Supplementary material for the EMNLP 2017 article: "What is the Essence of a Claim? Cross-Domain Claim Identification"

Johannes Daxenberger[†], **Steffen Eger**[†‡], **Ivan Habernal**[†], **Christian Stab**[†], **Iryna Gurevych**[†‡]
[†]Ubiquitous Knowledge Processing Lab (UKP-TUDA)
Department of Computer Science, Technische Universität Darmstadt
[‡]Ubiquitous Knowledge Processing Lab (UKP-DIPF)
German Institute for Educational Research and Educational Information
http://www.ukp.tu-darmstadt.de

1 Experimental Setup: Detailed Description of Hand-Crafted Features

This part of the supplementary material describes the hand-crafted features we used in more detail.

Structure Features: Structure features capture the position, the length and the punctuation of a sentence. First, we define two binary features which indicate if the current sentence is the first or last sentence in the paragraph in which it is contained. These feature are motivated by the findings of Stab and Gurevych (2017) who found that structural properties of argument components are effective for distinguishing the argumentative function of argument components. In addition, Peldszus and Stede (2016) found that 43% of claims appear in the first sentence. Second, we add the number of tokens of the sentence to our feature set which proved to be indicative for identifying argumentatively relevant sentences (Biran and Rambow, 2011; Moens et al., 2007). Finally, we adopted the punctuation features from Mochales-Palau and Moens (2009).

Lexical Features: As lexical features, we employ lowercased unigrams. We assume that these features are helpful for detecting claims since they capture discourse connectors like "*therefore*", "*thus*", or "*hence*" which frequently signal the presence of claims. We consider the most frequent 4,000 unigrams as binary features.

Syntactic Features: To account for grammatical information at the sentence level, we include information about the part-of-speech and parse tree for each sentence. Following Stab and Gurevych (2017), we add binary POS n-grams (the 2000 most frequent, $2 \leq n \leq 4$) and constituent parse tree production rules (4000 most frequent, minimum occurrence 5) as originally suggested by Lin et al. (2009). Additionally, to account for the frequency of POS tags, we include a feature which counts the occurrence of each part-of-speech per sentence.

Discourse Features: Cabrio et al. (2013) suggested that the relation between parts of discourse (e.g. connectives such as "because") can be helpful to determine argumentative content. As this finding is affirmed by Stab and Gurevych (2017), we include discourse features with the help of the Penn Discourse Treebank (PDTB) styled end-to-end discourse parser as presented by Lin et al. (2014). We include discourse relations extracted from the parser output as a triple of i) the type of relation, ii) whether the relation is implicit or explicit, and iii) whether the current sentence is part of the first or the second discourse argument (or both).

Embedding Features: We represent each sentence as a summation of its word embeddings (Guo et al., 2014). These simple yet powerful latent semantic representation features have been found predictive in related work (Habernal and Gurevych, 2015, 2017). In particular, we use pretrained 300 dimensional GoogleNews word embeddings.[1]

2 Cross-Domain Experiments: Full Results

Table 1 displays the results of cross-domain experiments, for all source domains. Precisely, we list results for the six best in-domain systems, according to average F_1 scores.

[1] https://code.google.com/archive/p/word2vec/

Train ↓ Test →	**MT**		**OC**		**PE**		**VG**		**WD**		**WTP**		**Avg**	
							LR+Lexical						*55.9*	*27.2*
MT	–	–	53.6	16.5	55.3	28.3	53.4	25.3	54.8	14.5	53.3	20.8	54.1	21.1
OC	62.3	46.4	–	–	57.3	48.0	59.8	39.2	54.2	12.0	57.3	26.5	58.2	34.4
PE	61.5	46.5	54.6	17.8	–	–	53.8	32.9	53.1	11.5	54.7	24.2	55.5	26.6
VG	62.7	47.0	57.4	20.4	57.2	47.4	–	–	50.2	10.0	55.6	25.4	56.6	30.0
WD	56.7	24.1	54.3	17.5	55.0	22.8	51.4	12.7	–	–	53.5	20.2	54.2	19.4
WTP	61.8	46.8	56.3	19.1	56.4	46.4	55.3	35.2	53.2	11.7	–	–	56.6	31.9
							LR-Embeddings						*56.1*	*28.0*
MT	–	–	54.3	17.4	52.0	30.0	56.3	34.5	55.1	14.5	52.6	21.4	54.0	23.5
OC	58.4	43.8	–	–	56.7	46.9	59.0	38.4	54.3	12.4	57.3	27.2	57.1	33.7
PE	58.6	37.0	55.0	18.2	–	–	53.8	20.9	53.6	13.0	54.5	21.0	55.1	22.0
VG	64.5	49.8	57.1	21.6	57.0	45.2	–	–	54.3	13.0	55.3	25.1	57.7	31.0
WD	63.3	41.5	55.7	19.5	55.9	31.5	55.0	23.6	–	–	53.7	21.2	56.7	27.5
WTP	57.7	41.6	56.0	19.9	56.2	42.5	57.2	35.8	52.8	11.6	–	–	56.0	30.3
							LR-Structure						*56.0*	*27.8*
MT	–	–	52.7	15.6	51.3	29.9	56.2	34.7	55.6	15.0	51.7	20.6	53.5	23.2
OC	59.0	44.2	–	–	56.3	46.7	58.8	38.3	54.2	12.3	57.7	27.5	57.2	33.8
PE	57.5	35.3	54.8	17.7	–	–	54.0	21.1	53.7	13.2	54.3	20.3	54.9	21.5
VG	65.6	51.3	57.0	21.3	56.8	44.9	–	–	54.5	13.2	55.1	24.8	57.8	31.1
WD	62.8	39.1	55.5	19.2	55.6	29.8	55.3	24.3	–	–	53.5	21.0	56.5	26.7
WTP	58.2	41.8	56.1	20.2	56.7	42.5	57.8	36.5	52.7	11.6	–	–	56.3	30.5
							LR-Syntax						*56.2*	*25.9*
MT	–	–	53.4	16.3	55.2	29.0	55.3	28.4	55.1	14.9	53.1	21.0	54.4	21.9
OC	63.8	48.7	–	–	57.9	47.8	59.1	38.5	54.1	12.2	57.3	27.2	58.4	34.9
PE	60.7	40.7	53.5	9.0	–	–	55.7	24.6	53.1	12.3	53.1	13.6	55.2	20.0
VG	67.3	53.2	56.9	21.1	58.2	45.6	–	–	51.9	11.0	55.8	25.7	58.0	31.3
WD	57.9	19.9	53.9	16.9	55.6	19.7	51.6	9.9	–	–	53.2	18.9	54.5	17.1
WTP	62.5	45.7	56.0	20.0	56.8	39.4	56.0	34.4	53.2	12.3	–	–	56.9	30.3
							LR All features						*56.2*	*27.9*
MT	–	–	53.9	17.0	51.9	29.5	56.1	34.2	55.1	14.5	52.5	21.2	53.9	23.3
OC	60.0	45.1	–	–	56.7	47.0	58.6	38.0	54.1	12.2	57.7	27.5	57.4	34.0
PE	58.1	36.3	54.6	17.3	–	–	54.1	21.4	54.0	13.5	54.4	20.4	55.0	21.8
VG	65.8	51.4	57.3	21.7	57.0	45.1	–	–	54.5	13.1	55.1	24.8	57.9	31.2
WD	62.6	38.5	55.4	19.0	56.0	30.1	55.1	23.3	–	–	53.6	20.9	56.5	26.3
WTP	58.0	41.7	56.1	20.3	56.8	42.6	59.1	38.0	52.2	11.2	–	–	56.5	30.8
							CNN:rand						*55.0*	*17.9*
MT	–	–	51.0	7.4	56.9	22.1	57.2	15.7	52.4	9.4	49.4	10.9	53.4	13.1
OC	57.1	39.7	–	–	56.4	42.8	58.9	37.3	54.6	13.2	58.4	28.9	57.1	32.4
PE	59.8	18.0	54.2	9.5	–	–	57.5	18.7	55.5	15.9	54.7	16.0	56.3	15.6
VG	68.7	51.5	55.8	19.2	57.0	32.0	–	–	51.7	10.5	54.7	22.0	57.6	27.0
WD	64.4	3.5	51.3	1.3	41.3	0.0	44.5	0.0	–	–	46.7	0.0	49.6	1.0
WTP	58.5	26.6	56.8	15.4	56.0	18.5	55.3	19.4	52.9	11.6	–	–	55.9	18.3
Majority bsl	42.9	0.0	48.0	0.0	41.3	0.0	44.5	0.0	48.6	0.0	46.7	0.0	45.3	0.0
Random bsl	47.5	30.6	50.5	14.0	51.0	38.4	51.0	29.3	49.3	9.3	50.3	20.2	49.9	23.6

Table 1: Cross-domain experiments, results only for selected systems. For each test dataset (column head) we show two scores: *Macro* F_1 score (left-hand column) and F_1 score for claims (right-hand column).

References

Or Biran and Owen Rambow. 2011. Identifying justifications in written dialogs. In *Fifth IEEE International Conference on Semantic Computing (ICSC)*, pages 162–168, Palo Alto, CA, USA.

Elena Cabrio, Sara Tonelli, and Serena Villata. 2013. *From Discourse Analysis to Argumentation Schemes and Back: Relations and Differences*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Jiang Guo, Wanxiang Che, Haifeng Wang, and Ting Liu. 2014. Revisiting Embedding Features for Simple Semi-supervised Learning. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 110–120, Doha, Qatar.

Ivan Habernal and Iryna Gurevych. 2015. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2127–2137, Lisbon, Portugal.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, 43(1):125–179.

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 343–351, Singapore.

Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2014. A PDTB-Styled End-to-End Discourse Parser. *Natural Language Engineering*, 20(2):151–184.

Raquel Mochales-Palau and Marie-Francine Moens. 2009. Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pages 98–107, Barcelona, Spain.

Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. 2007. Automatic detection of arguments in legal texts. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 225–230, Stanford, CA, USA.

Andreas Peldszus and Manfred Stede. 2016. An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation*, pages 801–815, Lisbon, Portugal.

Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, pages in press, preprint available at arXiv:1604.07370.