Towards Detecting Counter-considerations in Text

Andreas Peldszus
Applied Computational Linguistics
FSP Cognitive Science
University of Potsdam
peldszus@uni-potsdam.de

Manfred Stede
Applied Computational Linguistics
FSP Cognitive Science
University of Potsdam
stede@uni-potsdam.de

Abstract

Argumentation mining obviously involves finding support relations between statements, but many interesting instances of argumentation also contain counter-considerations, which the author mentions in order to preempt possible objections by the readers. A counter-consideration in monologue text thus involves a switch of perspective toward an imaginary opponent. We present a classification approach to classifying counter-considerations and apply it to two different corpora: a selection of very short argumentative texts produced in a text generation experiment, and a set of newspaper commentaries. As expected, the latter pose more difficulties, which we investigate in a brief error anaylsis.

1 Introduction

The exchange of argument and objection is obviously most typical for dialogue, but to a good extent it is also present in monologue text: Authors do not only provide justifications for their own position – they can also mention potential objections and then refute or outweigh them. In this way they demonstrate to have considered the position of "the other side", which altogether is designed to reinforce their own position. We use the term 'counter-consideration' in a general sense to cover all such moves of an author, no matter whether they are directed at the conclusion of the text or at an intermediate argument, or at some support relation, and irrespective of whether they are explicitly refuted by the author or merely mentioned and left outweighed by the mass of arguments in favour of the main claim.[1]

For an author, presenting a counter-consideration involves a switch of perspective by temporarily adopting the opposing viewpoint and then moving back to one's own. This is a move that generally requires some form of explicit linguistic marking so that the reader can follow the line of argumentation. The kinds of marking include explicit belief attribution followed by a contrastive connective signaling the return ("Some people think that X. However, this ..."), and there can also be quite compact mentions of objections, as in "Even though the project is expensive, we need to pursue it, because..."

Detecting counter-considerations is thus a subtask of argumentation mining. It involves identifying the two points of perspective switching, which we henceforth call a move from the *proponent role* to the *opponent role* and back. Thus the task can be operationalized as labelling segments of argumentative text in terms of these two roles. Then, counter-considerations are segments labeled as "opponent".

We study this classification problem using two different corpora: a colletion of user-generated short "microtexts", where we expect the task to be relatively easy, and a set of argumentative newspaper pieces that explicitly argue in favour of or against a particular position ('ProCon'). These texts are longer and more complex, and the opponent role can be encoded in quite subtle ways, so that we expect

[1] Govier (2011) discusses the role of such counter-considerations in 'pro and con' argumentation in more depth. Also, for a comprehensive overview of different notions of objections in argument analysis, see Walton (2009).

Proceedings of the 2nd Workshop on Argumentation Mining, pages 104–109,
Denver, Colorado, June 4, 2015.

the classification to be more difficult.

After looking at related work, Section 3 describes our corpora and the machine learning experiments. In Section 4, we evaluate the results and discuss the most common problems with the ProCon texts, and Section 5 concludes.

2 Related work

The majority of work on text-oriented argumentation mining concentrates on identifying just the "gist" of arguments, i.e., premises and conclusions. This holds, for example, for the well-known early approach of Mochales Palau and Moens (2009), and for the follow-up step on scheme classification (on top of detected premises/conclusions) by Feng and Hirst (2011).

Among the few approaches that do consider counter-considerations, Kang and Saint-Dizier (2014) analyze technical douments (largely instructional text), where the notion of exception to an argument plays a role, but its function is quite different from the perspective-switching that we discuss here.

Ong et al. (2014) work on student essays, which are somewhat more similar to "our" genres. Their task includes the recognition of sentence types (CurrentStudy, Hypothesis, Claim, Citation) and of *support* and *oppose* relations between sentences. For the complete task, the authors use eight hand-coded rules performing string matching using word lists and numbers (for identifying the year of a citation); thus the approach is geared toward finding relationships specifically between citations and will not generalize well to the broad class of counter-considerations.

A support/oppose distinction is also made by Stab and Gurevych (2014), who annotated a corpus of 90 essays (1673 sentences) with the central claim of the text (90 instances), claims of paragraph-size units (429), and premises (1033). Claims are marked with an attribute 'for' (365) or 'against' (64), but the authors do not report numbers on the stance of premises. Note however, that the stance of premises could be inferred by the relation structure, i.e. the sequence of supposing and opposing relations. Of the 1473 relations in the corpus, 161 are opposing. As the proportion of 'against' claims is also relatively low, the authors restrict their classification task, again, to the 'for' claims and the support relations.

Looking beyond the argumentation mining literature, elaborate approaches to subjectivity analysis are also relevant to us, as found in the *appraisal theory* of Martin and White (2005), whose multidimensional analysis also covers a speaker's consideration of conflicting standpoints. Appraisal is a very comprehensive scheme that is difficult to annotate (Read and Carroll, 2012a); thus its automatic classification is hard, as experiments by Read and Carroll (2012b) show. Our smaller task of role identification addressed here can be considered a subproblem of appraisal analysis.

3 Classification study

3.1 Corpora

As stated earlier, we worked with two different corpora in order to study the difference in task difficulty for short and simple "user-generated" texts versus newspaper articles.

The "argumentative microtext" corpus (Peldszus and Stede, 2015) is a new, freely available collection of 112 very short texts that were collected from human subjects, originally in German. Subjects received a prompt on an issue of public debate, usually in the form of a yes/no question (e.g., "Should shopping malls be open on Sundays?"), and they were asked to provide their answer to the question along with arguments in support. They were encouraged to also mention potential counter-considerations. The target length suggested to the subjects was five sentences. After the texts were collected, they were professionally translated to English, so that the corpus is now available in two languages. An example of an English text is:

> Health insurance companies should naturally cover alternative medical treatments. Not all practices and approaches that are lumped together under this term may have been proven in clinical trials, yet it's precisely their positive effect when accompanying conventional 'western' medical therapies that's been demonstrated as beneficial. Besides, many general practitioners offer such counselling and treatments in parallel anyway - and who would want to question their broad expertise?

The annotation of argumentation structure (common to both language versions) follows the scheme outlined in Peldszus and Stede (2013), which in turn is based on the work of Freeman (1991), and it includes different types of support and attack relations. The argumentative role per segment can be inferred from the relational structure. 21.7% of the 576 individual discourse segments bear the opponent role. As reported in Peldszus (2014), naive and untrained annotators reached an agreement of κ=.52 in distinguishing proponent and opponent on a subset of the corpus, while expert annotators achieved perfect agreement.

The ProCon corpus consists of 124 texts taken from a "pro and contra" column of the German newspaper *Der Tagesspiegel*. The setting for the content is essentially the same: A "should we do X or not / Is X good or bad / ..." question on an issue of public interest. The texts, however, are written by journalists, and a pro and a contra article appear next to each other in the paper (but they don't refer to each other). Typically they are 10-12 sentences long. While the microtexts are manually segmented, we use an automatic segmentation module for German to split the ProCon texts. This is a statistical system trained on a similar corpus, which aims at identifying clause-size segments on the output of a dependency parser (Bohnet, 2010). Segmentation leads to 2074 segments, which have then been annotated with the proponent/opponent label by two expert annotators. 8.3% of the individual 2074 segments bear the opponent role. Agreement between these experts had been tested on 24 manually segmented ProCon texts and resulted in κ=.74. Table 1a summarizes the corpus statistics.

To get a clearer picture of the distribution of opponent segments, we study their frequency and position in the individual texts: Table 1b shows the number of texts by the number (n) of included opponent segments, and Table 1c gives the percentage of opponent segments occurring in the first to fifth chunk of the text. While there is clear tendency for opponent segments to appear in the opening of a ProCon text, they are more equally spread in the microtexts.

3.2 Experiments

Feature sets We compare three different feature sets: two simple bag-of-word models as baselines and one model with additional features from automatic linguistic analysis. The first model (B) only extracts binary features for each lemma occurring in the target segment. The second model (B+C) additionally extracts these features from the preceding and the subsequent segment, thus providing a small context window. The full model (B+C+L) adds parsing-based features for the whole context window, such as pos-tags, lemma- and pos-tag-based dependency-parse triples, the morphology of the main verb (Bohnet, 2010), as well as lemma-bigrams. Discourse connectives are taken from a list by Stede (2002) and used both as individual items and as indicating a coherence relation (Cause, Contrast, etc.). Furthermore, we use some positional statistics such as relative segment position, segment length, and punctuation count.

Approach The goal is to assign the labels 'proponent' and 'opponent' to the individual segments. We trained a linear log-loss model using stochastic gradient descent learning as implemented in the Scikit learn library (Pedregosa et al., 2011). The learning rate is set to optimal decrease, and the class weights are adjusted according to class distribution. We used a nested 5x3 cross validation (CV), with the inner CV for tuning the hyper parameters (the regularization parameter alpha and the number of best features to select) and the outer CV for evaluation. We optimize macro averaged F1-score. The folding is stratified, randomly distributing the texts of the corpus while aiming to reproduce the overall label distribution in both training and test set.

All results are reported as average and standard deviation over the 50 folds resulting from 10 iterations of 5-fold cross validation. We use the following metrics: Cohen's Kappa κ, Macro average F1, Precision, Recall and F1 for the opponent class.

Results The performance of the classifiers is shown in Table 2.[2] Comparing the results for the two datasets confirms our assumption that the task is much harder on the ProCon texts. When comparing the different models, we observe that the simple baseline model without context performs poorly; adding context improves the results significantly.

[2]Similar results for an earlier version of the microtext corpus for this and other argumentation mining tasks have been presented in Peldszus (2014).

	microtexts	ProCon
texts	112	124
segments	576	2074
segments (proponent)	451	1902
segments (opponent)	125	172
segments per text	5.1±0.8	16.9±3.1
opp. seg. per text	1.1±0.7	1.4±1.5

(a) general statistics (averages with std. dev.)

n	microtexts	ProCon
0	15	46
1	74	32
2	18	16
3	5	17
4		6
5		3
6		3

(b) opponent frequency

p	microtexts	ProCon
1/5	16.0%	35.5%
2/5	23.2%	18.6%
3/5	17.6%	19.1%
4/5	28.8%	12.8%
5/5	14.4%	11.6%

(c) opponent position

Table 1: Corpus statistics: For details see Section 3.1.

The full featureset (B+C+L) always yields best results, except for a small drop of precision on the ProCon texts. The improvement of the full model over B+C is significant for the microtexts ($p < 0.003$ for κ, F1 macro and opponent F1, using Wilcoxon signed-rank test over the 50 folds), but not significant for the ProCon texts.

Feature selection mostly supports the classification of the ProCon texts, where the mass of extracted features impairs the generalization. Typically only 25 features were chosen. For the microtexts, reducing the features to the 50 best-performing ones still yields good but not the best results. One reason for the difference in feature selection behaviour between the datasets might be that the proportion of proponent and opponent labels is more skewed for the ProCons than for the microtexts. Another reason might be the richer set of expressions marking the role switch in the ProCon texts.

A common observation for both corpora is that the connective *aber* (‘but’) in the subsequent segment is the best predictor for an opponent role. Other important lexical items (also as part of dependency triples) are the modal particles *natürlich* (‘of course’, ‘naturally’) and *ja* (here in the reading: ‘as is well-known’), and the auxiliary verb *mögen* (here: ‘may’). All of these occur in the opponent role segment itself, and they have in common that they “color” a statement as something that the author concedes (but will overturn in the next step), which corresponds to the temporary change of perspective. As for differences between the corpora, we find that the connective *zwar*, which introduces a concessive minor clause, is very important in the microtexts but less prominent in ProCon. We attribute this to the microtext instruction of writing rather short texts, which supposedly leads the students to often formulating their counter-considerations as compact minor clauses, for which *zwar* (‘granted that’) is the perfect marker. Presumably for the same reason, we observe that the concessive subordinator *obwohl* (‘although’) is among the top-10 features for microtexts but not even among the top-50 for ProCon. In ProCon, the group of connectives indicating the Contrast coherence relation is a very good feature, and it is absent from the microtext top-50; recall, though, that the single connective *aber* (‘but’) is their strongest predictor, and the very similar *doch* is also highly predictive.

4 Discussion and error analysis

Proponent/Opponent role identification is not an easy classification task. For the microtexts, we regard the results as fairly satisfactory. For ProCon, there is a significant drop in F1 macro, and even more so for the opponent prec/rec/F1. This was in principle to be expected, but we wanted to know reasons and thus performed a qualitative error analysis.

Segmentation. As pointed out, ProCon texts have been automatically segmented, which leads to a number of errors that generate some of the classification problems; we found, however, that this is only a small factor.

There are other points to remark on segmentation, though. First, we find 37 cases where more than one opponent role segment appear in a sequence (mostly two of them, but ranging up to six), as compared to 68 cases of individual segments. The sequences pose problems for segment-wise classification focusing on perspective *change* signals, especially when the context window is small. Many of

	microtexts			ProCon		
	B	B+C	B+C+R	B	B+C	B+C+R
κ	.375±.109	.503±.080	.545±.098	.187±.064	.320±.078	.323±.091
F1 macro	.685±.056	.751±.040	.772±.049	.588±.033	.659±.040	.660±.047
opponent P.	.548±.097	.647±.081	.668±.096	.428±.165	.370±.063	.361±.074
opponent R.	.474±.146	.575±.084	.626±.108	.163±.054	.400±.109	.422±.117
opponent F1	.497±.101	.604±.065	.640±.081	.225±.064	.378±.073	.382±.083

Table 2: Results for role-identification, reported as average and standard deviation

the sequences occur right at the beginning of the text, where the author provides an extended description from the opponent's view, and then switches to his own perspective. Correctly identifying complete sequences would require a deeper analysis of cohesive devices for finding continuation or break of topic/perspective/argumentative orientation.

Also, notice that many of the sequences actually contain argumentative sub-structure, where, for example, the possible objection is first backed up with purported evidence and then refuted.

Here, the question of segmentation grain-size arises. In the present annotation, we do not label segments as 'opponent role' when they include not only the opponent's objection but also the author's refutation or dismissal. This is because on the whole, the segment conveys the author's (proponent's) position. A translated example from the corpus is: "Not convincing at all is the argument that to the government, teachers should be worth more than a one-Euro-job." Besides such cases of explicit dismissal, we find, for instance, concessive PPs that include an opposing argument: "Despite the high cost, the building must be constructed now." We leave it to future work to dissect such complex segments and split them into an opponent and a proponent part.

Connectives. Contrastive connectives are very good indicators for changing back from the opponent role to the proponent role, but unfortunately they occur quite frequently also with other functions. There are 105 opponent segments or sequences thereof in the corpus, but 195 instances of the words *aber* and *doch*, which are the most frequent contrastive connectives. Therefore, their presence needs to be correlated with other features in order to serve as reliable indicators.

Language. While our focus in this paper was on the performance difference between the German microtexts and the ProCon texts, we want to mention that the overall classification results for microtexts do hardly differ between the German and the English version. This leads us to expect that for English pro/contra commentaries, we would also obtain results similar to those for German.

5 Conclusion

Counter-considerations may be regarded as not the most important aspects of an argumentation, but in many essayistic text genres, they constitute rhetorical moves that authors quite frequently advance to strengthen their points. After all, refuting a potential objection is in itself an argument in support of the conclusion. Almost two thirds of the newspaper pro/contra texts in our corpus have counter-considerations, and so we think these devices are definitely worth studying in order to arrive at complete argumentation analyses.

Casting the problem as a segment classification task, we obtained good results on our corpus of microtexts, whereas we see room for improvement for the longer and more complex pro/contra newspaper texts. Our error analysis identified several directions for future work, which will also include testing a sequence labelling approach to see whether the regularities in signalling perspective changes can be captured more easily, especially for the many cases of contiguous sequences of opponent role segments.

Acknowledgments

We are grateful to the anonymous reviewers for their thoughtful comments and suggestions on improving the paper. The first author was supported by a grant from Cusanuswerk.

References

Bernd Bohnet. 2010. Very high accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 89–97, Stroudsburg, PA, USA. Association for Computational Linguistics.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, pages 987–996, Stroudsburg, PA, USA. Association for Computational Linguistics.

James B. Freeman. 1991. *Dialectics and the Macrostructure of Argument*. Foris, Berlin.

Trudy Govier. 2011. More on counter-considerations. In *Proceedings of International Conference of the Ontario Society for the Study of Argumentation (OSSA)*, pages 1–10, Windsor/Ontario.

Juyeon Kang and Patrick Saint-Dizier. 2014. A discourse grammar for processing arguments in context. In *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 43–50. IOS Press.

James R. Martin and Peter R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan, Houndsmills/New York.

Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Law (ICAIL 2009), Barcelona, Spain*, pages 98–109. ACM.

Nathan Ong, Diane Litman, and Alexandra Brusilovsky. 2014. Ontology-based argument mining and automatic essay scoring. In *Proceedings of the First Workshop on Argumentation Mining*, pages 24–28, Baltimore, Maryland, June. Association for Computational Linguistics.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Andreas Peldszus and Manfred Stede. 2013. From argument diagrams to automatic argument mining: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31.

Andreas Peldszus and Manfred Stede. 2015. An annotated corpus of argumentative microtexts. In *Proceedings of the First Conference on Argumentation*, Lisbon, Portugal, June. to appear.

Andreas Peldszus. 2014. Towards segment-based recognition of argumentation structure in short texts. In *Proceedings of the First Workshop on Argumentation Mining*, Baltimore, U.S., June. Association for Computational Linguistics.

Jonathon Read and John Carroll. 2012a. Annotating expressions of appraisal in english. *Language Resources and Evaluation*, 421–447(3).

Jonathon Read and John Carroll. 2012b. Weakly-supervised appraisal analysis. *Linguistic Issues in Language Technology*, 8(2).

Christian Stab and Iryna Gurevych. 2014. Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1501–1510. Dublin City University and Association for Computational Linguistics.

Manfred Stede. 2002. DiMLex: A Lexical Approach to Discourse Markers. In Vittorio Di Tomaso Alessandro Lenci, editor, *Exploring the Lexicon - Theory and Computation*. Edizioni dell'Orso, Alessandria, Italy.

Douglas Walton. 2009. Objections, rebuttals and refutations. In *Proceedings of International Conference of the Ontario Society for the Study of Argumentation (OSSA)*, pages 1–10, Windsor/Ontario.