

# 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 attention 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  $\Rightarrow$  1,760 reviews in total
- Annotation was performed by 4 annotators
  - $\approx 20\%$  of all reviews were annotated twice (**agreement of 0.72 Fleiss  $\kappa$** )

## 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)
- $\approx 70\%$  of all reviews have a 1- or 5-star rating

## Examples

**Max Mustermann** 20.März 2016  
★★★★★

**Super Reisebegleiter** Funktioniert einwandfrei ist einfach zu bedienen und dank den Offline-Karten ideal zum Reisen.

**Petra Musterfrau** 17.Februar 2016  
★★★★★

**Echt Schrott fürs Tablet (7")** Extrem lange Ladezeit, und Bildgröße paßt sich nicht an. Nicht spielbar auf dem Tablet.

- Positive evaluation of the app and its design
 

Sehr gute und schnelle App. Das Design finde ich sehr hübsch.
- Mixed evaluation concerning different aspects of an app
 

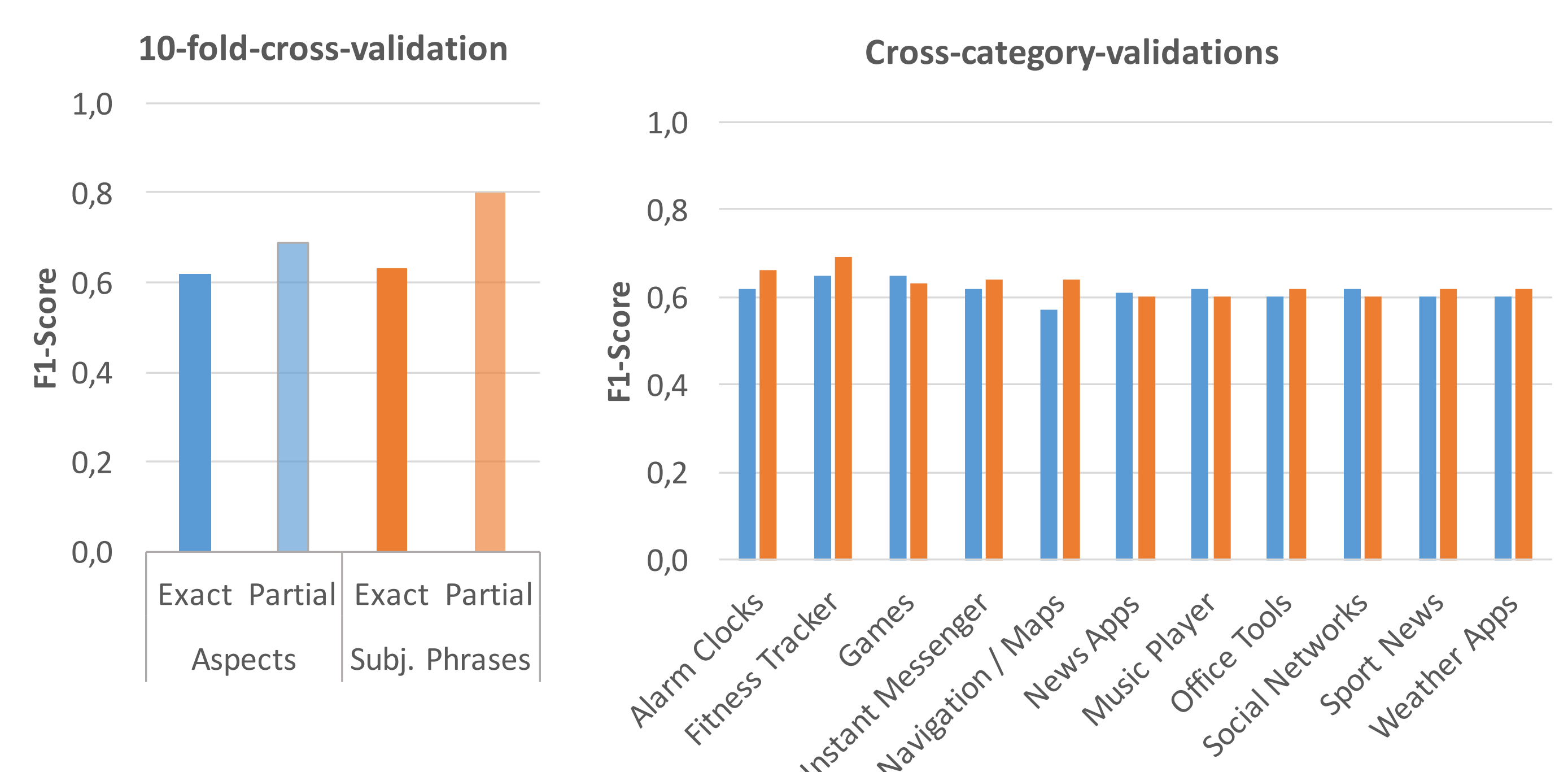
Runtastic ist einfach der hammer 😊, aber die angezeigte Werbung nervt mich total.

## Corpus Statistics

	Alarm Clocks	Fitness Tracker	Games	Instant Messenger	Navigation / Maps	News Apps	Music Player	Office Tools	Social Networks	Sport News	Weather Apps	Total
# Reviews	160	160	160	160	160	160	160	160	160	160	160	1,760
Avg. length	18.6	17.8	17.2	16.9	21.0	22.0	21.6	23.0	13.5	16.9	14.6	18.5
# Aspects	255	233	190	200	232	287	240	255	149	221	225	2,487
# Subj. phrases	393	339	357	327	378	375	383	368	315	374	350	3,959
# Positive	309	250	230	159	233	211	228	221	158	211	253	2,463
# Negative	78	87	123	163	139	158	146	137	152	161	89	1,433
# Relations	217	176	146	146	202	237	182	193	104	173	193	1,969
Fleiss' $\kappa$	0.73	0.74	0.68	0.73	0.73	0.71	0.69	0.71	0.73	0.72	0.74	0.72
$F_1$ subj. phrases	0.70	0.73	0.61	0.64	0.77	0.71	0.65	0.68	0.67	0.71	0.76	0.69
$F_1$ aspects	0.86	0.82	0.67	0.67	0.79	0.81	0.73	0.79	0.79	0.77	0.82	0.78
$F_1$ relations	0.71	0.71	0.44	0.43	0.68	0.64	0.53	0.58	0.38	0.68	0.80	0.62

## 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



## References

- [1] Claudia Iacob and Rachel Harrison. Retrieving and analyzing mobile apps feature requests from online reviews. In *Proceedings of the 10th IEEE Working Conference on Mining Software Repositories*, 2013.
- [2] Roman Klinger and Philipp Cimiano. The USAGE review corpus for fine-grained multi-lingual opinion analysis. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 2014.
- [3] Walid Maalej and Hadeer Nabil. Bug report, feature request, or simply praise? on automatically classifying app reviews. In *Proceedings of the IEEE 23rd International Requirements Engineering Conference*, 2015.



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