GerVADER - A German adaptation of the VADER sentiment analysis tool for social media texts

Karsten Michael Tymann, Matthias Lutz, Patrick Palsbröker, and Carsten Gips

FH Bielefeld University of Applied Sciences, Minden, Germany
`ktymann@fh-bielefeld.de`, `matthias.lutz@fh-bielefeld.de`,
`patrick.palsbroeker@fh-bielefeld.de`, `carsten.gips@fh-bielefeld.de`
`https://www.fh-bielefeld.de`

Abstract. For the English language sentiment analysis tools are fairly popular. One is called VADER [1] which offers a rather simple process for sentiment classification. Due to its lexicon-based approach with a design focus on social media texts, no additional training data is required. In this paper the process of creating VADER is applied to build a German adaptation which is called GerVADER. The paper will present the concept of VADER and how a German version can be built within reasonable time. GerVADER uses SentiWS as a starting point for the lexicon, combines it with language independent parts of the VADER lexicon and copies the process of having users rate the words intensity and polarity. The next step is comprised of comparing the algorithmically changes due to the natural differences in language between German and English. Then GerVADER is compared to the results of the SB10k [2] corpus classification which contains more than 9000 human labeled tweets. Finally GerVADER is tested with parts of the SCARE [3] dataset which contains reviews for mobile apps. The results show that GerVADER lacks some additional work to increase its classification accuracy, but it promises better results considering how well the original performed.

Keywords: VADER· German sentiment analysis· SB10k · SCARE

1 Introduction

Sentiment analysis is often based on machine learning which requires lots of data and sometimes even additional human work, e.g. for labeling the data beforehand. In the German language collecting reasonable amounts of data for machine learning is quite difficult, since not many work has been done in the field yet. This was the motivation of this work to build an own corpus and label a reasonable amount of it for training purposes. Especially for the domain of social media and micro blogging the internet is lacking up-to-date German corpora to

bootstrap one's sentiment analysis tool. While there exist corpora like SB10k or DAI [6], those are not available to the public and can be difficult to obtain in its entirety, given how companies like Twitter handle their data policy. Even if one obtains the corpora it still requires lots of research and additional work to get a running sentiment analysis tool for the German language.

Another crucial factor for a sentiment analysis tool can be lexicons. For the German language there are multiple lexicons free to use. SentiWS [4] for example is a German lexicon with polarity and intensity. For every word there are multiple grammatical forms, e.g. the plural form of the word.

GermanPolarityClues (GPC) [5] is another resource for a German dictionary, however the project seems abandoned and superseded by SentiWS. Nevertheless, it is stated that in tests GPC reached a F1 score of up to 0.88.

Both of the mentioned lexicons are not adapted to the social media domain and lack multiple linguistic features that are common in such a domain.

This paper will show how GerVADER builds upon the good results of VADER with its own lexicon. VADER reached great classification accuracy for microblogging platforms (up to $F1 = 0.96$) and was able to score better results than human raters in some cases. VADER is free to use, requires no knowledge of machine learning and can be easily executed and expanded with Python or its multiple adaptations in other programming languages.

It will be shown how GerVADERs lexicon was built with the SentiWS lexicon and parts of VADERs lexicon. In the next step the lexicon has been rated by a crowd and then cleaned of ambiguous data. Afterwards the grammatical and lexical heuristics used in VADER (e.g. negation words) have been manually adjusted to the German language. Then GerVADER has been tested on parts of the SB10k corpus as well as on a subset of app reviews of the SCARE corpus. GerVADER scores mediocre ratings (F1=0.36-0.70) depending on the test corpus and allows future tools to compete against it in the mentioned datasets. The scores hint that GerVADER has still unexploited potential and that the German language might need additional grammatical rules for an improved VADER adaptation. Especially the correct classificaton of negative sentiments is lacking as well as further testing of GerVADER with different corpora and its usefulness in other domains.

GerVADER was developed as part of a student project and will be free to download and to use like VADER [12].

2 Background

In this section the original VADER tool for the English language will be described. Furthermore, the SentiWS lexicon, which was used as a basis for Ger-VADERs lexicon, as well as the corpora SB10k and SCARE, which were used for evaluating GerVADER, will be introduced.

2.1 VADER

VADER is short for "Valence Aware Dictionary and sEntiment Reasoner" and is available under the MIT License. The tool was published in 2014 and is especially focused on social media texts. It uses a lexicon driven approach as well as additional heuristics for rating the input. Since VADER is not a machine learning approach it offers consistent ratings and requires no training data. VADER achieved some remarkable scores for multiple domains such as tweets, movie or product reviews. The development of VADER can be split into seven steps:

Gather lexical features of established sentiment lexicons: The creators of VADER first researched for existing sentiment lexicons like LIWC, ANEW and GI. They took parts (words) of it and integrated them into their own VADER lexicon.

Gather lexical features characteristic for microblogging domains: Texts on social media and other microblogging platforms have their own unique characteristics. The creators gathered emoticons, domain specific words and other abbreviations from these platforms and integrated them into the lexicon.

Rate lexical feature candidates: In this step the creators gathered a crowd in order to rate the words individually in their intensity and polarity. Every rater received batches of 25 words, in which five words have been intentionally integrated that function as a gold-standard validation. If a user rates three or more of the five gold-standard validation words wrong, the whole batch will be discarded. The gold-standard words have been manually set and do not seem to be available for download. In [1] it is stated that for good results the participants have received financial compensation. Also, the raters have been carefully selected. Every rater had to pass a reading comprehension test and took part in an online sentiment training. At the end of the rating process VADER had more than 9000 words rated with 10 individual ratings each in the range of very negative (-4) to neutral (0) to positive (+4).

Filtering: In this step the lexicon has been cleaned from inconclusive words. These words were either rated neutral in total by the crowd, or the crowd was divided over the polarity and intensity of the word. This means that the standard deviation of the word was 2.5 or higher and thus resulting in a value that cannot be trusted for a sentiment classification. After the filtering the lexicon contained more than 7.500 words.

Building human heuristics: VADER contains five heuristics that can shift or boost the sentiment of a sentence. These heuristics include punctuation marks, capitalization (words in all caps), booster words (negative and positive, e.g. words like "amazingly"), contrastive conjunctions and words that negate a sentence (e.g. "not", "won't"). When a sentence is being rated these keywords are identified and can shift or impact the rating.

Evaluate heuristics: In order to evaluate how much the gathered heuristics can influence a sentiment, the authors have conducted a controlled experiment with 30 tweets that have been manually modified into different versions which include the features explained in subsection 2.1. Those tweets have been mixed into other tweets and were again rated by a crowd. As a result, the authors

were able to analyze how much a lexical or grammatical feature can impact the sentiment of a tweet. The findings were integrated into VADERs heuristics.

Evaluation and results: In the last step VADER was tested in four different domains against other established lexicon-based approaches. The domains included social media text, movie and product reviews as well as newspaper articles. The results have shown that VADER outperforms every other lexicon in every domain and is even able to outperform human raters in the domain of social media texts. Against machine learning models (NB, ME, SVM) VADER was ranked first in three of the four domains. Only in the movie domain the Naive Bayes and Maximum Entropy methods (both trained on movie corpora) were able to reach better results than VADER ($F1 = 0.75$ vs. $F1 = 0.61$). In summary VADER achieves good scores in the social media domain. Another benefit of the approach is that VADER does not require any training and rates consistently. On the downside VADER does not detect irony as well as longer more complex sentences might be rated wrongly, since the heuristics only apply to small ranges of words. Additionally, VADER does not detect phrases but will rate every single word individually which can lead to wrong conclusions.

2.2 SentiWS

SentiWS is a German lexicon for sentiment analysis. It offers 1.644 positive and 1.827 negative words with a polarity and intensity range from -1 to 1. All words are given in their base form and additionally in variations like plural form for nouns or tenses for verbs (see Table 1). Therefore, while only the basic form is rated one can easily transfer the rating to its grammatical variations. SentiWS has been lastly updated at the end of 2018 and is therefore an up-to-date resource for a German sentiment lexicon.

Table 1. SentiWS extract: Positive words

Word	POS	Polarity	Forms
Anspruch	NN	0.0040	Anspruchs,Anspruches, ...
agil	ADJX	0.1959	agilstes,agilster, ...
ehren	VVINF	0.0612	ehret,ehrest, ...

2.3 SB10k

SB10k is a German Twitter corpus with almost 10.000 tweets. It consists of human labeled tweets that can be used for machine learning algorithms for sentiment analysis. In the paper the authors compared how two different classifiers (SVM and CNN) performed on the SB10k corpus as well as on two other additional German corpora (DAI and MGS [7]). Their results showed that the best rating for the SB10k corpus was a $F1 = 0.65$ classification. The corpus is freely available for download (as collection of the relevant Twitter IDs) [2]. SB10ks corpus and classification results will serve as benchmark in section 4.

2.4   SCARE

SCARE is another corpus that offers around 800.000 app reviews for different app categories from the Google Play Store. The reviews have an id, a star rating from 1 to 5, a review headline and a review text. Some categories and reviews will be used for evaluating GerVADERs performance in the app review domain.

3   Process

While VADER started from scratch, for GerVADER some steps can be copied while other steps have to be replicated accordingly for the German language. Just like VADER the development of GerVADER starts with the creation of the lexicon, followed by the ratings of the crowd, filtering of ambiguous words, an adaptation of the booster and negation words including some smaller changes on the source code and lastly the classification tests.

3.1   Constructing the initial lexicon

The initial lexicon is based on the SentiWS dataset. Only the basic forms of the words have been taken into consideration thus resulting in a sum of 3471 words (1.644 positive, 1.827 negative). Additionally, unique German terms have been added to the lexicon that are commonly used in slang expressions and on social media platforms (see Fig. 1). Note however that only single words have been taken account of since phrases cannot be part of the lexicon. Two sources were taken into consideration for the additional words.

One of them is from Langenscheidt which is a German company for language and language-to-language dictionaries. Langenscheidt [8] is in Germany also known for their annual ranking of teenager slang words ("Jugendwörter des Jahres" [9]). Every year Langenscheidt collects words that are commonly and exclusively used by teenagers and young adults. This contest results in the publication of the Top10 "Jugendwörter des Jahres". The Top3 words of the years 2008-2017 have been added to the lexicon as well as the Top10 words of the year 2018 [10]. Only single words have been taken into consideration.

The second source is a website that collects German slang words [11]. Several single words have been manually selected and were added to the lexicon.

Both sources added together have contributed 80 words to the initial lexicon resulting in an overall size of the initial lexicon of 3.546 words. All words are contained in the same lexicon without any initial polarity rating. In the next step a crowd contributed with their individual ratings to the lexicon (see Fig. 1).

For the validation lexicon that is used for validating a user rated batch, a manually written lexicon was created. It consists of more than 100 words. The most positive and most negative words in VADER (e.g. words like "death" or "hell") were collected and manually translated by the author. The gold standard words did not block off any lexicon words so that a word like "Tod" (i.e. "death") was still able to be rated by the user.
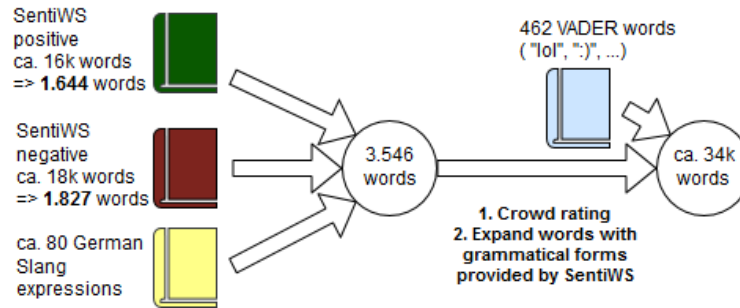
Fig. 1. GerVADER lexicon creation

3.2   Crowd-rating of the lexicon

For GerVADER the crowd consisted of fellow students and friends. The crowd was introduced to the project but did not have to pass any tests or trainings. The list of participants has not been shared so that no rater knew who the other raters were in order to block any communication within the crowd.

For the rating platform a custom-made application has been developed. The raters have been given access to the website with everyone receiving a username and a password. On the site were two main sections.

The first section was a tutorial section where the functionality of the site was briefly explained.

The second section was the component of rating the words in which the user was introduced to a randomly generated batch consisting of 25 words. The server kept track of every raters progress and returned within the batch 20 words that the rater had not yet rated plus 5 gold standard words. The user was then able to rate the words according to the -4 to +4 scale. After rating a batch, the user was allowed to submit it. The server then checked whether the ratings were valid and reviewed whether the gold standard words have been rated correctly. If three or more words of the gold standard words were rated significantly different, the batch has been dropped without notifying the user. The ratings were then not saved so that the user was still able to rate the words in another batch.

Since no participant was financially compensated, motivation was a huge factor. In order to tackle this problem each rater was linked with an animal image. On the main page the number of already rated words was shown as well as the animal pictures of raters who have rated one or more batches on the according day. Thus, feelings of competition and cooperation were invoked. This update was done one week after the release and increased the participation rate of the crowd significantly. Furthermore, a graphic showing the number of rated words by each rater was created periodically and sent to every user via email. Both steps were necessary since the participation was overall not enough to create a lexicon with every word having 10 individual ratings. Within one month all words received round about 7 individual ratings.

3.3   Generating the final lexicon

Similar to VADER, words that have a neutral rating were filtered out for the final lexicon. Also, words with a standard deviation of 2.5 or higher have been removed. For the finalization of the lexicon there have been two additional steps.

The first step involved the VADER lexicon. Since most emoticons and English abbreviations are also common in German social media texts, more than 800 words of this type have been added to GerVADER with their original intensity. Therefore, the users had not to rate common terms like "lol" or ":)" (see Fig. 1).

The second step was to take into consideration that VADER does not do any kind of pre-processing. So the words from the lexicon are directly compared with the words of the sentence being analyzed. However, SentiWS offers for every word multiple grammatical forms. Therefore, those grammatical forms have been added to the lexicon, meaning that every single one of them represents its own word in the lexicon. They received the same rating as the basic form. With the expansion of the lexicon by grammatical forms the size of the lexicon increased to more than 34.000 words.

3.4   Adapting the VADER heuristics

The heuristics are part of the source code and are in some parts not exclusive to the English language. Characteristics like capitalization of words or punctuation marks convey the same meaning for both the German and English languages.

Only 3 of the 5 heuristics had to be adapted for the German language. These heuristics revolve around booster, negation and contrast conjunction words. For GerVADER the English words have been simply translated to German.

3.5   Evaluation and additional steps

Before evaluation, another adjustment has been done to the algorithm. While comparing the words from the text to the lexicon words, the inspected word is transformed to all lower cases. Since the lexicon however contains words with upper and lower cases that are identical and just differ from their POS tag (e.g. "Anstieg" *noun* and "anstieg" *verb*), the lowercase transformation had to be edited. In the lexicon, words can have a different intensity whether it is for example a noun (written in capital in German) or a verb. However, users usually do not care about the correct capitalization of the words. Thus, the following adjustment has been made:

1. Check if the currently inspected word can be found in the lexicon
2. If not, transform the word to all lower cases and recheck the lexicon
3. If not, only capitalize the first letter of the word and recheck the lexicon

At the end of this process if the word has not been found in the lexicon in any of these steps, the next word will be inspected. This adjustment allows to match more words without relying on the user to write the word correctly

regarding the capitalization. Apart from that adjustment no further adjustments have been done on the VADER algorithm.

For the performance evaluation of GerVADER, the SB10k and parts of the SCARE corpus have been used. Since tweets can be deleted, only 7.000+ tweets (about 70%) of the original SB10k corpus could be collected. Therefore, comparing the results to the original results of the authors is not fully reliable. Additionally, in [2] it is left open on which 10% of the corpus the authors have tested on. To conquer this problem, GerVADER comes with the corpus as well as another dataset containing only 10% of the SB10k corpus. For future work this will allow for comparison.

For the SCARE corpus a selection of review categories has been made. Since the classification labels of user reviews are given as stars (1-5), before being able to classify the data with GerVADER the star ratings first have to be translated to as either positive, negative or neutral. To do this as simply as possible, 1+2-star ratings are interpreted as negative, 3 as neutral and 4+5 as positive. The headline will be merged into the comment. If the headline has no punctuation mark at the end of the headline, a dot will be put in between the headline and review text. The idea is to prevent that words from the headline influence the sentiment of the review comment. However, in some cases this might be a problem if the user started a sentence in the headline and continued it in the text, but we assume that these cases are neglectable rare. Only two app review categories will be tested.

4   Results

For testing the performance of GerVADER the obtained SB10k corpus, 10% of the SB10k corpus as well as parts of the SCARE corpus were taken into consideration (see Table 2). For every test the precision, recall and F1 score for every label (pos, neg, neu) is measured. The total F1 score that will be the deciding factor in how well GerVADER and the other classifiers perform is only calculated with the positive and negative F1 score. Additionally, the sum of $F1_{pos}$, $F1_{neg}$ and $F1_{neu}$ will be calculated and called $F1_3$.

Starting with the SB10k corpus, the corpus is already human-labeled and requires no additional work. The results for 7476 tweets (pos, neg, neu) show an overall F1 score of 39,42% (see Table 2, No. 1). The positive F1 score is 43,54% while the negative F1 score is 35,30%. Overall the results show a good recall for positive tweets and a good precision for neutral tweets, however the numbers for the other criteria and labels are less than 50% (see Fig. 2, No. 1). These low numbers decrease the F1 scores significantly. The numbers show that while most positive tweets are classified correctly, the numbers for the negative and neutral tweets are less accurate. Especially for the neutral tweets many tweets have been wrongly classified as positive. The negative tweets however have been almost evenly distributed on all three labels which hints at a problem with detecting negation words. GerVADER achieves better results with filtering out the neutral statements beforehand. In these test cases the F1 score is circa 64%

Table 2. GerVADER results - NN = no neutrals (filtered beforehand), N = neutral, P = positive, $F1_3 = (F1_{pos} + F1_{neg} + F1_{neu})/3$

No.	Classifier	Corpus		F1pos	F1neg	F1neutral	F1	F1-3
		Training	Test					
1	GerVADER	-	SB10k	43,54%	35,30%	40,69%	**39,42%**	39,87%
2	GerVADER	-	SB10k (NN)	74,50%	53,73%	-	**64,12%**	
3	GerVADER	-	SB10k (10%)	44,52%	37,64%	42,01%	**41,08%**	41,39%
4	GerVADER	-	SB10k (10%, NN)	73,15%	55,23%	-	**64,19%**	
5	SVM	SB10k	SB10k (10%)	66,16%	47,80%	81,32%	**56,98%**	65,09%
6	CNN	SB10k	SB10k (10%)	71,46%	58,72%	81,18%	**65,09%**	70,45%
7	SVM	MGS	SB10k (10%)	67,77%	53,23%	80,20%	**60,50%**	67,07%
8	CNN	MGS	SB10k (10%)	63,94%	58,21%	70,66%	**61,08%**	64,27%
9	GerVADER	-	SportNews	85,41%	55,05%	12,71%	**70,23%**	51,06%
10	GerVADER	-	SportNews (NN)	88,07%	57,78%	-	**72,93%**	
11	GerVADER	-	SportNews (NN, N merged into P)	90,72%	57,78%	-	**74,25%**	
12	GerVADER	-	News Apps	80,63%	58,14%	11,78%	**69,39%**	50,18%
13	GerVADER	-	News Apps (NN)	83,73%	60,75%	-	**72,24%**	
14	GerVADER	-	News Apps (NN, N merged into P)	85,77%	60,75%	-	**73,26%**	

(see Table 2, No. 2+4). Remember however that GerVADER is still rating some tweets as neutral, so this classification option is still enabled. Only the neutral statements have been filtered out, the process has been kept the same.

1) GerVADER SB10k

			PREDICTED				
			pos	neg	neu		
			4034	1499	1943		F1
ACTUAL	pos	1717	**1252**	133	332		43,54%
	neg	1130	392	**464**	274		35,30%
	neu	4629	2390	902	**1337**		40,69%
	sum	7476			F1-3 39,84%	F1	**39,42%**

Fig. 2. GerVADER detailed results (see Table 2, No. 1)

Comparing the results with the results by the SB10k authors, one can see that GerVADER does not outperform any classifier (see Table 2, No. 5-8). We have to take into consideration however, that the original test corpus is not publicly available, therefore one can question the results. Nonetheless classifiers that have not trained on the SB10k corpus still reach 20% better results.

In another test GerVADER has been tested on some parts of the SCARE corpus (see Table 2, No. 9-14). Tests have been run with reviews referring to news

and sport news apps. Since the reviews are given with a comment as well as with a star rating, the structure of the reviews had been altered. 1-2-star reviews are interpreted as negative reviews, 3 as neutral and 4-5 as positive reviews.

With the sport news apps results you can see that GerVADER classifies 70% correctly. Especially the F1 score for positive labels is very good with more than 85% (see Table 2, No. 9). If we sort out the neutral reviews before classifying it, meaning that we just classify 1, 2, 4 or 5-star ratings, so either negative or positive labeled reviews, the F1 score raises to 72% (see Table 2, No. 10). Again, GerVADER still rates reviews as neutral. If we now assume that neutral reviews are always positive, since one can suggest that a user would rather review an app that he likes than one he dislikes, we can merge the neutral numbers into the positive ratings in both labeling cases. We then achieve a F1 score of 74,25% with an $F1_{pos}$ score of 90,72% (see Table 2, No. 11).

For the news apps almost equal results have been achieved with the same 3 classification tests (see Table 2, No. 12-14).

In summary one can see that the classifications differ from domain to domain. Comparing with the original VADER classification the F1 scores are significantly lower for multiple reasons that will be discussed in the next chapter.

5   Discussion and future work

Although GerVADER has been tested in domains that VADER achieved its best results, the tests show in some regards rather bad scores. The reason why the overall F1 scores are that low is the classification of neutral and negative texts.

Concerning the negative texts, one has to ask why so many of the data is labeled as positive. Actual negative texts are almost as often rated positive as they are also rated negative. Additionally, also the neutral prediction is in many tests almost as high as both the negative and positive classification. Therefore the $F1_{neg}$ score for negative texts is in all tests much lower than the $F1_{pos}$ score. The reason why this is the case is how in German sentences the negation word "nicht" can often occur at the end of the sentence. It can be very common that the negation is at the end of the sentence (after the verb) while in English the negation word is always in pair with the verb(s).

GerVADER for example does not detect the difference between the following sentences:

1. Ich mag das. (meaning: I like it - i like it)
2. Ich mag das nicht. (meaning: I don't like it - i like it not)

In both cases GerVADER detects the word "mag" as a positive word and therefore calculates the overall sentiment as positive. While the negation word "nicht" is detected, the sentiment is not shifted. The reason is that only the sentiment ratings **after** the negation word are influenced. So, if the negation word appears at the end of the sentence it has no impact on the overall sentiment.

This is a flaw in the current status of GerVADER that needs to be addressed. Such an adjustment however requires an overhaul of the algorithm in order to

be more suitable for the German language. Because English does not have such negated sentences, such a logic is not implemented in VADER. If GerVADER detects the negation words at the end of the sentences, the classification scores will most likely increase for negated sentences. If this is the case, the overall F1 score will also improve. Therefore, one can conclude that GerVADER in its current state biggest flaw is the detection of negative sentences.

Moreover, VADER and therefore GerVADER in its current state have problems with the negation of longer sentences, even if the negation word is at the beginning of the sentence. For example, the following sentence is wrongly rated:

1. Ich finde **nicht**, dass diese Menschen wirklich **freundlich** sind. (rated positive, should be negative)

While the negation word "nicht" is recognized, there are no following words with a sentiment rating except for the word "freundlich" (rated positive). But the occurrence of the word is too far away from the negation word to shift the sentiment of the word. So, for longer sentences the words following the negation word are not shifted if there are too many words in between. In combination with the already mentioned problem with negation words, this gives more insight why many negative texts are falsely rated positive.

Other than that, GerVADER also needs some other improvements. The booster and negation words are only translated from the English original. It lacks a real adaptation for the language. Additionally, there are no phrases covered. Therefore, phrases like "Alles in Butter" are not detected. If you split up the phrase there are no words with a sentiment rating, but read as a phrase it has a positive meaning (meaning "all right").

Furthermore, the lexicon might need more words. Especially German slang words and words that are commonly used in social media texts might be lacking. Moreover, the rating process for the current lexicon could be continued, since a larger crowd promises more trustworthy results.

Lastly a wider benchmark might show better how useful GerVADER in its present state really is.

6 Conclusion

This paper has shown that GerVADER has potential to become another useful tool in the sentiment analysis for the German language. While the results compared to VADER are kind of underwhelming, GerVADER promises many areas for improvement. Given how little the algorithm has actually been altered, the results for the app reviews clearly show that even in its current form it already achieves remarkable results. Just like other lexicon approaches, the lexicon can be easily expanded. Since no machine learning is needed, GerVADER can be basically used plug-and-play and achieves much faster results without needing any sort of training. This also results in a consistent rating, meaning that a text will not be rated differently, where unlike in machine learning approaches the

text might be classified differently depending on the training data. If the proposed adjustments are made, GerVADER might be one of the most viable tools for classifying sentiments. In any case its speed and consistency are two of its biggest strengths, plus the lexicon can be useful for other researches. Given how well the role model performs, there is no reason to doubt that GerVADER will achieve comparable results in the future. GerVADER is publicly available and will be further worked on [12].

References

1. Hutto, C.J. & Gilbert, E.E.: VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14)(2014)
2. Cieliebak, Mark and Deriu, Jan and Egger, Dominic and Uzdilli, Fatih: A Twitter Corpus and Benchmark Resources for German Sentiment Analysis. Social NLP @ EACL. https://doi.org/10.18653/v1/W17-1106
3. Mario Sänger, Ulf Leser, Steffen Kemmerer, Peter Adolphs, and Roman Klinger: SCARE – The Sentiment Corpus of App Reviews with Fine-grained Annotations in German. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).
4. R. Remus, U. Quasthoff & G. Heyer: SentiWS - a Publicly Available German-language Resource for Sentiment Analysis. In Proceedings of the 7th International Language Ressources and Evaluation (LREC'10), pp. 1168-1171. (2010)
5. Ulli Waltinger: Sentiment Analysis Reloaded: A Comparative Study On Sentiment Polarity Identification Combining Machine Learning And Subjectivity Features. In Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST '10). (2010)
6. Bütow, F., Lommatzsch, A., Ploch, D.: Creation of a German Corpus for Internet News Sentiment Analysis. Project report, Berlin Institute of Technology, AOT (2016)
7. Igor Mozetic, Miha Grcar, and Jasmina Smailovic: Multilingual Twitter Sentiment Classification: The Role of Human Annotators. PloS one, 11(5):e0155036. (2016)
8. Wikipedia, Langenscheidt. https://www.langenscheidt.com/ Last accessed 22 Jan 2019
9. Wikipedia, Jugendwort des Jahres. https://de.wikipedia.org/wiki/Jugendwort_des_Jahres_(Deutschland) Last accessed 22 Jan 2019
10. Langenscheidt, Jugendwort des Jahres 2018.https://www.langenscheidt.com/jugendwort-des-jahres Last accessed 22 Jan 2019
11. CoolSlang, German Slang Dictionary. https://www.coolslang.com/ Last accessed 22 Jan 2019
12. Tymann, K.: GerVADER. https://github.com/KarstenAMF/GerVADER