Aspect-Oriented Opinion Mining from User Reviews in Croatian

Goran Glavaš, Damir Korenčić, Jan Šnajder
User review

Really laudable! Food was delivered 15 minutes early. We ordered pizza which was filled with extras, well-baked, and very tasteful.

Rating: 6/6

- Aspect-oriented opinion mining
- Construction of opinion lexicon
  - product aspects
  - opinion clues
- Extraction of opinionated aspects
- Prediction of overall review opinion
Really laudable! Food was delivered 15 minutes early. We ordered pizza which was filled with extras, well-baked, and very tasteful.

- Lexicon
  - aspects: food, deliver, pizza
  - clues: laudable, early, filled, well-baked, tasteful

- Opinionated aspects
  - (deliver, early)
  - (pizza, filled), (pizza, well-baked), (pizza, tasteful)

- Review opinion
  - positive
  - 6/6
Preprocessing

- Spell checking with GNU Aspell
- Lemmatization [Šnajder et al., 2008]
- POS tagging [Agić et al., 2008]
- Dependancy parsing [Agić, 2012]
Opinion lexicon

- Candidates for positive/negative clues are lemmas that appear much more frequently in positive/negative reviews
- Aspect candidates are lemmas that frequently co-occur with opinion clues
- Manual filtering of the initial lists of candidates
Opinionated aspects

- Pairing of aspects with the opinion clues that target them
- Polarity of the (aspect, clue) pair can be inverted
  - *the pizza is never cold*
  - *cold pizza vs. cold ice-cream*
- Generate all the (aspect, clue) candidate pairs within a sentence
- Supervised classification of candidates into *paired* or *not paired* classes
Opinionated aspects

- Basic features
  - distance, sentence length, number of aspects and clues
  - punctuation, other aspects and clues in between, order

- Lexical features
  - lemmas of aspect and clue, bag of lemmas in between
  - conjunction of aspect or clue with another aspect or clue

- Part-of-speech features
  - POS tags, tags in between, before and after the pair
  - agreement of gender and number

- Syntactic dependency features
  - relation labels along the path from the aspect to the clue
  - is the aspect syntactically the closest to the clue?
  - is the clue syntactically the closest to the aspect?
Opinionated aspects

- Reviews crawled from pauza.hr
- Trained on 200 sentences, 1406 aspect-clue pairs
- Tested on 70 sentences, 308 aspect-clue pairs
- libSVM [Chang & Lin, 2011] for classification
- Baseline assigns to each aspect the closest opinion clue within the sentence
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>31.8</td>
<td>71.0</td>
<td>43.9</td>
</tr>
<tr>
<td>Basic</td>
<td>77.2</td>
<td>76.1</td>
<td>76.6</td>
</tr>
<tr>
<td>Basic+Lex</td>
<td>78.1</td>
<td><strong>82.6</strong></td>
<td>80.3</td>
</tr>
<tr>
<td>Basic+Lex+POS</td>
<td>80.9</td>
<td>79.7</td>
<td>80.3</td>
</tr>
<tr>
<td>Basic+Lex+POS+Syntax</td>
<td><strong>84.1</strong></td>
<td>80.4</td>
<td><strong>82.2</strong></td>
</tr>
</tbody>
</table>

- Models with linguistic features outperform Basic model.
- No significant difference between linguistic feature sets.
Overall opinion prediction

- Review polarity prediction – binary classification
- Review rating prediction – regression
- Features
  - tf-idf weighted bag-of-word representation of the review
  - number of tokens in the review
  - number of positive and negative emoticons
  - number and the lemmas of positive and negative clues
  - number and lemmas of positively and negatively opinionated aspects
Overall opinion prediction

- 3310 reviews, 100K tokens
- For polarity prediction we consider ratings $\leq 2.5$ as negative and $\geq 4$ as positive
- libSVM [Chang & Lin, 2011] for classification and regression
- Baseline – bag-of-words model
## Opinionated aspects

### Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Review polarity</th>
<th>Review rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pos F1</td>
<td>Neg F1</td>
</tr>
<tr>
<td>BoW</td>
<td>94.1</td>
<td>79.1</td>
</tr>
<tr>
<td>BoW+E</td>
<td>94.4</td>
<td>80.3</td>
</tr>
<tr>
<td>BoW+E+A</td>
<td>95.7</td>
<td>85.2</td>
</tr>
<tr>
<td>BoW+E+C</td>
<td>95.7</td>
<td>85.6</td>
</tr>
<tr>
<td>BoW+E+A+C</td>
<td>96.0</td>
<td>86.2</td>
</tr>
</tbody>
</table>

E – emoticons; A – opinionated aspects; C – opinion clues

- aspect and clue features outperform the BoW baseline
- no significant difference between aspect and clue features
Conclusion

- We presented a method for aspect-oriented opinion mining from domain-specific user reviews in Croatian.
- Supervised model with linguistic features is effective for assigning opinions to the product aspects.
- Opinion clues and opinionated aspects improve prediction of overall review polarity and rating.
- Future work:
  - Evaluation of the method on other domains
  - Aspect-based opinion summarization
Thanks for your attention!

Text Analysis and Knowledge Engineering Lab
www.takelab.hr
References


