

Native sentiment analysis tools vs. translation services - Comparing GerVADER and VADER

Karsten Michael Tymann, Louis Steinkamp, Oxana Zhurakovskaya, and
Carsten Gips

FH Bielefeld University of Applied Sciences, Minden, Germany
ktymann@fh-bielefeld.de, louis.steinkamp@fh-bielefeld.de,
oxana.zhurakovskaya@fh-bielefeld.de, carsten.gips@fh-bielefeld.de
<https://www.fh-bielefeld.de>

Abstract. VADER is a rule-based sentiment analysis tool for English texts with a social media focus. GerVADER is a German adaptation of VADER, which was developed following the steps of VADER's development process. VADER showed high F1 scores especially for the social media domain, whereas the German adaptation achieved much lower results within the same domain, although on other test data. In this work we examine the question of whether these differences are language-specific. Therefore we apply an improved version of GerVADER to German texts and compare the results with the application of VADER to the same texts that are automatically translated into English. The benchmarking showed, that the translation combined with VADER achieves up to 5% higher F1 scores in all test cases, which can be explained by the translation tools automatic fixing of flawed sentences. However, native language tools can still be viable, since it saves time and costs and does not need another dependency to a third party service.

Keywords: VADER· GerVADER· sentiment analysis· translation

1 Introduction

Sentiment analysis describes the process of automatically rating texts or sentences with a sentiment value. The sentiment value ranges from negative, to neutral to positive and can be expressed as a numeric value or a classification in one of the three sentiment categories. Compared to machine learning based approaches classification can also be done by rule-based algorithms which have the advantage that they do not require any training data. However, developing a rule-based tool requires linguistic knowledge and is significantly more dependent on the target language. Hence one can not simply transfer one language's features, e.g. German, to another language, e.g. English. The languages can differ in their grammar and overall sentence structure, making it very error prone

Copyright © 2020 by the paper's authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

to simply transfer the algorithm to another language. How can negation be detected, or what is the meaning of a specific punctuation? Are there words that can have different meanings in different contexts and how can you derive those? Machine learning approaches simply train on lots of annotated data and extract the features with techniques like word embedding, but for a rule-based approach the developer has to design the process of detecting word patterns themselves. While a rule-based tool does not need bootstrap data, it will still need some form of lexicon with individual words or phrases to derive sentiments from. These lexicons are usually created by humans and are often rated by different people to offer an average sentiment value.

In this work, which was part of a student project at Bielefeld University of Applied Sciences, we will first discuss some improvements to our rule-based sentiment analysis tool GerVADER [5]. Second, we examine the question whether the effort to develop adaptations in the native language like GerVADER is worthwhile at all or whether one could achieve similarly good results by translating the corpora to be examined into English and using VADER [2] subsequently for the analysis. In order to investigate on this, we translated our test corpora to the English language with 3rd party tools and benchmarked the data with the English tool VADER.

2 Background

2.1 VADER & GerVADER

VADER is an abbreviation for “Valence Aware Dictionary and sEntiment Reasoner” and is free to use. It is a rule based tool for analyzing sentiments for English sentences. VADER managed to outperform other lexicon-based approaches as well as machine learning models. Especially in the domain of social media VADER achieved high F1 scores of up to 96%. [2]

GerVADER is an adaptation of the VADER tool for the German language. The process of VADER has been replicated in some steps, such as the crowd rating for the lexicon which is based on the SentiWS lexicon [3], while others have been simply transferred to the German language, such as the heuristics. GerVADER is as well free to use. [5]

2.2 Benchmarking corpora

SCARE is a corpus consisting of Google Play Store reviews. The reviews are categorized by their star ratings (1 to 5) and are split into 11 app categories. In total there are over 800.000 user reviews. [4]

The SB10k corpus consists of German tweets that are humanly labeled into the three sentiment categories: positive, neutral, negative. It consists of almost 10.000 tweets. Both corpora will be used for benchmarking purposes. [1]

2.3 Translation tools

For translating our test data, we have mostly relied on Googles translation service. One of the datasets has been additionally translated with MyMemory.

Googles translation service is based on the Google Neural Machine Translation (GNMT) system. Its hybrid model consists of a Transformer [6] encoder and RNN decoder. The learning is based on sequence-to-sequence neural network learning and is a mix between character and word-delimited models. [7]

MyMemory¹ is a large collection of Translation Memories that are collected and provided by humans and organizations. The translations are saved as words or sequences in databases which can then be matched by the users input. As of now there are over 4 billion human contributions.

3 Process

The process section is divided into two subsections. In the first we analyze the flaws of GerVADER and how we improved the algorithm. In the second subsection we will give insight on the benchmarking itself.

3.1 Flaws in GerVADER

The analysis of the initial version of GerVADER [5] showed three flaws that promised room for improvement. Firstly the negation detection is inaccurate and can not be converted to the German language by simply translating the negation keywords (e.g. 'not'). Secondly booster words (e.g. 'super', 'very') are sometimes the only words with a sentiment meaning in a sentence, but they do not get noticed, since they simply serve as booster for following words valences. Thus the sentences receive a neutral rating, although the booster word itself might carry sentiment meaning (e.g. 'super'). Thirdly misspelled words do not get noticed, since the words have to be written exactly like in the lexicon. For every problem case we developed test corpora so that we were able to tell whether our changes improved the overall rating. Details on the changes can be found on GitHub².

3.2 VADER vs. GerVADER

Both VADER tools cover their own languages. A question however arises whether it even makes sense to translate a sentiment analysis tool to one's native language. To investigate on this we translated our test corpora with Google and MyMemory.

For MyMemory only the SB10k corpus was used, whereas for Google Translate we tested multiple corpora (SB10k, SCARE). Additionally we constructed a SCARE_Balanced corpus, consisting of 400 positive, 400 negative and 400 neutral reviews of each SCARE corpus file. Thus a balanced file of all 3 sentiments is built, consisting of 13.200 entries.

¹ MyMemory by translated LABS <https://mymemory.translated.net/>

² GitHub - GerVADER <https://github.com/KarstenAMF/GerVADER>

Table 1. Benchmarking results for GerVADER and VADER (F1-3: mean of positive, negative, neutral F1-scores)

Tool	Corpus	F1	F1-3
GerVADER2.0	SB10k	39,63%	38,97%
VADER (Google)	SB10k	42,91%	44,20%
VADER (MyMemory)	SB10k	42,95%	44,44%
GerVADER2.0	SportNews (SCARE)	70,26%	50,98%
VADER (Google)	SportNews (SCARE)	71,62%	51,27%
GerVADER2.0	SCARE_Balanced (SCARE)	55,37%	44,06%
VADER (Google)	SCARE_Balanced (SCARE)	58,52%	45,69%

4 Results

GerVADER improved in all three areas for our test corpora by adjusting the original algorithm rules as well as adding new features such as fuzzy-matching. Thus the overall classification score of German texts has increased.

When comparing VADER with GerVADER Table 1 shows that VADER outperforms GerVADER of up to 5%. Even with the improvements in GerVADER, it is still outmatched. This is due to the fact, that translation tools do not just translate sentences word by word, but also consist of features such as fuzzy-matching and entity or POS (part-of-speech) tagging. Those are features, that we have partly integrated into GerVADER as well. Googles API is pre-trained with methods of Deep Learning on millions of data and adjusted to word and phrase sequences and not a simple word-to-word translation. Therefore it goes beyond a simple translation mechanism. As a result, spelling errors are corrected and the overall structure of the sentences is adapted to the desired output language. Thus it is not just a simple translation but also a text correction, which may explain the better results for VADER in the benchmark.

5 Conclusion & Future work

While GerVADER has been improved, looking at the translation comparison the question arises whether GerVADER serves any purpose. Developing a native language adopted tool is challenging and has lots of potential for creating new flaws. It requires linguistic knowledge in the target language, but this allows one to address language specific characteristics more appropriately than with a translation. Also, the translation service is an additional dependency, which can be problematic in factors of costs, time and complexity. With the current version it might not be worth it to trade GerVADER for VADER for a maximum of 5% F1 score improvement. However, translation tools handle the linguistic features for the developer and are therefore an interesting research topic for VADER and similar tools that are available in several languages.

References

- [1] Mark Cieliebak, Jan Deriu, Dominic Egger, and Fatih Uzdilli. “A Twitter Corpus and Benchmark Resources for German Sentiment Analysis.” In: *Proceedings of the 4th International Workshop on Natural Language Processing for Social Media (SocialNLP 2017)*, Valencia, Spain, 2017. 2017. DOI: 10.18653/v1/W17-1106.
- [2] C. Hutto and Eric Gilbert. *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*. 2014. URL: <https://www.aiai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109>.
- [3] R. Remus, U. Quasthoff, and G. Heyer. “SentiWS - a Publicly Available German-language Resource for Sentiment Analysis.” In: *Proceedings of the 7th International Language Resources and Evaluation (LREC’10)*, pp. 1168-1171. (2010). 2010.
- [4] Mario Sanger, 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 2016)*. Portorož, Slovenia, 2016. ISBN: 978-2-9517408-9-1. URL: <https://www.aclweb.org/anthology/L16-1178/>.
- [5] Karsten Michael Tymann, Matthias Lutz, Patrick Palsbroker, and Carsten Gips. “GerVADER - A German adaptation of the VADER sentiment analysis tool for social media texts.” In: *In Proceedings of the Conference “Lernen, Wissen, Daten, Analysen” (LWDA 2019), Berlin, Germany, September 30 - October 2, 2019*. 2019. URL: http://ceur-ws.org/Vol-2454/paper_14.pdf.
- [6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. “Attention Is All You Need”. In: *CoRR* abs/1706.03762 (2017). arXiv: 1706.03762. URL: <http://arxiv.org/abs/1706.03762>.
- [7] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation”. In: *CoRR* abs/1609.08144 (2016). arXiv: 1609.08144. URL: <http://arxiv.org/abs/1609.08144>.