OPFI: A Tool for Opinion Finding in Polish

Aleksander Wawer
Institute of Computer Science, Polish Academy of Science
ul. Jana Kazimierza 5, 01-238 Warszawa, Poland
axw@ipipan.waw.pl

Abstract

The paper contains a description of OPFI: Opinion Finder for the Polish Language (Wawer, 2016), a freely available tool for opinion target extraction. The goal of the tool is opinion finding: a task of identifying tuples composed of sentiment (positive or negative) and its target (about what or whom is the sentiment expressed). OPFI is not dependent on any particular method of sentiment identification and provides a built-in sentiment dictionary as a convenient option. Technically, it contains implementations of three different modes of opinion tuple generation: one hybrid based on dependency parsing and CRF, the second based on shallow parsing and the third on deep learning, namely GRU neural network. The paper also contains a description of related language resources: two annotated treebanks and one set of tweets.

Keywords: sentiment analysis, opinion targets, opinion tuples, opinion mining

The paper is a description of OPFI: Opinion Finder for the Polish Language (Wawer, 2016), a freely available tool for opinion target extraction. It is organized as follows: Section 1. is an introduction to the problem of opinion target extraction. Section sec:Overview contains a brief description of opinion target identification methods, bundled in OPFI. Section 4. is a technical overview of OPFI and a discussion of input formats and involved tools. Section 5. provides a brief evaluation of selected OPFI algorithms and models. Finally, Section 6. concludes the paper.

1. Introduction

Opinion extraction is about identifying and categorizing opinions expressed in a piece of text (as for example, sentiments expressed in a word, phrase, sentence or document). Opinions are expressed about topics or objects, as for example positive or negative opinions about a product. Typically, opinion extraction consists of multiple sub-tasks, such as identifying opinions and their polarities, often called sentiment analysis, followed by finding their targets (topics or objects of opinions).

The goal of the latter sub-task, called opinion target extraction, consists of identifying words towards which an opinion (sentiment) is expressed. Their character, meaning and syntactic role can vary and is influenced by corpus type. In the domain of product reviews, opinion targets are usually aspect terms, sometimes called attributes, denoting properties of evaluated entity (eg.: This camera has great zoom but poor battery). It is also the case that it is entity itself that is the target of an opinion (eg.: I hate this camera). In other types of texts, opinion targets are more varied. They might be syntactically expressed not only as nouns and noun phrases, but as verbs referring to various states and activities (eg.: I do not like swimming).

The formulation of opinion target extraction problem that we follow is similar to (Qiu et al., 2011), where authors also consider multiple types of opinion targets, including entities and their aspects. To extract opinion targets, they apply two simple dependency patterns that are matched against sentiment words. The patterns are used in an iterative algorithm that uses dependency parse information and seed lexicons, to discover sentiment and opinion target vocabulary in iterative fashion.

A similar task to opinion target extraction, but less broad, is aspect-based opinion mining (Pontiki et al., 2015). It aims to address the shortcomings of message-level (tweet-level or sentence-level) opinion classification, where only one major sentiment value is assigned to a text, ignoring the possibility of reverse polarities (as for example in: great zoom but poor battery).

Known approaches to opinion target extraction include not only syntactic pattern methods such as in (Qiu et al., 2011; Poria et al., 2014; Gindl et al., 2013), but also sequence-labeling algorithms, such as Conditional Random Fields (CRF) (Jakob and Gurevych, 2010).

2. Overview

The problem of identification opinion targets in OPFI is solved by analyzing syntactic links between sentiment expressions and candidates for opinion targets, extracted from syntactic structure. Identification of opinion targets is therefore strongly linked to prior syntactic parsing and sentiment identification in a sentence. However, OPFI has been designed to be independent from these steps, by supporting two broad parsing classes: shallow and dependency, and any parser within these two classes.

Our approach, implemented in the application OPFI, is based on three methods, of choice to the user.

The first method is an approach sketched in (Wawer, 2015b). In this method, opinion target identification is performed using two steps. First, a set of dependency patterns is applied to identify possible opinion target candidates. Second, the list is processed by a Conditional Random Field (CRF) tagger (Lafferty et al., 2001). The feature space is designed with domain-independency in mind, and consists only of part-of-speech features (not lexical).

The second method is an adaptation of recently more and more popular paradigm of deep learning. Namely, it is based on a GRU neural network described by (Cho et al., 2014) and implemented in Keras neural networking toolbox (`http://keras.io`). The GRU model uses a very different feature space from the one used in the first hybrid

solution.. The GRU network modelis trained on word2vec (Mikolov et al., 2013b; Mikolov et al., 2013a) word embedding vectors. The set of vectors trained on Polish Wikipedia and National Corpus of Polish, used in tool, may be also downloaded from the OPFI home page.

The third method is based on shallow parsing. In this case, opinion target identification is done using a set of shallow grammatical structures known to be potentially good indicators for opinion targets. The selection of usable patterns was based on (Wawer, 2015a).

3. Datasets

3.1. Description

In this section we present detailed description of corpora and treebanks used to train and evaluate OPFI, containing annotations of opinions, opinion targets and relations between the two types of elements. All three corpora were annotated manually using BRAT annotation framework (Stenetorp et al., 2011).

Reviews: 1000 sentences from product review corpus. This corpus was compiled from reviews of two types of products, perfumes and clothes. The sentences were selected in a semi-random fashion, containing sentiment words and opinion target words (for this specific corpus, a dictionary of product aspects was previously available). The sentences were then parsed using the MaltEval parser model for the Polish language (Wróblewska and Woliński, 2011). This dataset provides information about correctness of dependency parse tree. For training and evaluating OPFI we selected only sentences with dependency structure verified by linguists as correct (without serious errors).

Twitter: 500 tweets random-selected from the database gathered for the TrendMiner project (www.trendminer.eu). The tweets were collected over a period of 6 months, from feeds related to journalism and political sphere.

Treebank: 1000 sentences from Składnica - a treebank of Polish (Wróblewska, 2012; Wróblewska and Woliński, 2011). This treebank is a result of parsing 20000 Polish sentences with the syntactic parser Świgra. For every sentence, the parser generated all possible syntactic parse trees predicted by the rules of its grammar. Then, linguists selected one correct parse tree for each sentence. This resulted in over 8000 sentences with correct constituency structure. For our experiments with opinion targets, we used a version of Składnica converted automatically to dependency structure. Finally, we identified a subset of 2000 sentences with known sentiment words (using the dictionary) and then, random-selected a half of this subset for opinion target annotation.The problem of parse correctness does not appear in this dataset, as all these sentences are from manually disambiguated Składnica treebank.

The resources are available to download from the OPFI home page (`http://zil.ipipan.waw.pl/OPTA`)

3.2. Annotation Quality

We verified annotation quality by double annotation of a random subset of the Reviews corpus. Results presented in Table 1 demonstrate generally high levels of agreement, relatively the lowest for relation between sentiments (S) and targets (T). The analysis of reasons of behind difficulties in annotating relations between S and T demonstrated a number of problematic issues, where the relation is weak or indirect. For example, in: "I like(S) this perfume(T)'s bottle", the relation between perfume (target) and like (sentiment) is indirect, and it is arguable what target should be selected. Both syntactic and semantic criteria could be envisioned, such as always selecting noun phrase's heads, where applicable.

	Total	Agreed	Agreement
correctness of Targets	75	64	85%
correctness of Sentiments	75	70	93%
S related to T	54	42	77%

Table 1: Inter-annotator agreement.

4. Technical Aspects

The tool has been implemented in Python. The input is:

- For the method that uses CRF and dependency patterns, an extended CONLL format; the last column indicates words sentiment. OPFI comes bundled with machine learning models (such as CRF), used internally to raise the precision over syntactic patterns. Technically, this approach is based on CRFsuite tool (Okazaki, 2007) and lbfgs algorithm to train the models. The objective of CRF is to extract all targets of opinions from the dataset, using several groups of features, mostly syntactic.
- For the method based on shallow parser, a JSON or SOAP format defined according to multiservice specification (Ogrodniczuk and Lenart, 2012). Technically, this method is implemented as a set of scripts for post-processing shallow parser output using a sentiment dictionary.
- For the GRU based method, plain text format without any lemmatization or pre-processing. In this case, it is assumed that all relevant information are encoded in word2vec vectors available for all meaningful orthographic word forms.

OPFI is bundled with a default sentiment dictionary, a copy of the dictionary available from `http://zil.ipipan.waw.pl/SlownikWydzwieku`. However, any method of sentiment identification can be used with the tool. For instance, a user could alternatively use any available resource or algorithm of phrase-level or word-level sentiment identification such as CRF-based or perhaps different sentiment dictionary.

5. Evaluation

Table 2 contains two performance quality metrices: precision and recall, computed for dataset type indicated in the last column.

Method	Precision	Recall	Dataset
shallow	0.62	0.47	twitter
	0.59	0.47	reviews
dependency+CRF	0.32	0.27	twitter
	0.82	>0.73*	reviews
GRU neural net	0.1	0.30	treebank

Table 2: Evaluation of OPFI methods.

In general, the above table should be treated with caution, as the training and evaluation methods are not fully comparable.

In the case of hybrid dependency patterns+CRF method, CRF part was evaluated in a 10-fold cross-validation scenario on the review dataset. However, however, for dependency pattern induction, we used all of the review dataset at once. In our view, this is reflected in possibly too optimistic reporting of recall. The precision measurement, which is driven mostly by CRF part of the hybrid, should not become too much affected.

For the evaluation of shallow method, we measured its error on all sentences and tweets from each dataset. No cross-validation was necessary due to the fact that this method is heurisitc-based and does not involve machine learning of any kind.

The GRU network was trained on the largest dataset, reviews, but tested on Składnica treebank. Such scenario tests mostly domain independence, and as it turned out, it did not avoid the issue of domain-dependency (becoming overly attached to domain-specific lexicons).

6. Conclusions

Opinion target identification fills a gap between fine-grained sentiment analysis (which typically means word-level and phrase-level sentiment recognition) and information extraction, by producing tuples of opinions and their targets.

This paper describes a tool (OPFI) for opinion target identification in the Polish language. OPFI was trained and evaluated on three different datasets (corpora or treebanks). It is highly versatile as it supports usage with any sentiment extraction method. It includes three different techniques (algoritms): the first one hybrid, based on dependency patterns and CRF, the second based on shallow grammar rules. One can use any method depending on type of input texts and expected quality of parsing. The third approach in OPFI is based on a GRU neural network.

Generally, for processing twitter data one can recommend shallow method as the most error tolerant. It achieves reasonable precision and good recall. For processing more clean texts and longer sentences such as in reviews, the method of choice becomes the one based on dependency patterns and CRF. This is due to more clean texts where the benefits of dependency parsing can become apparent. The GRU method is highly experimental at this point and will be the subject of further fine-tuning.

OPFI can be used for large scale processing of Polish texts, such as tweets or product reviews, to seek for opinions expressed about politicians, products, or their aspects, in order to aggregate them into meaningful knowledge.

7. Acknowledgements

This work was funded by the National Science Centre of Poland grant nr UMO-2012/05/N/ST6/03587

8. Bibliographical References

Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259.

Gindl, S., Weichselbraun, A., and Scharl, A. (2013). Rule-based opinion target and aspect extraction to acquire affective knowledge. In *Proceedings of WWW'13 workshop on Multidisciplinary Approaches to Big Social Data Analysis*.

Jakob, N. and Gurevych, I. (2010). Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 1035–1045, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *Proceedings of Workshop at ICLR*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Proceedings of NIPS*.

Ogrodniczuk, M. and Lenart, M. (2012). Web Service integration platform for Polish linguistic resources. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012*, pages 1164–1168. ELRA.

Okazaki, N. (2007). Crfsuite: a fast implementation of conditional random fields (crfs).

Pontiki, M., Galanis, D., Papageogiou, H., Manandhar, S., and Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, SemEval 2015.

Poria, S., Cambria, E., Ku, L.-W., Gui, C., and Gelbukh, A. (2014). A rule-based approach to aspect extraction from product reviews. In *Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP)*, August.

Qiu, G., Liu, B., Bu, J., and Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27, March.

Stenetorp, P., Topić, G., Pyysalo, S., Ohta, T., Kim, J.-D., and Tsujii, J. (2011). Bionlp shared task 2011: Supporting resources. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 112–120, Portland, Oregon, USA, June. Association for Computational Linguistics.

Wawer, A. (2015a). Comparing shallow and dependency syntactic analysis for opinion target extraction. In Zygmunt Vetulani et al., editors, *Proceedings of the 7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*. Poznań, Poland.

Wawer, A. (2015b). Towards domain-independent opinion target extraction. In *2015 IEEE 15th International Conference on Data Mining Workshops (SENTIRE 2015)*, Los Alamitos, CA, USA. IEEE Computer Society.

Wróblewska, A. and Woliński, M. (2011). Preliminary experiments in Polish dependency parsing. volume 7053 of *Lecture Notes in Computer Science*, pages 279–292. Springer-Verlag.

Wróblewska, A. (2012). Polish dependency bank. *Linguistic Issues in Language Technology*, 7(1).

9. Language Resource References

Aleksander Wawer. (2016). *OPFI: Opinion Finder*. Institute of Computer Science, Polish Academy of Sciences, http://zil.ipipan.waw.pl/OPTA, 1.0.