SCARE – The Sentiment Corpus of App Reviews with Fine-grained Annotations in German

Mario Sänger, Ulf Leser, Peter Adolphs, Steffen Kemmerer, Roman Klinger

Summary

- We make the Sentiment Corpus of App Reviews with Fine-grained Annotations in German publicly available.
- Subjective phrases, application aspects and their relations are annotated in 1,760 German mobile application reviews from the Google Play Store.

Motivation

- Sentiment Analysis / Opinion Mining has gained a lot of intention in the last decade
  - Only limited work on analyzing user reviews in app stores is available [1, 3]
- User reviews form a rich information resource for app developers
  - User opinions, bug reports and feature requests are recorded in the reviews
  - Incorporating this feedback may have influence on the success of the app
- No review corpora for fine-grained sentiment analysis exist

Annotation

- Token-/Phrase-based annotation (inspired by the USAGE corpus [2]) of
  - Application aspects (e.g. design, usability)
    - with additional information if corresponding to a different app or feature request
  - Subjective (evaluative) phrases
    - with a polarity (positive, negative, neutral)
- Annotation of target aspects of subjective expressions
- Sampling of 160 user reviews per category ⇒ 1,760 reviews in total
- Annotation was performed by 4 annotators
  - ≈ 20% of all reviews were annotated twice (agreement of 0.72 Fleiss κ)

Corpus Statistics

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Reviews</th>
<th>Avg. length</th>
</tr>
</thead>
<tbody>
<tr>
<td># Reviews</td>
<td>160</td>
<td>200</td>
</tr>
<tr>
<td># Aspects</td>
<td>255</td>
<td>228</td>
</tr>
<tr>
<td># Subj. phrases</td>
<td>393</td>
<td>368</td>
</tr>
<tr>
<td># Positive</td>
<td>309</td>
<td>221</td>
</tr>
<tr>
<td># Negative</td>
<td>78</td>
<td>152</td>
</tr>
<tr>
<td># Relations</td>
<td>217</td>
<td>193</td>
</tr>
</tbody>
</table>

| Fleiss’ κ       | 0.73    | 0.71        |
|                 | 0.73    | 0.71        |
|                 | 0.69    | 0.67        |

| F1 sub. phrases | 0.70    | 0.65        |
| F1 asp. aspects | 0.96    | 0.77        |
| F1 relations    | 0.71    | 0.38        |

F1 scores

Prediction Baseline

- Probabilistic model based on linear-chain conditional random fields
  - Features: token-based, polarity lexicons, word embeddings and context token
  - Homogeneous performance across different app categories

Examples

- Positive evaluation of the app and it’s design
  - Super Reisebegleiter funktioniert einwandfrei, ist einfach zu bedienen und dank den Offline-Karten ideal zum Reisen.

- Mixed evaluation concerning different aspects of an app
  - Runtastic ist einfach der hammer J, aber die angezeigte Werbung nervt mich total.

Review Collection

- Focus on 11 application categories (e.g. instant messengers, fitness tracker, games, social network platforms, music players, news applications)
- Collection of over 800,000 German user reviews from the Google Play Store
- Reviews are very short (avg. 17 tokens)
  - Negative reviews are longer than positive ones (25 vs. 13 tokens)
  - ≈ 70% of all reviews have a 1- or 5-star rating

References


The annotated corpus and the complete review data set is available at http://www.romanklinger.de/scare/.