

ÚFAL: Using Hand-crafted Rules in Aspect Based Sentiment Analysis on Parsed Data

Kateřina Veselovská, Aleř Tamchyna

Charles University in Prague, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské náměstí 25, Prague, Czech Republic

{veselovska,tamchyna}@ufal.mff.cuni.cz

Abstract

This paper describes our submission to SemEval 2014 Task 4¹ (aspect based sentiment analysis). The current work is based on the assumption that it could be advantageous to connect the subtasks into one workflow, not necessarily following their given order. We took part in all four subtasks (aspect term extraction, aspect term polarity, aspect category detection, aspect category polarity), using polarity items detection via various subjectivity lexicons and employing a rule-based system applied on dependency data. To determine aspect categories, we simply look up their WordNet hypernyms. For such a basic method using no machine learning techniques, we consider the results rather satisfactory.

1 Introduction

In a real-life scenario, we usually do not have any golden aspects at our disposal. Therefore, it could be practical to be able to extract both aspects and their polarities at once. So we first parse the data, bearing in mind that it is very difficult to detect both sources/targets and their aspects on plain text corpora. This holds especially for pro-drop languages, e.g. Czech (Veselovská et al., 2014) but the proposed method is still language independent to some extent. Secondly, we detect the polarity items in the parsed text using a union of two different existing subjectivity lexicons (see Section 2). Afterwards, we extract the aspect terms in the dependency structures containing polarity ex-

¹<http://alt.qcri.org/semeval2014/task4/>

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

pressions. In this task, we employ several hand-crafted rules detecting aspects based on syntactic features of the evaluative sentences, inspired by the method by Qiu et al. (2011). Finally, we identify aspect term categories with the help of the English WordNet and derive their polarities based on the polarities of individual aspects. The obtained results are discussed in Section 4.

2 Related Work

This work is related to polarity detection based on a list of evaluative items, i.e. subjectivity lexicons, generally described e.g. in Taboada et al. (2011). The English ones we use are minutely described in Wiebe et al. (2005) and several papers by Bing Liu, starting with Hu and Liu (2004). Inspired by Kobayashi et al. (2007), who make use of evaluative expressions when learning syntactic patterns obtained via pattern mining to extract aspect-evaluation pairs, we use the opinion words to detect evaluative structures in parsed data. The issue of target extraction in sentiment analysis is discussed in articles proposing different methods, mainly tested on product review datasets (Popescu and Etzioni, 2005; Mei et al., 2007; Scaffidi et al., 2007). Some of the authors take into consideration also product aspects (features), defined as product components or product attributes (Liu, 2006). Hu and Liu (2004) take as the feature candidates all noun phrases found in the text. Stoyanov and Cardie (2008) see the problem of target extraction as part of a topic modelling problem, similarly to Mei et al. (2007). In this contribution, we follow the work of Qiu et al. (2011) who learn syntactic relations from dependency trees.

3 Pipeline

Our workflow is illustrated in Figure 1. We first pre-process the data, then mark all aspects seen in the training data (still on plain text). The rest of the pipeline is implemented in Treex (Popel and

	Pattern		Example sentence
Subj _{aspect}	Pred _{copula}	PAdj	The food was great.
Subj _{aspect}	Pred _{copula}	PNoun	The coconut juice is the MUST!
Subj _{aspect}	Pred Adv _{eval}		The pizza tastes so good.
Attr _{eval}	Noun _{aspect}		Nice value.
Subj _{aspect}	Pred _{eval}		Their wine sucks.
Subj _{source}	Pred _{eval}	Obj _{aspect}	I liked the beer selection.

Table 1: Syntactic rules.

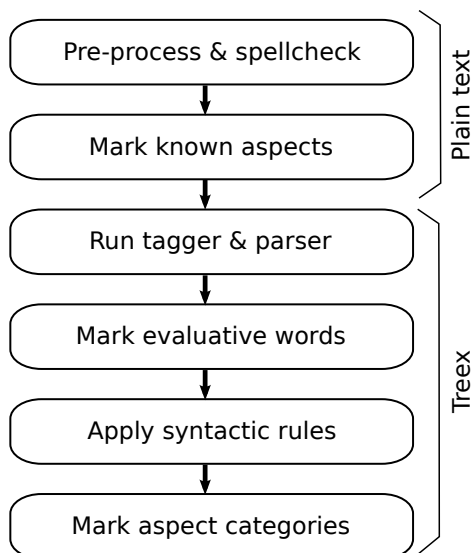


Figure 1: Overall schema of our approach.

Žabokrtský, 2010) and consists of linguistic analysis (tagging, dependency parsing), identification of evaluative words, and application of syntactic rules to find the evaluated aspects. Finally, for restaurants, we also identify aspect categories and their polarity.

3.1 Data

We used the training and trial data provided by the organizers. During system development, we used the trial section as a held-out set. In the final submission, both datasets are utilized in training.

3.2 Pre-processing

The main phase of pre-processing (apart from parsing the input files and other simple tasks) is running a spell-checker. As data for this task comes from real-world reviews, it contains various typos and other small errors. We therefore implemented a statistical spell-checker which works in two stages:

1. Run Aspell² to detect typos and obtain suggestions for them.
2. Select the appropriate suggestions using a language model (LM).

We trained a trigram LM from the English side of CzEng 1.0 (Bojar et al., 2012) using SRILM (Stolcke, 2002). We binarized the LM and use the Lazy decoder (Heafield et al., 2013) for selecting the suggestions that best fit the current context. Our script is freely available for download.³

We created a list of exceptions (domain-specific words, such as “netbook”, are unknown to Aspell’s dictionary) which should not be corrected and also skip named entities in spell-checking.

3.3 Marking Known Aspects

Before any linguistic processing, we mark all words (and multiword expressions) which are marked as aspects in the training data. For our final submission, the list also includes aspects from the provided development sets.

3.4 Morphological Analysis and Parsing

Further, we lemmatize the data and parse it using Treex (Popel and Žabokrtský, 2010), a modular framework for natural language processing (NLP). Treex is focused primarily on dependency syntax and includes blocks (wrappers) for taggers, parsers and other NLP tools. Within Treex, we used the Morče tagger (Hajič et al., 2007) and the MST dependency parser (McDonald et al., 2005).

3.5 Finding Evaluative Words

In the obtained dependency data, we detect polarity items using MPQA subjectivity lexicon (Wiebe et al., 2005) and Bing Liu’s subjectivity clues.⁴

²<http://aspell.net/>

³<https://redmine.ms.mff.cuni.cz/projects/staspell>

⁴<http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>

	Task 1: aspect extraction			Task 2: aspect polarity	Task 3: category detection			Task 4: category polarity
	prec	recall	F-measure	accuracy	prec	recall	F-measure	accuracy
UFAL	0.50	0.72	0.59	0.67	0.57	0.74	0.65	0.63
best	0.91	0.82	0.84	0.81	0.91	0.86	0.88	0.83

Table 2: Results of our system on the Restaurants dataset as evaluated by the task organizers.

	Task 1: aspect extraction			Task 2: aspect polarity
	prec	recall	F-measure	accuracy
UFAL	0.39	0.66	0.49	0.57
best	0.85	0.67	0.75	0.70

Table 3: Results of our system on the Laptops dataset as evaluated by the task organizers.

We lemmatize both lexicons and look first for matching surface forms, then for matching lemmas. (English lemmas as output by Morče are sometimes too coarse, eliminating e.g. negation – we can mostly avoid their matching by looking at surface forms first.)

3.6 Syntactic Rules

Further, we created six basic rules for finding aspects in sentences containing evaluative items from the lexicons, e.g. *“If you find an adjective which is a part of a verbal predicate, the subject of its governing verb should be an aspect.”*, see Table 1. Situational functions are marked with subscript, PAdj and PNoun stand for adjectival and nominal predicative expressions.

Moreover, we applied three more rules concerning coordinations. We suppose that if we find an aspect, every member of a given coordination must be an aspect too.

The excellent mussels, puff pastry, goat cheese and salad.

Concerning but-clauses, we expect that if there is no other aspect in the second part of the sentence, we assign the conflict value to the identified aspect.

The food was pretty good, but a little flavorless.

If there are two aspects identified in the but-coordination, they should be marked with opposite polarity.

The place is cramped, but the food is fantastic!

3.7 Aspect Categories

We collect a list of aspects from the training data and find all their hypernyms in WordNet (Fellbaum, 1998). We hand-craft a list of typical hypernyms for each category (such as “cooking” or “consumption” for the category “food”). Moreover, we look at the most frequent aspects in the training data and add as exceptions those for which our list would fail.

We rely on the output of aspect identification for this subtask. For each aspect marked in the sentence, we look up all its hypernyms in WordNet and compare them to our list. When we find a known hypernym, we assign its category to the aspect. Otherwise, we put the aspect in the “anecdotes/miscellaneous” category. For category polarity assignment, we combine the polarities of all aspects in that category in the following way:

- all positive → positive
- all negative → negative
- all neutral → neutral
- otherwise → conflict

4 Results and Discussion

Table 2 and Table 3 summarize the results of our submission. We do not achieve the best performance in any particular task, our system overall ranked in the middle.

We tend to do better in terms of recall than precision. This effect is mainly caused by our decision to also automatically mark all aspects seen in the training data.

4.1 Effect of the Spell-checker

We evaluated the performance of our system with and without the spell-checker. Overall, the impact

is very small (f-measure stays within 2-decimal rounding error). In some cases its corrections are useful (“convient” → “convenient parking”), sometimes its limited vocabulary harms our system (“fettucino alfredo” → “fitting Alfred”). This issue could be mitigated by providing a custom lexicon to Aspell.

4.2 Sources of Errors

As we always extract aspects that were observed in the training data, our system often marks them in non-evaluative contexts, leading to a considerable number of false positives. However, using this approach improves our f-measure score due to the limited recall of the syntactic rules.

The usefulness of our rules is mainly limited by the (i) sentiment lexicons and (ii) parsing errors.

(i) Since we used the lexicons directly without domain adaptation, many domain-specific terms are missed (“flavorless”, “crowded”) and some are matched incorrectly.

(ii) Parsing errors often confuse the rules and negatively impact both recall and precision. Often, they prevented the system from taking negation into account, so some of the negated polarity items were assigned incorrectly.

The “conflict” polarity value was rarely correct – all aspects and their polarity values need to be correctly discovered to assign this value. However, this type of polarity is infrequent in the data, so the overall impact is small.

Having participated in all four tasks, our system can be readily deployed as a complete solution which covers the whole process from plain text to aspects and aspect categories annotated with polarity. Considering the number of tasks covered and the fact that our system is entirely rule-based, the achieved results seem satisfactory.

5 Conclusion and Future Work

In our work, we developed a purely rule-based system for aspect based sentiment analysis which can both detect aspect terms (and categories) and assign polarity values to them. We have shown that even such a simple approach can achieve relatively good results.

In the future, our main plan is to involve machine learning in our system. We expect that outputs of our rules can serve as useful indicator features for a discriminative learning model, along

with standard features such as bag-of-words (lemmas) or n -grams.

6 Acknowledgements

The research described herein has been supported by the by SVV project number 260 140 and by the LINDAT/CLARIN project funded by the Ministry of Education, Youth and Sports of the Czech Republic, project No. LM2010013.

This work has been using language resources developed and/or stored and/or distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).

References

- Ondřej Bojar, Zdeněk Žabokrtský, Ondřej Dušek, Petra Galuščáková, Martin Majliš, David Mareček, Jiří Maršík, Michal Novák, Martin Popel, and Aleš Tamchyna. 2012. The Joy of Parallelism with CzEng 1.0. In *Proc. of LREC*, pages 3921–3928. ELRA.
- Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.
- Jan Hajič, Jan Votrubec, Pavel Krbeč, Pavel Květoň, et al. 2007. The best of two worlds: Cooperation of statistical and rule-based taggers for czech. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*, pages 67–74.
- Kenneth Heafield, Philipp Koehn, and Alon Lavie. 2013. Grouping language model boundary words to speed k-best extraction from hypergraphs. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 958–968, Atlanta, Georgia, USA, June.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04*, pages 168–177, New York, NY, USA. ACM.
- 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)*.
- Bing Liu. 2006. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

- Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 523–530.
- Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. 2007. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, pages 171–180, New York, NY, USA. ACM.
- Martin Popel and Zdeněk Žabokrtský. 2010. TectoMT: Modular NLP Framework. In Hrafn Loftsson, Eiríkur Rögnvaldsson, and Sigrún Helgadóttir, editors, *IceTAL 2010*, volume 6233 of *Lecture Notes in Computer Science*, pages 293–304. Iceland Centre for Language Technology (ICLT), Springer.
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, pages 339–346, Stroudsburg, PA, USA.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Comput. Linguist.*, 37(1):9–27, March.
- Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin. 2007. Red opal: Product-feature scoring from reviews. In *Proceedings of the 8th ACM Conference on Electronic Commerce, EC '07*, pages 182–191, New York, NY, USA. ACM.
- Andreas Stolcke. 2002. SRILM – An Extensible Language Modeling Toolkit. In *Proc. Intl. Conf. on Spoken Language Processing*, volume 2, pages 901–904.
- Veselin Stoyanov and Claire Cardie. 2008. Topic identification for fine-grained opinion analysis. In *Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1, COLING '08*, pages 817–824, Stroudsburg, PA, USA.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Comput. Linguist.*, 37(2):267–307, June.
- Kateřina Veselovská, Jan Mašek, and Vladislav Kuboň. 2014. Sentiment detection and annotation in a treebank.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210.