



take[lab];

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

User review

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

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

- 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

- 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

- 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 ?

- 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

Model	Precision	Recall	F1
Baseline	31.8	71.0	43.9
Basic	77.2	76.1	76.6
Basic+Lex	78.1	82.6	80.3
Basic+Lex+POS	80.9	79.7	80.3
Basic+Lex+POS+Syntax	84.1	80.4	82.2

- models with linguistic features outperform Basic model
- no significant difference between linguistic feature sets

- 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

- 3310 reviews, 100K tokens
- For polarity prediction we consider ratings ≤ 2.5 as negative and ≥ 4 as positive
- libSVM [Chang & Lin, 2011] for classification and regression
- Baseline – bag-of-words model

Results

Model	Review polarity			Review rating	
	Pos F1	Neg F1	Avg F1	r	MAE
BoW	94.1	79.1	86.6	0.74	0.94
BoW+E	94.4	80.3	87.4	0.75	0.91
BoW+E+A	95.7	85.2	90.5	0.80	0.82
BoW+E+C	95.7	85.6	90.7	0.81	0.79
BoW+E+A+C	96.0	86.2	91.1	0.83	0.76

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

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

- 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!

take[lab];

Text Analysis and Knowledge Engineering Lab
www.takelab.hr

Agić, Ž. (2012).

K-best spanning tree dependency parsing with verb valency lexicon reranking.

In *Proceedings of 24th international Conference on Computational Linguistics (COLING 2012): Posters* (pp. 1–12).

Agić, Ž., Tadić, M., & Dovedan, Z. (2008).

Improving part-of-speech tagging accuracy for Croatian by morphological analysis.

Informatica, 32(4), 445–451.

Chang, C.-C. & Lin, C.-J. (2011).

Libsvm: a library for support vector machines.

ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 27.

Šnajder, J., Bašić, B., & Tadić, M. (2008).

Automatic acquisition of inflectional lexica for morphological normalisation.

Information Processing & Management, 44(5), 1720–1731.