the decision-making process of their potential customers.

31th Twente Student Conference on IT 5th of July, 2019, Enschede, The Netherlands.

Reading and evaluating these opinions by hand is time consuming, so computers can be of use. "Sentiment analysis deals with the computational treatment of opinion, sentiment, and subjectivity in text" [19]. In other words, computers are able to, to a certain extent, extract people's opinions, sentiments and emotions from text. Currently, there are two main approaches in sentiment analysis. The first approach is based on lexicons, dictionaries of words annotated with a positive or negative sentiment score (polarity) [21]. This lexicon is used to assess sentiment of a phrase or paragraphs by combining the polarities of the individual words in a text to one final sentiment score for the whole phrase or paragraph. In the second approach sentiment of a text is gathered with the use of Machine Learning methods. Examples of such methods are Naive Bayes, Support Vector Machines, Random Forest and Neural Networks [16, 12, 7]. This paper focuses on the first approach.

Most research on this approach is aimed at constructing sentiment analysis tools and lexicons. This often includes comparison with existing sentiment analysis tools or lexicons. A particularly large part of research is focused on the English and Arabic language [4, 21]. Sentiment analysis in other languages is growing, including Romanian, Chinese, German, Spanish and Japanese [17]. However, sentiment analysis for the Dutch language is not well developed yet. Research on lexicons for the Dutch language is mainly focused on the semantic of words instead of the sentiment of words [22, 8]. This means that there are not many reliable Dutch lexicons for sentiment analysis. Therefore an upcoming trend is the creation of a multilingual lexicon, which can be used for multiple languages at the same time [2], for example, a lexicon that can be used both for the English and the Dutch language. The creation of a multilingual lexicons is promising, but still lots of research needs to be done on developing these multilingual sentiment analysis tools and lexicons [9, 14]

Due to the underdeveloped sentiment analysis for Dutch, the Human Media Interaction (HMI) department of the University of Twente started a project to ameliorate sentiment analysis for the Dutch language. They make use of the only Dutch lexicon-based sentiment analysis tool that can be freely used and is readily available, called Pattern [20]. The lexicon of Pattern consists of roughly 4000 words and is based on research done on adjectives by Smedt and Daelemans [3]. The 4000 words of Pattern's lexicon are quite limited. Research done by Gatti, Guerini and Turchi [5] observed that the size of the lexicon positively influences the performance of sentiment analysis. More evidence for this hypothesis was gathered by research on Arabic sentiment analysis [1]. Since larger lexicons seem to enhance the performance of sentiment analysis, Pat-

tern's lexicon was extended in a preliminary experiment by adding a lexicon from Moors et al. (Moors' lexicon) [13]. This led to an extended lexicon (the Extended lexicon) of approximately 8200 words. The Extended lexicon was tested and showed indeed a slight increased performance. However, this was not in line with the expected increase.

1.1 Aim of research

A possible explanation for the small increase may have something to do with the goal of Moors' lexicon. The main aim of this lexicon was to form a "sample" of the Dutch language. When sampling a language, the aim is to find a set of words that represents a language. This means that Moors et al. were interested in words with all kinds of polarities, including neutral words. By adding Moors' lexicon to Pattern's lexicon, the Extended lexicon contains relatively more neutral words compared to Pattern's lexicon. However, neutral words are not as important for sentiment analysis as emotion words, words which indicate an emotion (strong polarity). Similarly, stop words (such as: a, the, and) are not as important as words which give meaning to a sentence for natural language processing. Stop word removal before processing even enhances the performance of natural language processing [5]. This might also be the case with the neutral words in sentiment analysis.

To our knowledge, no research is available which examines the influence of neutral words in sentiment analysis. However, the human brain processes emotion words (positive/negative) faster than neutral words [10], and while not directly related to sentiment analysis, it could be a sign of the higher importance of strongly emotional words. This paper examines the influence of neutral word removal on the performance of lexicon-based analysis. Since there is no research related to neutral words and sentiment analysis, it is not clear what can be considered as a neutral word. For now, the informal definition of neutral words is as follows: words that have not a strong emotion associated with it.

The research discussed in this paper is structured as follows: information on the different lexicons is available in section 2. The methodology of the research can be found in section 3. Section 4 contains a description of the results. The conclusions and future work are described in section 5.

2. LEXICONS

The words in an affective lexicon contain a sentiment score (polarity). For Pattern this polarity consists of a number between -1 and +1, rounded to one decimal, and it describes the positive or negative sentiment of a word. An example would be the Dutch word 'oncomfortabel' (meaning: 'uncomfortable'), which has a strong negative sentiment associated with it, which translates to a polarity of -0.8. For this research two lexicons were used: Pattern's lexicon and Moors' lexicon. Together they form the Extended lexicon. These lexicons are described below.

2.1 Pattern's lexicon

Pattern's lexicon is the lexicon that is used for the sentiment analysis tool called Pattern. The lexicon consists of 3918 words. The mean of this lexicon is -0.021. The mean is thus close to the polarity of 0.0. The standard deviation is 0.425. This means that most data points are located in the central part of the range (around polarity of 0.0). The frequency histogram, figure 1, shows that the lexicon is fairly symmetrical. These statistics give an indication that the data in the lexicon is close to a normal distribution.

Frequency polarity in lexicon		
Polarity	Pattern's lexicon	Moors' lexicon
-1.0 to -0.7	220 (5.6%)	152 (2.7%)
-0.6 to -0.4	828 (21.1%)	768 (17.9%)
-0.3 to -0.1	657 (16.8%)	835 (19.4%)
0.0	659 (16.8%)	589 (13.7%)
+0.1 to +0.3	678 (17.3%)	1320 (30.7%)
+0.4 to +0.6	632 (16.1%)	594 (13.8%)
+0.7 to +1.0	244 (6.2%)	93 (2.1%)

Table 1. Frequency polarity in lexicon

Some words (461 words) occur multiple times in Pattern's lexicon, due to the different senses of a word. These words are called polysemic words. The different senses of a polysemic word each have a polarity associated with it. This polarity can be different among the senses. For example, the word 'bitter'. This word has at least three different meanings: 1) the taste of bitter 2) unpleasant situation and 3) someone who is embittered. These different meanings all have a different polarity associated with them: 1) -0.1 2) -0.6 and 3) -0.8. Currently, Pattern does not make use of the context and is therefore not able to decide which sense of a polysemic word needs to be used. In order to deal with the multiple senses of a word, Pattern simply takes the average of the polarities as the polarity for predicting the sentiment of a text. For example, the Dutch word 'goedkoop' (meaning: 'cheap') is a polysemic word and the two senses both have the polarity of -0.6. What Pattern does is (-0.6 + -0.6)/2 = -0.6. Another example is the word 'angstig' (meaning: 'anxious'). This word is also a polysemic word, but the senses have in this case a different polarity: -0.2 and 0.0. In all cases, regardless of the polarities and context, Pattern computes the average, so (-0.2 + 0.0)/2 = -0.1. This final polarity is the polarity for that particular word. The polarities of all the words in a text together will result in the final sentiment score of a text.

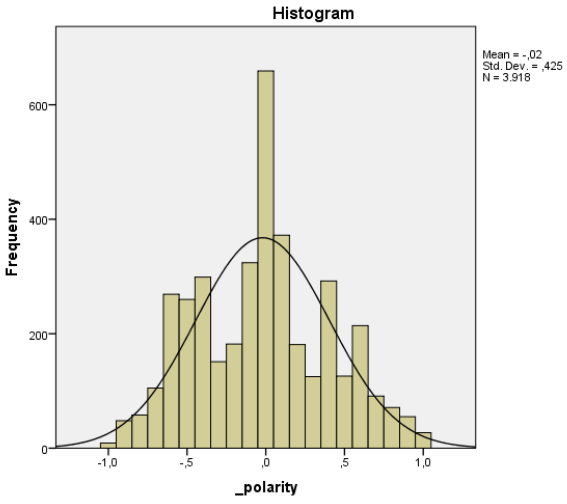

Figure 1. Pattern's lexicon polarity frequency

2.2 Moors' lexicon

Moors' lexicon is constituted to be a sample of the Dutch language. This lexicon consists of 4299 words. In this lexicon the polarity is expressed in a 7-point Likert scale, a value of 7 indicates a strong positive sentiment and 1 a strong negative sentiment. In order for Pattern to deal with this polarity it needs to be rescaled into a range of -1 to +1. This is done using the following formula:

$$P = (n - 4)/3 \quad (1)$$

The P is the polarity, number between -1 and +1. The n is the sentiment score expressed in Moors' lexicon, which is a number between 1 and 7.

After the polarity is rescaled, some statistics are drawn from this lexicon. The mean of this lexicon is -0.021 and the standard deviation is 0.353. This means that on average the polarity of the words are 0.353 points away from the mean. The frequency histogram, figure 2, shows that the lexicon is fairly symmetrical. These statistics indicate that the data in the lexicon is close to a normal distribution. Lastly, Moors' lexicon is not sense-disambiguated.

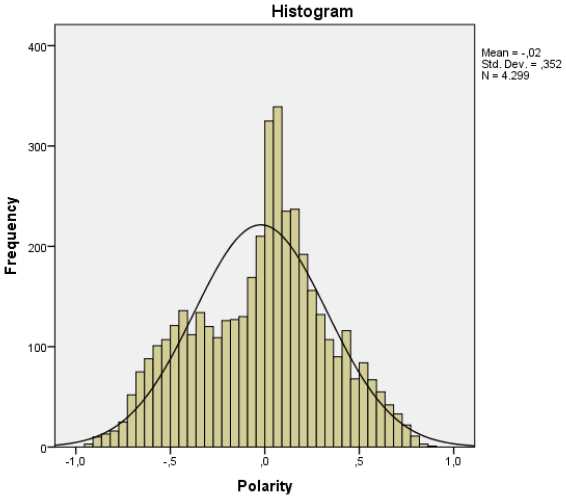

Figure 2. Moors' lexicon polarity frequency

2.3 Pattern's vs. Moors' lexicon

Moors' lexicon is slightly larger than Pattern's lexicon. Pattern's lexicon consists of 3,304 unique words and Moors' lexicon of 4,299 unique words. The mean values of the lexicons are both -0.021, but the standard deviation of Moors' lexicon (0,353) is smaller than the standard deviation of Pattern's lexicon (0.425). This shows that the polarities of Moors' lexicon are located closer to the mean compared to Pattern's lexicon, which indicates that Moors' lexicon contains more words that are around the polarity of 0.0 (neutral words). Another statistic that shows this is the frequency table (table 1). Pattern's lexicon contains 1,994 words (50.9%) which have a polarity between -0.3 to +0.3 and Moors' lexicon contains relatively more words in this range, namely 2,744 words (63.8%) . This shows Moors' lexicon contains relatively more neutral words compared to Pattern's lexicon.

2.4 Extended lexicon

The Extended lexicon is the combination of Pattern's lexicon and Moors' lexicon. This Extended lexicon has in total 8,217 words. However, the polysemic words from Pattern's lexicon need to be subtracted from this and that results into an Extended Lexicon of 7,603 words. Moreover, there is a word overlap between Pattern's lexicon and Moors' lexicon, which can be found in table 2. The overlap between those lexicons consists of 726 words, hence the Extended lexicon finally consists of 7,603-726 = 6,877 unique words. In the overlap between the two lexicons (726 words), some words have the same polarity in both lexicons (121 words) and some words have a different polarity in the two lexicons (605 words). Here again, when a word occurs multiple times, Pattern just simply takes the average of all polarities as the final polarity for that particular word.

Overlap Pattern's lexicon and Moors' lexicon	
Words total overlap	726
Words overlap: same polarity	121
Words overlap: different polarity	605

Table 2. Overlap Pattern's lexicon and Moors' lexicon

3. METHODOLOGY

In order to examine the influence of neutral word removal on the performance of sentiment analysis, neutral words were gradually removed and the performance of these lexicons were tested on 70,000 Dutch book reviews. As noted before, it is not clear from the literature what can be considered as a neutral word. Therefore, neutral words were removed using multiple thresholds. The following thresholds were tested: 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85 and 0.95. For example in the case of 0.15 threshold, all the words with a polarity between -0.15 and 0.15 were filtered from Moors' lexicon. After filtering Moors' lexicon, the two lexicons were put together and they formed a Threshold lexicon with a threshold of 0.15. As the example shows in this study we have chosen to only filter words from Moors' lexicon. We could also have chosen to filter from both lexicons, but then we would need more conditions in order to control for the interaction effect.

The Threshold lexicons were tested on 70,000 Dutch book reviews. In order to keep the circumstances equal among measurements, all Threshold lexicons were tested on the same reviews. Dutch book reviews are chosen in this study because of the reported 82% accuracy Pattern has on book reviews [6]. After performing sentiment analysis on these reviews with the use of Pattern, each review was given a sentiment score (polarity). In order to determine the performance of sentiment analysis, the mean absolute error (MAE) was computed. This error shows the average difference between the predicted sentiment scores from Pattern and the actual sentiment score which come with the Dutch book reviews. The lower the difference between the predicted and the actual value, the higher the performance of sentiment analysis. In this study, the measure MAE is chosen instead of accuracy, because the MAE also takes into account how far the predicted value lies from the actual value. As an example, let's assume that the actual sentiment score of a review is 0.8 and the predicted value is 0.4 and we have another tool who predicts the review with a score of 0.6. When the accuracy measure is used, it would only state that these predictions are wrong and thus does not include the fact that the prediction of 0.4 lies fur-

ther away from the actual value (0.8) than the prediction of 0.6. In this case, the tools will have the same accuracy score (performance), but actually the tool which predicted 0.6 performed better than the tool which predicted 0.4. By using the MAE, this difference in performance is included in the measure.

3.1 Dutch Book Reviews

The Dutch Book Reviews were taken from bol.com. These reviews consists of a title, body (text) and a star rating of 1-5. For this research solely the text is used in the prediction of Pattern. The star ratings are taken as the actual sentiment value of the text. A star rating of 5 indicates that this review has a strong positive attitude towards a book and a star rating of 1 a strong negative attitude towards a book. To enable comparison with the predicted polarity, the star rating were transformed to the scale of -1 and +1. Equation 2 was used for this transformation.

$$P = (s - 1)/2 - 1 \quad (2)$$

The parameters of the equation are as follows: P is the polarity, a number between -1 and +1 and s is the star rating, a number between 1 and 5. For example, a book review with a star rating of 5 will give a polarity of +1 and a star rating of 3 will give a polarity of 0.

4. RESULTS

The results show that when Moors' lexicon is added to Pattern's lexicon, the performance of the Extended lexicon without neutral word removal (Threshold lexicon 0.0) is lower than the performance of Pattern's lexicon as shown in table 3. The MAE for the Extended lexicon (Threshold lexicon 0.0) is 0.537 and the MAE of Pattern's lexicon 0.510, see table 3. This may seem counter intuitive, because a larger lexicon will give a higher performance according to previous findings in the literature [1, 5]. However, as discussed before, a possible cause of this lower performance is the high percentage of neutral words in this Threshold lexicon. Therefore it is in accordance with the expectations.

Performance sentiment analysis per lexicon	
Lexicon	Performance SA (MAE)
Pattern's lexicon	0.501
Moors' lexicon	0.534
Extended lexicon (no threshold)	0.537

Table 3. Performance sentiment analysis of the basic lexicons

When neutral words are filtered, the MAE decreases. This means that the predicted values come closer to the actual values and indicate a higher performance when more neutral words are filtered. This trend is in line with the expectation, that neutral word removal will enhance the performance of sentiment analysis.

The red line in figure 3 visualizes the performance (MAE) of Pattern's lexicon. All the MAE scores of Threshold lexicons below the red line have a higher performance than Pattern's lexicon and all the Threshold lexicons above this line have a lower performance compared to Pattern's lexicon. This shows that Threshold lexicons with a threshold lower than 0.35 have a lower performance than Pattern's lexicon and Threshold lexicons with a threshold of 0.35 or

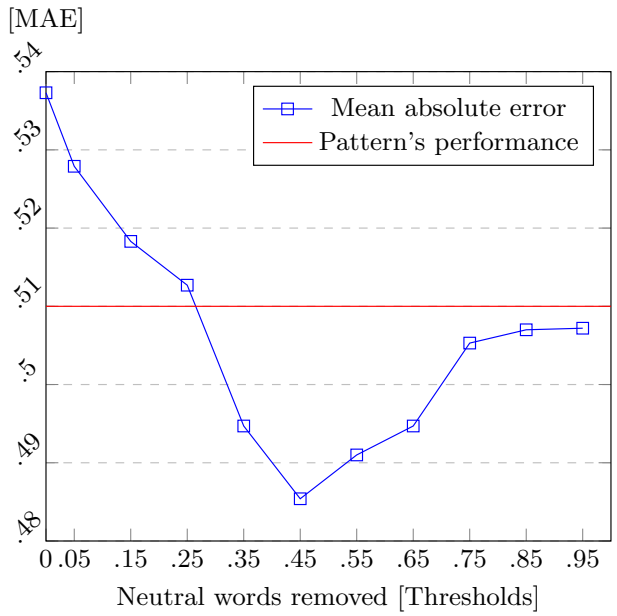

Figure 3. Influence of neutral word removal on the performance of sentiment analysis

higher have a higher performance and indicates thus an improved lexicon compared to Pattern's lexicon.

The lowest MAE (highest performance) is achieved with the 0.45 Threshold lexicon. The difference between this lexicon and Pattern's lexicon is an error of 0.510 - 0.485 = 0.025. This improvement is small, but significant ($p<0.001$), which was tested with the Wilcoxon Signed Rank Test. If words are filtered with a threshold higher than 0.45, the MAE increases again, which indicates a lower performance of sentiment analysis compared to the 0.45 Threshold lexicon. The possible cause of this increase is that too many crucial words, words which are needed to accurately determine the sentiment of a text are filtered. As shown in table 4, when going from the Threshold lexicon with 0.45 to Threshold lexicon 0.55, 205 positive and 237 negative words with a polarity between 0.4 and 0.6 and -0.4 and -0.6 are removed.

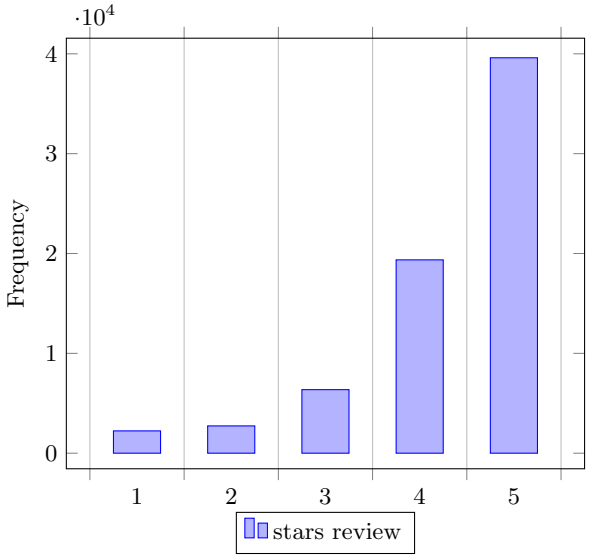

Figure 4. Frequency of stars in reviews

Polarity	Number of words in each lexicon											
	Threshold lexicons											Pattern's
	0.00	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	lexicon
-1.0 to -0.7	414	414	414	414	414	414	414	414	287	248	223	220
-0.6 to -0.4	1500	1500	1500	1500	1500	1256	1019	828	828	828	828	828
-0.3 to -0.1	1411	1411	1142	913	657	657	657	657	657	657	657	657
0.0	1051	659	659	659	659	659	659	659	659	659	659	659
+0.1 to +0.3	2146	2146	1491	1010	678	678	678	678	678	678	678	678
+0.4 to +0.6	1231	1231	1231	1231	1231	1003	798	632	632	632	632	632
+0.7 to +1.0	464	464	464	464	464	464	464	464	344	277	247	244
Total words	8217	7825	6901	6191	5603	5131	4689	4332	4085	3979	3924	3918

Table 4. Frequency polarity in Threshold lexicons

In this research the lexicons are tested on 70 thousand Dutch book reviews. As shown in figure 4 the amount of book reviews are not equally distributed among the stars, which results in an unbalanced set of data. There are way more book reviews which have 4 or 5 stars (polarity 0.5 or 1.0) than 1 or 2 stars (polarity -1.0 or -0.5). This unbalanced set gives a distorted view on the performance of each lexicon. For example, if a lexicon contains more positive words the performance of that lexicon will get a higher performance than a lexicon which contains more negative words, even though the same threshold of neutral words are filtered. To control for this variable, the Dutch book reviews were randomly sub-sampled to obtain a balanced dataset. The sub-sample consists of 11,180 reviews, with each star category containing 2,236 reviews. The Threshold lexicons were tested on the sub-sample and the results of this experiment can be found in figure 5.

When the unbalanced dataset is tested, the Threshold lexicon 0.0 (Extended lexicon without neutral word removal) has a lower performance (MAE of 0.565), compared to the performance of Pattern's lexicon (MAE of 0.524). The overall trend in this graph, see figure 4, is that neutral word removal will enhance sentiment analysis (SA) performance. All the improvements of the lexicons, as shown in figure 4, are significant ($p<0.001$), except for the Threshold of 0.95. This all seems to be in accordance with the previous results.

However, there are four main differences. The first one is that the overall performance of all lexicons is lower on the 11,180 book reviews compared to the 70 thousand book reviews, average MAE of 0.543 against an average MAE of 0.508.

Furthermore, all Threshold lexicons are above the red line. This means that all Threshold lexicons have a lower performance than Pattern's lexicon. This result is not in agreement with previous research that larger lexicons enhance the SA performance [1]. Moreover, this result also contradicts the previous results when testing on the unbalanced dataset. A possible explanation for this difference is not found.

The third main difference is that when the threshold lexicons are tested on the unbalanced dataset the performance decreases when the threshold is higher than 0.45, and this decrease cannot be found in the experiment done on the balanced dataset. There is not yet a possible cause we can devote to this phenomena.

The last main difference in results, is the outlier of Threshold lexicon 0.35. The trend seen in this graph that is that the MAE will go down, when more words are filtered from the lexicon. However, this not the case when going from Threshold lexicon 0.25 to Threshold lexicon 0.35, as figure 5, the MAE increases. A possible cause of this is that many words are filtered that are crucial for accurately predicting the sentiment scores.

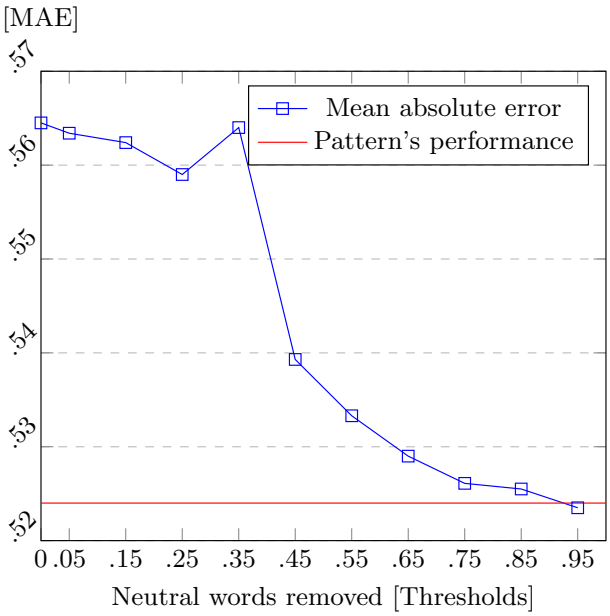

Figure 5. Influence of neutral word removal on the performance of sentiment analysis

5. CONCLUSION

In this research we looked at the influence of neutral word removal on sentiment analysis. Two similar experiments were performed, one on an unbalanced dataset and the other on a balanced sub-sample of the unbalanced dataset. From the results we cannot conclude that neutral word removal enhances the overall performance of sentiment analysis, due to the small enhancements in performance and the inexplicable differences between the results of the two experiments. The inexplicable differences referred to are the differences in trend with a threshold higher than 0.45 and the difference in improvement with a larger or a smaller lexicon. What we can conclude here is that Moors lexicon is not suited to perform sentiment analysis. The experiment on the balanced dataset showed that extending Pattern's lexicon with Moors' lexicon does not improve the performance of sentiment analysis.

This research has some limitations. One of the limitations

concerns the Dutch book review dataset. As explained earlier the star ratings are considered to be the indicator for the sentiment associated with the reviews. However, this does not necessarily have to be the case. This assumption was made in order to be able to conduct this research, because no other Dutch dataset was available. In order to provide more accurate research a new dataset needs to be created. A possible way to construct this is based on the concept of crowdsourcing. An audience is asked to rate text by assigning a sentiment score to it. If a few hundred pieces of texts are rated by a few hundred of people, the pieces of text together can then be used as a sentiment dataset. A quick way to gather all these ratings is by using the platform Amazon Mechanical Turk, similarly to what was done for the creation of a sentiment analysis dataset with Twitter data [15].

Another limitation is the sentiment analysis tool used. Pattern does not take of the context in which the words are used into account and therefore has difficulty to determine which instance of a word (including polarity) to choose from the lexicon. Hence, Pattern takes the average of all the instances, which may cause a large error in the predictions. Another point to make here is that the neutral word removal is performed on Moors' lexicon, thus Pattern's lexicon still contains neutral words. The words are not filtered from both Moors' lexicon and Pattern's lexicon in order to control for the interaction effect. Future research may filter Pattern's lexicon as well, but keeping in mind to evaluate the interaction effect of filtering both lexicons.

Other future work concerns the inexplicable differences in results between the unbalanced dataset and the balanced dataset. One of those differences is the difference in SA performance with a Threshold higher than 0.45. In the unbalanced dataset the SA performance, drop, whereas in the results of the balanced dataset the SA performance rises. The possible cause for this difference is not available. Another point of further work concerns the inexplicable difference on the overall SA performance for Threshold lexicons compared to Pattern's lexicon. The unbalanced dataset shows multiple Threshold lexicons which have a significant higher SA performance than Pattern's lexicon, while the balanced dataset results show that all Threshold lexicons have a lower SA performance. Possibly, more insight into the sentiment analysis process can help to identify possible causes for these phenomena. Such insight could be obtained by measuring the amount of words recognized by the Threshold lexicons (coverage).

6. REFERENCES

[1] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub. Arabic sentiment analysis: Lexicon-based and corpus-based. In *Proceedings of the 2013 IEEE Jordan Conference on Applied Electrical Engineering And Computing Technologies (AEECT)*, pages 1–6. IEEE, 2013.

[2] D. Bal, M. Bal, A. Van Bunningen, A. Hogenboom, F. Hogenboom, and F. Frasincar. Sentiment analysis with a multilingual pipeline. In *Proceedings of the International Conference on Web Information Systems Engineering*, pages 129–142. WISE, 2011.

[3] T. De Smedt and W. Daelemans. "Vreselijk mooi!" (terribly beautiful): a subjectivity lexicon for Dutch adjectives. In *Proceedings of the International Conference on Language Resources and Evaluation*, pages 3568–3572. LREC, 2012.

[4] N. El-Makky, K. Nagi, A. El-Ebshihy, E. Apady, O. Hafez, S. Mostafa, and S. Ibrahim. Sentiment analysis of colloquial Arabic tweets. In *Proceedings of the ASE International Conference on Social Informatics*, pages 1–9. SocInfo, 2014.

[5] L. Gatti, M. Guerini, and M. Turchi. SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, 7(4):409–421, 2016.

[6] J. Geertzen. software & demos: Brill. *URL http://cosmion. net/jeroen/software/brill/pos*, 2010. Accessed: 2019-05-10.

[7] A. Hasan, S. Moin, A. Karim, and S. Shamshirband. Machine learning-based sentiment analysis for Twitter accounts. *Mathematical and Computational Applications*, 23(1):11, 2018.

[8] A. Hogenboom, B. Heerschop, F. Frasincar, U. Kaymak, and F. de Jong. Multi-lingual support for lexicon-based sentiment analysis guided by semantics. *Decision Support Systems*, 62:43–53, 2014.

[9] M. Kaity and V. Balakrishnan. Building multilingual sentiment lexicons based on unlabelled corpus. In *Data Science Research Symposium 2018*, page 10, 2018.

[10] S.-T. Kousta, D. P. Vinson, and G. Vigliocco. Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112(3):473–481, 2009.

[11] B. Liu. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca*, 2010.

[12] O. R. Llombart. Using machine learning techniques for sentiment analysis. *published in June of*, 2017.

[13] A. Moors, J. De Houwer, D. Hermans, S. Wanmaker, K. Van Schie, A.-L. Van Harmelen, M. De Schryver, J. De Winne, and M. Brysbaert. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior research methods*, 45(1):169–177, 2013.

[14] B. Naderalvojoud, B. Qasemizadeh, L. Kallmeyer, and E. A. Sezer. A cross-lingual approach for building multilingual sentiment lexicons. In *International Conference on Text, Speech, and Dialogue*, pages 259–266. Springer, 2018.

[15] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov. Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1–18, 2016.

[16] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, pages 79–86. ACL, 2002.

[17] V. Perez-Rosas, C. Banea, and R. Mihalcea. Learning sentiment lexicons in Spanish. In *Proceedings of the International Conference on Language Resources and Evaluation*, volume 12, page 73. LREC, 2012.

[18] J. Prager et al. Open-domain question–answering. *Foundations and Trends® in Information Retrieval*, 1(2):91–231, 2007.

[19] S. Schrauwen. Machine learning approaches to sentiment analysis using the Dutch Netlog corpus. *Computational Linguistics and Psycholinguistics Research Center*, 2010.

[20] T. D. Smedt and W. Daelemans. Pattern for

Python. *Journal of Machine Learning Research*, 13(Jun):2063–2067, 2012.

[21] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[22] P. Vossen, I. Maks, R. Segers, and H. VanderVliet. Integrating lexical units, synsets and ontology in the Cornetto database. In *Proceedings of the International Conference on Language Resources and Evaluation*. LREC, 2008.