GerVADER

by: Karsten Michael Tymann
  Matthias Lutz
  Patrick Palsbröker
  Carsten Gips
Outline

1. VADER - Benefits and Concept
2. VADER - building block: encyclopedia
3. From VADER to GerVADER
4. GerVADER in six steps
5. Suitable corpura and test data
6. GerVADER Results
VADER – Benefits and concept

- From: C.J. Hutto and Eric Gilbert
- Rated English sentences positively, negatively or neutral (sentiment)
  - On a scale from -1 to 0 to +1
- MIT license, available on GitHub
- Easy to use: 1 Python script
- Without Machine Learning → Lexicon and rule-based analysis tool
- Evaluates individual records using
  - Single word sentiments (encyclopedia)
    → Rating of individual words (pos, neg, neu)
  - 5 Heuristics
    → Reinforcement, reduction, inversion of a sentiment
VADER – Benefits and concept

- Very good scores in social media domains
  ○ Designed specifically for this purpose

VADER is smart, handsome, and funny! positive
{'pos': 0.752, 'compound': 0.8439, 'neu': 0.248, 'neg': 0.0}

Today SUX! negative
{'pos': 0.0, 'compound': -0.5461, 'neu': 0.221, 'neg': 0.779}
## Comparison of VADER to SOTA

<table>
<thead>
<tr>
<th>Test Sets</th>
<th>3-Class Classification Accuracy (F1 scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tweets</td>
</tr>
<tr>
<td>VADER</td>
<td>0.96</td>
</tr>
<tr>
<td>NB (tweets)</td>
<td>0.84</td>
</tr>
<tr>
<td>ME (tweets)</td>
<td>0.83</td>
</tr>
<tr>
<td>SVM-C (tweets)</td>
<td>0.83</td>
</tr>
<tr>
<td>SVM-R (tweets)</td>
<td>0.65</td>
</tr>
<tr>
<td>NB (movie)</td>
<td>0.56</td>
</tr>
<tr>
<td>ME (movie)</td>
<td>0.56</td>
</tr>
<tr>
<td>NB (amazon)</td>
<td>0.69</td>
</tr>
<tr>
<td>ME (amazon)</td>
<td>0.67</td>
</tr>
<tr>
<td>SVM-C (amazon)</td>
<td>0.64</td>
</tr>
<tr>
<td>SVM-R (amazon)</td>
<td>0.54</td>
</tr>
<tr>
<td>NB (nyt)</td>
<td>0.59</td>
</tr>
<tr>
<td>ME (nyt)</td>
<td>0.58</td>
</tr>
</tbody>
</table>

See [VADER] in Sources.
VADER – Building block: Encyclopedia

- For each word a sentiment rating from -4 to 0 to +4
- Compiled from crowd ratings of participants
- Each rating with -4 to +4 (negative to positive word)
- Average value formed
- Word is recognized in sentence and sentiment assigned to it
From VADER to GerVADER

- **Idea:** German adaptation of the English method
- Same algorithm
- Exchange of the English encyclopedia with the German encyclopedia
- Partial adjustment of heuristics and some code lines in the algorithm
GerVADER - in six steps

1. Lexicon compilation
2. Crowd rating the lexicon words by polarity (-4, -3, ..., 0, +1, ..., +4) (Wisdom Of The Crowd)
3. Filter words -> Gold standard lexicon
4. Installation of 5-language heuristics as boost factors
5. e.g. !, !!, !!!, or negation...
6. Small adjustments
7. Testing VADER in different domains against common algorithms (benchmark)
1. **Compile encyclopedia**
   - Initial situation: SentiWS Lexicon
     - Lexicon including grammatical word forms
   - Supplemented by 80 German slang expressions
     - Slang common in social media channels
   → Encyclopedia for the evaluation of the Crowd: 3,546 words

2. **Crowd Rating**
   - Rating of the words (-4 to +4) by 10 headed crowd members
   - Approx. 7 reviews per word received
3. **Filter words → Gold standard lexicon**

- **Filter**: Words with neutral sentiment and high standard deviation (> 2.5) → Crowd disagree in rating
  - Hayvan 2.5 [+1, +1, -1, -4, +4, -4, +0, +0]

- **Expand**: Words extended to grammatical forms, sentiment adopted
  - agil = [agilstes, agilster, agileren, ... ]

- Added cross-language VADER words
  - 462 VADER words ("lol", "rofl", ...)
  - 3,500+ VADER Emojis
Compilation process: Review and mapping
4. Translating the 5-Heuristics from VADER

Punctuation ! can boost the sentiment of the sentence

1. **Punctuation**: ! vs !! vs !!! vs .
   - Punctuation ! can boost the sentiment of the sentence
   - "VADER is smart, handsome, and funny."
   - "VADER is smart, handsome, and funny!"

2. **ALL-CAPS**
   - Strengthens the sentiment of the word
   - "VADER is very smart, handsome, and funny."
   - "VADER is **VERY SMART**, handsome, and **FUNNY!!**"
4. Translating the 5-Heuristics from VADER

3. Booster Words
   - amplify/attenuate the next word
   - “It is good”
   - “It is extremely good”

4. Contrast-conjunction (contrast)
   - Follow-up part is given increased attention
   - “The food is great, but the service sucks”

5. Negating sentence
   - Sentiment tips over
   - "VADER is not smart, handsome, nor funny."

Just **Booster Words** and **Negating sentence** adapted for **GerVADER** → Words were manually translated
5. Small adaptions

- VADER transforms currently viewed word into lowercases
- but in German nouns are written in capital letters
- GerVADER encyclopedia has same words with different POS tag

<table>
<thead>
<tr>
<th>Word</th>
<th>1</th>
<th>0.8</th>
<th>[1,+1,+0,+3,+1,+1,+1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anstieg</td>
<td>1.1</td>
<td>0.8</td>
<td>[+1,+1,+0,+3,+1,+1,+1]</td>
</tr>
<tr>
<td>anstieg</td>
<td>0.7</td>
<td>0.7</td>
<td>[+0,+0,+0,+2,+1,+1]</td>
</tr>
</tbody>
</table>

therefore:

1. search currently viewed word in encyclopedia
2. if not found, lowercase() the word and search in lexicon
3. if not found, first letter capitalize() and search in lexicon
6. Benchmarking - Twitter Corpus + Google Play Corpus

**A Twitter Corpus and Benchmark Resources for German Sentiment Analysis**

- Paper that collected tweets with more than 9000 tweets and manually labeled them (SB10k Corpus)
  - Got only '7500' on request
- Google Play Store Reviews (SCARE Corpus), 800,000+ Reviews
  - Different categories, e.g. Sport News Apps
# GerVADER - Results

<table>
<thead>
<tr>
<th>No.</th>
<th>Classifier</th>
<th>Training Corpus</th>
<th>Test Corpus</th>
<th>F1pos</th>
<th>F1neg</th>
<th>F1neutral</th>
<th>F1</th>
<th>F1-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GerVADER</td>
<td>-</td>
<td>SB10k</td>
<td>43.54%</td>
<td>35.30%</td>
<td>40.69%</td>
<td>39.42%</td>
<td>39.84%</td>
</tr>
<tr>
<td>2</td>
<td>GerVADER</td>
<td>-</td>
<td>SB10k (NN)</td>
<td>74.50%</td>
<td>53.73%</td>
<td>-</td>
<td>64.12%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>GerVADER</td>
<td>-</td>
<td>SB10k (10%)</td>
<td>44.52%</td>
<td>37.64%</td>
<td>42.01%</td>
<td>41.08%</td>
<td>41.39%</td>
</tr>
<tr>
<td>4</td>
<td>GerVADER</td>
<td>-</td>
<td>SB10k (10%, NN)</td>
<td>73.15%</td>
<td>55.23%</td>
<td>-</td>
<td>64.19%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>SVM</td>
<td>SB10k</td>
<td>SB10k (10%)</td>
<td>66.16%</td>
<td>47.80%</td>
<td>81.32%</td>
<td>56.98%</td>
<td>65.09%</td>
</tr>
<tr>
<td>6</td>
<td>CNN</td>
<td>SB10k</td>
<td>SB10k (10%)</td>
<td>71.46%</td>
<td>58.72%</td>
<td>81.18%</td>
<td>65.09%</td>
<td>70.45%</td>
</tr>
<tr>
<td>7</td>
<td>SVM</td>
<td>MGS</td>
<td>SB10k (10%)</td>
<td>67.77%</td>
<td>53.23%</td>
<td>80.20%</td>
<td>60.50%</td>
<td>67.07%</td>
</tr>
<tr>
<td>8</td>
<td>CNN</td>
<td>MGS</td>
<td>SB10k (10%)</td>
<td>63.94%</td>
<td>58.21%</td>
<td>70.66%</td>
<td>61.08%</td>
<td>64.27%</td>
</tr>
<tr>
<td>9</td>
<td>GerVADER</td>
<td>-</td>
<td>SportNews</td>
<td>85.41%</td>
<td>55.05%</td>
<td>12.71%</td>
<td>70.23%</td>
<td>51.06%</td>
</tr>
<tr>
<td>10</td>
<td>GerVADER</td>
<td>-</td>
<td>SportNews (NN)</td>
<td>88.07%</td>
<td>57.78%</td>
<td>-</td>
<td>72.93%</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>GerVADER</td>
<td>-</td>
<td>SportNews (NN, N merged into P)</td>
<td>90.72%</td>
<td>57.78%</td>
<td>-</td>
<td>74.25%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>GerVADER</td>
<td>-</td>
<td>News Apps</td>
<td>80.63%</td>
<td>58.14%</td>
<td>11.78%</td>
<td>69.39%</td>
<td>50.18%</td>
</tr>
<tr>
<td>13</td>
<td>GerVADER</td>
<td>-</td>
<td>News Apps (NN)</td>
<td>83.73%</td>
<td>60.75%</td>
<td>-</td>
<td>72.24%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>GerVADER</td>
<td>-</td>
<td>News Apps (NN, N merged into P)</td>
<td>85.77%</td>
<td>60.75%</td>
<td>-</td>
<td>73.26%</td>
<td></td>
</tr>
</tbody>
</table>
# GerVADER - Results

## SB10k Corpus

<table>
<thead>
<tr>
<th></th>
<th>PREDICTED</th>
<th></th>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>f1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pos</td>
<td>neg</td>
<td>neu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4034</td>
<td>1499</td>
<td>1943</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTUAL</td>
<td>pos</td>
<td>neg</td>
<td>neu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1717</td>
<td>1252</td>
<td>133</td>
<td>332</td>
<td>0,3103</td>
<td>43,54%</td>
</tr>
<tr>
<td>neg</td>
<td>1130</td>
<td>392</td>
<td>464</td>
<td>274</td>
<td>0,3095</td>
<td>35,30%</td>
</tr>
<tr>
<td>neu</td>
<td>4629</td>
<td>2390</td>
<td>902</td>
<td>1337</td>
<td>0,6881</td>
<td>40,08%</td>
</tr>
<tr>
<td></td>
<td>7476</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F1: 39,42%
F1(3): 39,84%

## SB10k Corpus - Neutral tweets previously filtered out

<table>
<thead>
<tr>
<th></th>
<th>PREDICTED</th>
<th></th>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>f1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pos</td>
<td>neg</td>
<td>neu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1644</td>
<td>597</td>
<td>606</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTUAL</td>
<td>pos</td>
<td>neg</td>
<td>neu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1717</td>
<td>1252</td>
<td>133</td>
<td>332</td>
<td>0,7615</td>
<td>74,50%</td>
</tr>
<tr>
<td>neg</td>
<td>1130</td>
<td>392</td>
<td>464</td>
<td>274</td>
<td>0,7772</td>
<td>53,73%</td>
</tr>
<tr>
<td>neu</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0,7859</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2847</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F1: 64,12%
F1(3):
GerVADER - Results

- SB10k can only be compared moderately, since it is not known on which 10% the authors have tested.
- SCARE positive, negative, neutral star classification (1-2 neg; 3 neu; 4-5 pos)
  - **positive statements** are quite well recognized
  - **negative statements** are almost equally distributed on all three labels
  - **neutral statements** are often recognized as positive
Negative statements are almost equally distributed on all 3 labels

GerVADER recognizes negation words in long sentences, but has no effect:

- ‘Ich finde **nicht**, dass diese Menschen wirklich **freundlich** sind.’
  → positive, but should be negative

Negative words only affect the nearest neighbours

- between **nicht** and **freundlich** too great a distance
- with long sentences or funny sentences this leads to wrong classifications
Negative statements are almost equally distributed on all 3 labels

What happens here?

6 negative Ach **verdammt. Nich** groß genug! -.- rated positive

7 negative Ach **verdammt. Groß** genug! -.- rated positive

8 negative Ach **verdammt. Nicht Groß** genug! -.- rated negative

9 negative **Verdammt!** -.- rated neutral

**verdammt & -.-** = is not recognized   **Nich** = is not a deposited negation word

-.- = nicht Teil des Lexikons

→ some words are not recognized, if they are also stored as **booster words**
→ incorrect spelling and words written together are not recognized either
Conclusion
- Copura often lacks context information and missing irony recognition
- Boosters and negation words must be translated/selected more adequately
- German negation ("nicht" at end of sentence) and long sentences with negation are not recognized
- Contrast conjunction "aber" is not even recognized yet
- Subset of emoticons are not recognized or missing in the lexicon
- Potential for improvement, especially with regard to the algorithm

But:
- SCARE Corpus Benchmark already promises good results
- Fixing the above problems could bring GerVADER much closer to VADER results
Sources

more details in paper sources

SentiWS:
In: Proceedings of the 7th International Language Ressources and Evaluation (LREC’10), pp. 1168-1171, 2010

http://wortschatz.uni-leipzig.de/de/download

SCARE:

http://www.romanklinger.de/scare/
Sources

more details in paper sources

SB10k:


https://www.researchgate.net/publication/315362150_A_Twitter_Corpus_and_Benchmark_Resources_for_German_Sentiment_Analysis

VADER:


https://github.com/cjhutto/vaderSentiment