

Aspect-Oriented Opinion Mining from User Reviews in Croatian

Goran Glavaš* Damir Korenčić† Jan Šnajder*

*University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia

†Ruđer Bošković Institute, Department of Electronics
Bijenička cesta 54, 10000 Zagreb, Croatia

{goran.glavas, jan.snajder}@fer.hr damir.korencic@irb.hr

Abstract

Aspect-oriented opinion mining aims to identify product aspects (features of products) about which opinion has been expressed in the text. We present an approach for aspect-oriented opinion mining from user reviews in Croatian. We propose methods for acquiring a domain-specific opinion lexicon, linking opinion clues to product aspects, and predicting polarity and rating of reviews. We show that a supervised approach to linking opinion clues to aspects is feasible, and that the extracted clues and aspects improve polarity and rating predictions.

1 Introduction

For companies, knowing what customers think of their products and services is essential. Opinion mining is being increasingly used to automatically recognize opinions about products in natural language texts. Numerous approaches to opinion mining have been proposed, ranging from domain-specific (Fahrni and Klenner, 2008; Qiu et al., 2009; Choi et al., 2009) to cross-domain approaches (Wilson et al., 2009; Taboada et al., 2011), and from lexicon-based methods (Popescu and Etzioni, 2007; Jijkoun et al., 2010; Taboada et al., 2011) to machine learning approaches (Boiy and Moens, 2009; Go et al., 2009).

While early attempts focused on classifying overall document opinion (Turney, 2002; Pang et al., 2002), more recent approaches identify opinions expressed about individual product aspects (Popescu and Etzioni, 2007; Fahrni and Klenner, 2008; Mukherjee and Liu, 2012). Identifying opinionated aspects allows for aspect-based comparison across reviews and enables opinion summarization

for individual aspects. Furthermore, opinionated aspects may be useful for predicting overall review polarity and rating.

While many opinion mining systems and resources have been developed for major languages, there has been considerably less development for less prevalent languages, such as Croatian. In this paper we present a method for domain-specific, aspect-oriented opinion mining from user reviews in Croatian. We address two tasks: (1) identification of opinion expressed about individual product aspects and (2) predicting the overall opinion expressed by a review. We assume that solving the first task successfully will help improve the performance on the second task. We propose a simple semi-automated approach for acquiring domain-specific lexicon of opinion clues and prominent product aspects. We use supervised machine learning to detect the links between opinion clues (e.g., *excellent*, *horrible*) and product aspects (e.g., *pizza*, *delivery*). We conduct preliminary experiments on restaurant reviews and show that our method can successfully pair opinion clues with the targeted aspects. Furthermore, we show that the extracted clues and opinionated aspects help classify review polarity and predict user-assigned ratings.

2 Related Work

Aspect-based opinion mining typically consists of three subtasks: sentiment lexicon acquisition, aspect-clue pair identification, and overall review opinion prediction. Most approaches to domain-specific sentiment lexicon acquisition start from a manually compiled set of aspects and opinion clues and then expand it with words satisfying certain co-occurrence or syntactic criteria in a domain-specific corpus (Kanayama and Nasukawa, 2006; Popescu and Etzioni, 2007; Fahrni and Klenner, 2008; Mukherjee and Liu, 2012). Kobayashi et

al. (2007) extract aspect-clue pairs from weblog posts using a supervised model with parts of dependency trees as features. Kelly et al. (2012) use a semi-supervised SVM model with syntactic features to classify the relations between entity-property pairs. Opinion classification of reviews has been approached using supervised text categorization techniques (Pang et al., 2002; Funk et al., 2008) and semi-supervised methods based on the similarity between unlabeled documents and a small set of manually labeled documents or clues (Turney, 2002; Goldberg and Zhu, 2006).

Sentiment analysis and opinion mining approaches have been proposed for several Slavic languages (Chetviorkin et al., 2012; Buczynski and Wawer, 2008; Smrž, 2006; Smailović et al., 2012). Methods that rely on translation, using resources developed for major languages, have also been proposed (Smrž, 2006; Steinberger et al., 2012). Thus far, there has been little work on opinion mining for Croatian. Glavaš et al. (2012) use graph-based algorithms to acquire a sentiment lexicon from a newspaper corpus. Agić et al. (2010) describe a rule-based method for detecting polarity phrases in financial domain. To the best of our knowledge, our work is the first that deals with aspect-oriented opinion mining for Croatian.

3 Aspect-Oriented Opinion Mining

Our approach consists of three steps: (1) acquisition of an opinion lexicon of domain-specific opinion clues and product aspects, (2) recognition of aspects targeted by opinion clues, and (3) prediction of overall review polarity and opinion rating.

The linguistic preprocessing includes sentence segmentation, tokenization, lemmatization, POS-tagging, and dependency parsing. We use the inflectional lexicon from Šnajder et al. (2008) for lemmatization, POS tagger from Agić et al. (2008), and dependency parser from Agić (2012). As we are dealing with noisy user-generated text, prior to any of these steps, we use GNU Aspell tool¹ for spelling correction.

Step 1: Acquisition of the opinion lexicon. We use a simple semi-automatic method to acquire opinion clues and aspects. We identify candidates for positive clues as lemmas that appear much more frequently in positive than in negative reviews (we determine review polarity based on user-assigned

rating). Analogously, we consider as negative clue candidates lemmas that occur much more frequently in negative than in positive reviews. Assuming that opinion clues target product aspects, we extract as aspect candidates all lemmas that frequently co-occur with opinion clues. We then manually filter out the false positives from the lists of candidate clues and aspects.

Unlike some approaches (Popescu and Etzioni, 2007; Kobayashi et al., 2007), we do not require that clues or aspects belong to certain word categories or to a predefined taxonomy. Our approach is pragmatic – clues are words that express opinions about aspects, while aspects are words that opinion clues target. For example, we treat words like *stići* (*to arrive*) and *sve* (*everything*) as aspects, because they can be targets of opinion clues, as in “*pizza je stigla kasno*” (“*pizza arrived late*”) and “*sve super!*” (“*everything’s great!*”).

Step 2: Identifying opinionated aspects. We aim to pair in each sentence the aspects with the opinion clues that target them. For example, in “*dobra pizza, ali lasanje su užasne*” (“*good pizza, but lasagna was terrible*”), the clue *dobra* (*good*) should be paired with the aspect *pizza*, and *užasne* (*terrible*) should be paired with *lasanje* (*lasagne*).

In principle, the polarity of an opinion is determined by both the opinion clue and the aspect. At an extreme, an aspect can invert the prior polarity of an opinion clue (e.g., “*cold pizza*” has a negative, whereas “*cold ice-cream*” has a positive polarity). However, given that no such cases occurred in our dataset, we chose not to consider this particular type of inversion. On the other hand, the polarity of an opinion may be inverted explicitly by the use of negations. To account for this, we use a very simple rule to recognize negations: we consider an aspect-clue pair to be negated if there is a negation word within a ± 3 token window of the opinion clue (e.g., “*pizza im nikad nije hladna*” – “*their pizza is never cold*”).

To identify the aspect-clue pairs, we train a supervised model that classifies all possible pairs within a sentence as either *paired* or *not paired*. We use four sets of features:

(1) *Basic features*: the distance between the aspect and the clue (in number of tokens); the number of aspects and clues in the sentence; the sentence length (in number of tokens); punctuation, other aspects, and other clues in between the aspect and the clue; the order of the aspect and the clue (i.e.,

¹<http://aspell.net/>

which one comes before);

(2) *Lexical features*: the aspect and clue lemmas; bag-of-words in between the aspect and the clue; a feature indicating whether the aspect is conjoined with another aspect (e.g., “*pizza i sendvič su bili izvrsni*” – “*pizza and sandwich were amazing*”); a feature indicating whether the clue is conjoined with another clue (e.g., “*velika i slasna pizza*” – “*large and delicious pizza*”);

(3) *Part-of-speech features*: POS tags of the aspect and the clue word; set of POS tags in between the aspect and the clue; set of POS tags preceding the aspect/clue; set of POS tags following the aspect/clue; an agreement of gender and number between the aspect and the clue;

(4) *Syntactic dependency features*: dependency relation labels along the path from the aspect to the clue in the dependency tree (two features: a concatenation of these labels and a set of these labels); a feature indicating whether the given aspect is syntactically the closest to the given clue; a feature indicating whether the given clue is syntactically the closest to given aspect.

Step 3: Predicting overall review opinion. We use extracted aspects, clues, and aspect-clue pairs to predict the overall review opinion. We consider two separate tasks: (1) prediction of review polarity (positive or negative) and (2) prediction of user-assigned rating that accompanies a review. We frame the first task as a binary classification problem, and the second task as a regression problem. We use the following features for both tasks:

(1) *Bag-of-word (BoW)*: the standard tf-idf weighted BoW representation of the review;

(2) *Review length*: the number of tokens in the review (longer reviews are more likely to contain more opinion clues and aspects);

(3) *Emoticons*: the number of positive (e.g., “:)”) and negative emoticons (e.g., “: (”);

(4) *Opinion clue features*: the number and the lemmas of positive and negative opinion clues;

(5) *Opinionated aspect features*: the number and the lemmas of positively and negatively opinionated aspects.

4 Evaluation

For experimental evaluation, we acquired a domain-specific dataset of restaurant reviews² from

²Available under CC BY-NC-SA license from <http://takelab.fer.hr/cropinion>

(HR) *Zaista za svaku **pohvalu!** Jelo su nam dostavili 15 minuta **ranije.** Naručili smo pizzu koja je bila prepuna dodataka, **dobro pečena,** i vrlo **ukusna.***

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

Rating: 6/6

Table 1: Example of a review (text and rating)

Pauza.hr,³ Croatia’s largest food ordering website. The dataset contains 3310 reviews, totaling about 100K tokens. Each review is accompanied by an opinion rating on a scale from 0.5 (worst) to 6 (best). The average user rating is 4.5, with 74% of comments rated above 4. We use these user-assigned ratings as gold-standard labels for supervised learning. Table 1 shows an example of a review (clues are bolded and aspects are underlined). We split the dataset into a development and a test set (7:3 ratio) and use the former for lexicon acquisition and model training.

Experiment 1: Opinionated aspects. To build a set on which we can train the aspect-clue pairing model, we sampled 200 reviews from the development set and extracted from each sentence all possible aspect-clue pairs. We obtained 1406 aspect-clue instances, which we then manually labeled as either *paired* or *not paired*. Similarly for the test set, we annotated 308 aspect-clue instances extracted from a sample of 70 reviews. Among the extracted clues, 77% are paired with at least one aspect and 23% are unpaired (the aspect is implicit).

We trained a support vector machine (SVM) with radial basis kernel and features described in Section 3. We optimized the model using 10-fold cross-validation on the training set. The baseline assigns to each aspect the closest opinion clue within the same sentence. We use stratified shuffling test (Yeh, 2000) to determine statistical significance of performance differences.

Results are shown in Table 2. All of our supervised models significantly outperform the closest clue baseline ($p < 0.01$). The *Basic+Lex+POS+Synt* model outperforms *Basic* model (F-score difference is statistically significant at $p < 0.01$), while the F-score differences between *Basic* and both *Basic+Lex* and *Basic+Lex+POS* are pairwise significant at $p < 0.05$. The F-score

³<http://pauza.hr/>

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+Synt	84.1	80.4	82.2

Table 2: Aspect-clue pairing performance

Model	Review polarity			Review rating	
	Pos	Neg	Avg	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

Table 3: Review polarity and rating performance

differences between *Basic+Lex*, *Basic+Lex+POS*, and *Basic+Lex+POS+Synt* are pairwise not statistically significant ($p < 0.05$). This implies that linguistic features increase the classification performance, but there are no significant differences between models employing different linguistic feature sets. We also note that improvements over the *Basic* model are not as large as we expected; we attribute this to the noisy user-generated text and the limited size of the training set.

Experiment 2: Overall review opinion. We considered two models: a classification model for predicting review polarity and a regression model for predicting user-assigned rating. We trained the models on the full development set (2276 reviews) and evaluated on the full test set (1034 reviews). For the classification task, we consider reviews rated lower than 2.5 as negative and those rated higher than 4 as positive. Ratings between 2.5 and 4 are mostly inconsistent (assigned to both positive and negative reviews), thus we did not consider reviews with these ratings. For classification, we used SVM with radial basis kernel, while for regression we used support vector regression (SVR) model. We optimized both models using 10-fold cross-validation on the training set.

Table 3 shows performance of models with different feature sets. The model with bag-of-words features (*BoW*) is the baseline. For polarity classification, we report F1-scores for positive and negative class. For rating prediction, we report Pearson

correlation (r) and mean average error (MAE).

The models that use opinion clue features (*BoW+E+C*) or opinionated aspect features (*BoW+E+A* and *BoW+E+A+C*) outperform the baseline model (difference in classification and regression performance is significant at $p < 0.05$ and $p < 0.01$, respectively; tested using stratified shuffling test). This confirms our assumption that opinion clues and opinionated aspects improve the prediction of overall review opinion. Performance on negative reviews is consistently lower than for positive reviews; this can be ascribed to the fact that the dataset is biased toward positive reviews. Models *BoW+E+A* and *BoW+E+C* perform similarly (the difference is not statistically significant at $p < 0.05$), suggesting that opinion clues improve the performance just as much as opinionated aspects. We believe this is due to (1) the existence of a considerable number (23%) of unpaired opinion clues (e.g., *užasno* (*terrible*) in “*Bilo je užasno!*” (“*It was terrible!*”)) and (2) the fact that most opinionated aspects inherit the prior polarity of the clue that targets them (also supported by the fact the *BoW+E+A+C* model does not significantly outperform the *BoW+E+C* nor the *BoW+E+A* models). Moreover, note that, in general, user-assigned ratings may deviate from the opinions expressed in text (e.g., because some users chose to comment only on some aspects). However, the issue of annotation quality is out of scope and we leave it for future work.

5 Conclusion

We presented a method for aspect-oriented opinion mining from user reviews in Croatian. We proposed a simple, semi-automated approach for acquiring product aspects and domain-specific opinion clues. We showed that a supervised model with linguistic features can effectively assign opinions to the individual product aspects. Furthermore, we demonstrated that opinion clues and opinionated aspects improve prediction of overall review polarity and user-assigned opinion rating.

For future work we intend to evaluate our method on other datasets and domains, varying in level of language complexity and correctness. Of particular interest are the domains with aspect-focused ratings and reviews (e.g., electronic product reviews). Aspect-based opinion summarization is another direction for future work.

Acknowledgments

This work has been supported by the Ministry of Science, Education and Sports, Republic of Croatia under the Grant 036-1300646-1986 and Grant 098-0982560-2566.

References

- Željko Agić, Marko Tadić, and Zdravko Dovedan. 2008. Improving part-of-speech tagging accuracy for Croatian by morphological analysis. *Informatika*, 32(4):445–451.
- Željko Agić, Nikola Ljubešić, and Marko Tadić. 2010. Towards sentiment analysis of financial texts in Croatian. In Nicoletta Calzolari, editor, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Željko Agić. 2012. K-best spanning tree dependency parsing with verb valency lexicon reranking. In *Proceedings of 24th international Conference on Computational Linguistics (COLING 2012): Posters*, pages 1–12.
- Erik Boiy and Marie-Francine Moens. 2009. A machine learning approach to sentiment analysis in multilingual web texts. *Information retrieval*, 12(5):526–558.
- Aleksander Buczynski and Aleksander Wawer. 2008. Shallow parsing in sentiment analysis of product reviews. In *Proceedings of the Partial Parsing workshop at LREC*, pages 14–18.
- Iliia Chetviorkin, Pavel Braslavskiy, and Natalia Loukachevich. 2012. Sentiment analysis track at romip 2011. *Dialog*.
- Yoonjung Choi, Youngho Kim, and Sung-Hyon Myaeng. 2009. Domain-specific sentiment analysis using contextual feature generation. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 37–44. ACM.
- Angela Fahrni and Manfred Klenner. 2008. Old wine or warm beer: Target-specific sentiment analysis of adjectives. In *Proc. of the Symposium on Affective Language in Human and Machine, AISB*, pages 60–63.
- Adam Funk, Yaoyong Li, Horacio Saggion, Kalina Bontcheva, and Christian Leibold. 2008. Opinion analysis for business intelligence applications. In *Proceedings of the first international workshop on Ontology-supported business intelligence*, page 3. ACM.
- Goran Glavaš, Jan Šnajder, and Bojana Dalbelo Bašić. 2012. Semi-supervised acquisition of Croatian sentiment lexicon. In *Text, Speech and Dialogue*, pages 166–173. Springer.
- Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, pages 1–12.
- Andrew B Goldberg and Xiaojin Zhu. 2006. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, pages 45–52. Association for Computational Linguistics.
- Valentin Jijkoun, Maarten de Rijke, and Wouter Weerkamp. 2010. Generating focused topic-specific sentiment lexicons. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 585–594. Association for Computational Linguistics.
- Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP '06*, pages 355–363, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Colin Kelly, Barry Devereux, and Anna Korhonen. 2012. Semi-supervised learning for automatic conceptual property extraction. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics, CMCL '12*, pages 11–20, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. 2007. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1065–1074.
- Arjun Mukherjee and Bing Liu. 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 339–348. Association for Computational Linguistics.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.
- Ana-Maria Popescu and Oren Etzioni. 2007. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer.

- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international joint conference on Artificial intelligence*, pages 1199–1204.
- Jasmina Smailović, Miha Grčar, and Martin Žnidaršič. 2012. Sentiment analysis on tweets in a financial domain. In *Proceedings of the 4th Jozef Stefan International Postgraduate School Students Conference*, pages 169–175.
- Pavel Smrž. 2006. Using WordNet for opinion mining. In *Proceedings of the Third International WordNet Conference*, pages 333–335. Masaryk University.
- Jan Šnajder, Bojana Dalbelo Bašić, and Marko Tadić. 2008. Automatic acquisition of inflectional lexica for morphological normalisation. *Information Processing & Management*, 44(5):1720–1731.
- Josef Steinberger, Mohamed Ebrahim, Maud Ehrmann, Ali Hurriyetoglu, Mijail Kabadjov, Polina Lenkova, Ralf Steinberger, Hristo Tanev, Silvia Vázquez, and Vanni Zavarella. 2012. Creating sentiment dictionaries via triangulation. *Decision Support Systems*.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.
- Peter D Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 35(3):399–433.
- Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th conference on Computational linguistics-Volume 2*, pages 947–953. Association for Computational Linguistics.