

Mining and Analyzing User Feedback

How can user feedback in app reviews be analyzed automatically? How can Twitter be used as a channel for user feedback and how can tweets be analyzed automatically?

Luc Boillat und Laura Salathe
Group 3

Social Computing Approaches for RE

Potential Benefits

- Steady income of potential requirements and feedback for iterations
 - End users (Social media, App Stores)
 - Other Stakeholders (Issue tracker)
- *Can* aid traditional requirements elicitation without much effort

Limitations

- Threats to validity of information
- *Can* result in great efforts when using unfiltered data

What does existing software lack to support RE?

- Give experienced users option to specify/categorize feedback

Why using user feedback?

- Online platforms are a very rich source for user feedback
- Previous research found that user feedback is essential for software quality and for identifying ideas for improvement

But..

- There is a lot of data to choose from: 30,000 tweets for popular applications such as Facebook and Snapchat on Twitter
- Problem: How to find the information you are looking for?



Maalej and Nabil (2015) and Guzmán et al. (2017)

Authors	Walid Maalej and Hadeer Nabil	Emitza Guzman, Mohamed Ibrahim and Martin Glinz
Data Source	Google Store and Apple AppStore	Twitter
Number of observations	Google Store: 146,057 reviews for 40 apps: AppStore: ~1.1 million reviews for 1100 apps 50% paid and 50% free	68,108 collected tweets from Twitter Dropbox, Spotify, Slack
Classification algorithms	Naive Bayes Multinomial logistic regression (MaxEnt) Decision Tree	Multinomial Naive Bayes (MNB)
Topic modeling algorithms		Biterm Topic Model (BTM) Latent Dirichlet Allocation (LDA)

Maalej and Nabil (2015)

Goals:

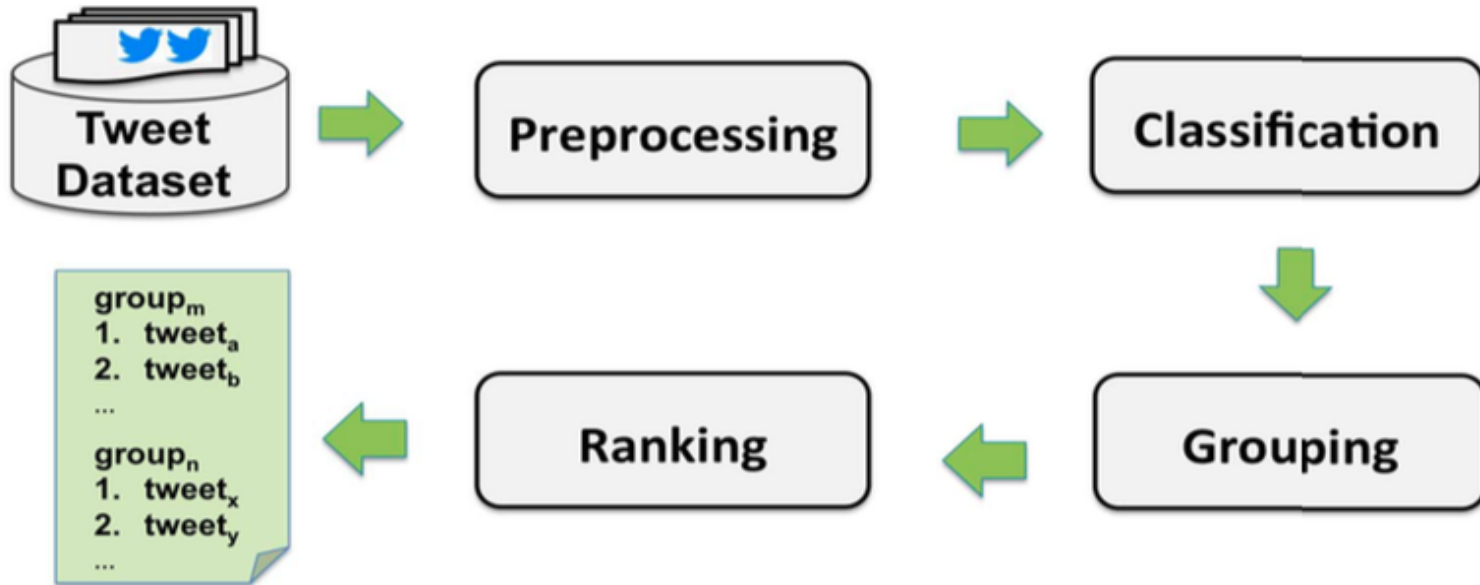
- Propose an approach to systematically analyze explicit, informal user feedback such as reviews published in App Stores
- Classify app reviews to bug reports, feature requests, user experiences, and ratings

Guzmán, Ibrahim and Glinz (2017)

Goals:

- present ALERTme(A Little biRd Told me)
 - classify tweets into categories related to software evolution
 - group tweets according to their content
 - rank tweets according to their relevance
- present empirical evidence that state-of-the-art techniques have a good performance when classifying and grouping tweets
- empirically assess that our novel technique for ranking tweets yields promising results

ALERTme approach



Methods used to automatically classifying reviews

- String Matching
- Bag of words (BOW)

Increasing accuracy by using natural language processing techniques (NLP):

- Stopwords
- Lemmatization /Stemming
- or review metadata

Some examples from the data analyzed by Maalej and Nabil (2015)

I wish I could give it zero stars

I've not been able to use the app because it keeps crashing! What a PAIN!

Few requests

Hey Dropbox team! Please add the ability to rename files, as well as the ability for gifs to play I know you can do it!!

Some examples from the data analyzed by Maalej and Nabil (2015)

```
"comment": "I've not been able to use the app because it keeps crashing! What a PAIN!",
"rating": 1,
"past": 1,
"stopwords_removal": "i've not been able to use app because keeps crashing! what pain!",
"reviewer": null,
"id": 250,
"stemmed": "i've not been abl to us the ap beca it keep crashing! what a pain!",
"fee": null,
"title": "I wish I could give it zero stars",
"label": "Bug",
"future": 0,
"lemmatized_comment": "i've not be able to use the app because it keep crashing! what a pain!",
"sentiScore": -4,
"sentiScore_neg": -4,
"reviewId": 5445,
"stopwords_removal_nltk": "i've able use app keeps crashing! pain!",
"present_simple": 2,
"dataSource": "RE2014_app_and_play_store_apps",
"appId": null,
"date": null,
"sentiScore_pos": 2,
"present_con": 1,
"length_words": 20,
"stopwords_removal_lemmatization": "i've not be able to use app because keep crashing! what pain!"
```

```
"comment": "Hey Dropbox team! Please add the ability to rename files, as well as the ability for gifs to play I know you can do it!!",
"rating": 4,
"past": 0,
"stopwords_removal": "hey dropbox team! please add ability to rename files, as well as ability for gifs to play know can do it!!",
"reviewer": null,
"id": 233,
"stemmed": "hey dropbox team! plea ad the abl to renam files, as wel as the abl for gif to play i know you can do it!!",
"fee": null,
"title": "Few requests",
"label": "Feature",
"future": 0,
"lemmatized_comment": "hey dropbox team! please add the ability to rename files, a well a the ability for gifs to play i know you can do it!!",
"sentiScore": 2,
"sentiScore_neg": -1,
"reviewId": 2453,
"stopwords_removal_nltk": "hey dropbox team! please add ability rename files, well ability gifs play know it!!",
"present_simple": 5,
"dataSource": "RE2014_app_and_play_store_apps",
"appId": null,
"date": null,
"sentiScore_pos": 2,
"present_con": 0,
"length_words": 29,
"stopwords_removal_lemmatization": "hey dropbox team! please add ability to rename files, a well a ability for gifs to play know can do it!!",
"Exclude": "\u0000"
```

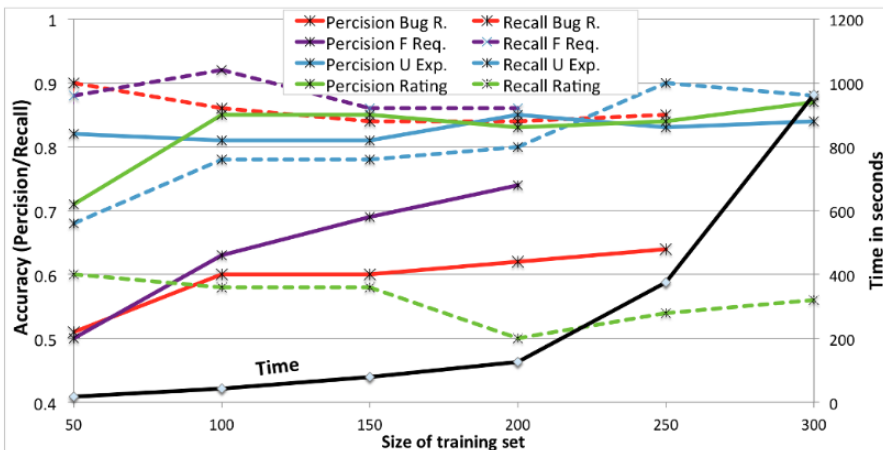
Results

-For classification: 150-200 seems a good size of the training sets for each type, allowing for a high accuracy while saving resources

-can be classified to bug reports, feature requests, user experiences, and ratings (praise or dispraise) with a high accuracy between 70-97%.

-stopword removal and lemmatization should be used carefully, since every word in a short informal review can be informative

-ALERTme is able to classify tweets automatically into the improvement request and other categories with a F-Measure of 0.79.



Source: Maalej and Nabil (2015)

Problems and discussion

- Truth set created by humans, but humans make mistakes

Solution: *Use multiple coders, provide coding guide, encourage discussion*

- Missed other machine learning features, algorithms, or metrics

But: *they are not claiming the completeness of the results -> More research needed*

- Not all end Users can be reached with these techniques e.g. old people, people living in countries with restricted internet access

But: *still a lot of information can be gained this way*

References

Guzmán, E., M. Ibrahim, M. Glinz (2017). A Little Bird Told Me: Mining Tweets for Requirements and Software Evolution. 25th IEEE International Requirements Engineering Conference (RE'17), Lisbon, Portugal. 11–20.

Maalej, W., H. Nabil (2015). Bug Report, Feature Request, or Simply Praise? On Automatically Classifying App Reviews. *23th IEEE International Requirements Engineering Conference (RE'15)*, Ottawa, Canada. 116–125.