Datenbank Spektrum
https://doi.org/10.1007/s13222-020-00341-z

SCHWERPUNKTBEITRAG

Relational and Fine-Grained Argument Mining

The LMU Munich project *ReMLAV* within the DFG Priority Program *RATIO “Robust Argumentation Machines”*

Dietrich Trautmann[1] · Michael Fromm[2] · Volker Tresp[2] · Thomas Seidl[2] · Hinrich Schütze[1]

Received: 5 February 2020 / Accepted: 29 May 2020
© The Author(s) 2020

Abstract
In our project *ReMLAV*, funded within the DFG Priority Program *RATIO* (http://www.spp-ratio.de/), we focus on relational and fine-grained argument mining. In this article, we first introduce the problems we address and then summarize related work. The main part of the article describes our research on argument mining, both coarse-grained and fine-grained methods, and on same-side stance classification, a relational approach to the problem of stance classification. We conclude with an outlook.

Keywords Argument Mining · Stance Classification · Relational Machine Learning

1 Introduction

In the project *ReMLAV*, funded within the DFG Priority Program *RATIO* (http://www.spp-ratio.de/), the *Center for Information and Language Processing (CIS)* and the *Chair for Database Systems and Data Mining (DBS)* at LMU Munich join forces to work on argument mining, an important problem in computational argumentation. Argument mining is the task of extracting argumentative sentences from large document collections to support argument search engines. We address two aspects of argument mining: argument extraction and stance classification.

Argument extraction is the core task of argument mining by identifying those parts of a document that are argumentative. We address this problem on two levels, on the sentence-level (coarse-grained) and on the token-level (fine-grained). For sentence-level argument extraction (Sect. 3.1.1), our research focuses on representations that capture different types of information that can support this task. Sentences as a whole are classified as, e.g., argumentative vs. non-argumentative. For token-level argument extraction (Sect. 3.1.2), we formalize the problem as sequence labeling which is a novel argument mining approach. Each token in the document is labeled, e.g., as argumentative vs. non-ar-gumentative. Argumentative segments are then the set of tokens consisting of maximum sequences that are labeled as argumentative.

The second problem we address is stance classification, i.e., the classification of an argumentative segment or sen-

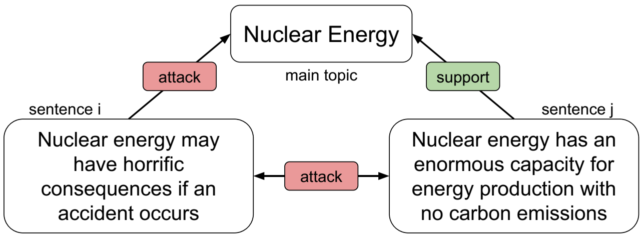
Nuclear Energy
main topic
attack
support
sentence i
sentence j
Nuclear energy may have horrific consequences if an accident occurs
attack
Nuclear energy has an enormous capacity for energy production with no carbon emissions

Fig. 1 Argumentative *sentences i* and *j* and the *main topic* [31], with *support* and *attack relations* between them

✉ Dietrich Trautmann
dietrich@cis.lmu.de

Michael Fromm
fromm@dbs.ifi.lmu.de

Volker Tresp
tresp@dbs.ifi.lmu.de

Thomas Seidl
seidl@dbs.ifi.lmu.de

Hinrich Schütze
inquiries@cislmu.org

[1] Center for Information and Language Processing, LMU Munich, Munich, Germany

[2] Database Systems and Data Mining, LMU Munich, Munich, Germany

Published online: 12 June 2020

tences with either a PRO label (arguing for a topic or point of view) or with a CON label (arguing against the topic). One important concept in this context are argumentative relations. Fig. 1 shows examples for relations between argumentative sentences and the topic “nuclear energy”. The relations are in this case supporting and attacking relations. Additionally, we develop methods to improve the overall stance classification with relational information, such as *same-side* and *not-same-side* in the same-side stance classification task (Sect. 3.2).

2 Related Work

2.1 Argumentation Schemes

A foundation for argument mining is an *argumentation sche-me*. An argumentation scheme defines what kind of arguments exist and the properties and relationships between them. Consequently, the main emphasis in argument mining lies in detecting argument components of argumentation schemes [12, 14, 16, 20, 27] and the relations between them [17, 27]. Different argumentation schemes of varying complexity have been suggested [8, 26, 30, 33].

However, many argument components (e.g., claims, prem-ises) do not generalize well across text types. Some works [6] show that it is not sufficient to train a single claim-detection model. Often the agreement between annotators during the dataset creation is low, since argumentation is a complex, highly subjective task [12]. Certain argument components (e.g., backing and warrant [30]) are often only implicitly stated [12]. Therefore, researchers have defined simpler and more tractable argumentation schemes.

In the simplest case, the argumentation scheme only differentiates between argumentative and non-argumentative text units. In a slightly more complex setting, stance information is also considered [28]. Computational argumentation models trained on these simpler argumentation schemes are often better applicable to a broader range of text genres. Based on these simpler schemes, two argument search engines, *ArgumenText*[1] [25] and *args*[2] [32] have been realized, where users can search a broad range of documents for certain topics.

Given the success of simpler argumentation schemes, we adopt them for our work.

2.2 Relational Machine Learning

A novel aspect of our approach is to model sets of arguments as graphs where each argument is a node and edges between arguments are relations like “attack” and “support”, as shown in Fig. 1. This relational model allows us to make inferences about arguments in the context of related arguments, inferences that would not be possible if we looked at each argument in isolation.

Relational data is gaining in importance in machine learning. The literature review by Nickel et al. [18], with an emphasis on knowledge graph construction, discusses many current models and datasets for relational machine learning. One of the successful models presented is RESCAL [19], which is based on tensor factorization. This model works over triples of subject, predicate and object, with the predicate describing the relation between the subject and the object. This and similar models have been trained over large knowledge graphs such as YAGO [29], DBpedia [2] and Freebase [4]. This approach could conceivably also be applied to argument graphs, but this is not trivial. For example, subjects and objects in knowledge graphs generally occur in many different relations, but most arguments in text are unique if they are represented as sequences of words.

In this article, we adopt a simpler approach to relational information: we build a graph of arguments where known edges are either same-side (both PRO or both CON) or not-same-side (one is PRO, one is CON). By incorporating new arguments into this graph, we can infer their stance.

3 Argument Mining Tasks

For argument mining, a substantial text collection is required. Many large topic-specific textual corpora can readily be retrieved from the Internet. In addition, one can exploit Internet search engines to discover and download news or discussion documents. There are also crawled web data such as *Common Crawl*[3] that can be indexed with tools like *Elasticsearch*[4]. Other resources include the *Open Web Text* [11] corpus, which is based on documents (urls) submitted to the social media platform *Reddit*[5].

Argument mining models, which are trained on annotated datasets, can be applied on the previously mentioned corpora to extract argumentative sentences. The level of granularity varies in those models and two important ones are models that are trained on the sentence-level (coarse-grained) and on the token-level (fine-grained). In our approaches, the goal is to classify whether units (sentences or tokens) are supporting (PRO), attacking (CON) or neutral (NON) toward a controversial topic. Token-level models support extracting argumentative segments that are often

[1] www.argumentsearch.com.

[2] www.args.me.

[3] http://commoncrawl.org/.

[4] https://www.elastic.co/elasticsearch/.

[5] https://www.reddit.com/.

addressing only one specific aspect of larger arguments and thus can be more useful in further downstream applications. Fine-grained models also support capturing several segments with-in a sentence that address different aspects and have different stances.

Stance classification is of central importance in argument mining, e.g., in an argument search engine that gives the user PRO arguments on one side and CON arguments on the other. Stance classification is hard because it typically requires a lot of detailed world and background knowledge as well as larger context. We approach stance classification through same-side stance classification. Pairs of argumentative paragraphs, sentences or segments are classified as being on the same-side (same stance toward a topic) or not. The graph of all arguments (with same-side and non-same-side edges) is then exploited for more accurate stance classification.

3.1 Argument Extraction

3.1.1 Sentence-Level Models

In previous work [9], some of us addressed the problem of topic-focused argument extraction on the sentence-level. Examples of the type of sentences that we extract can be seen in Fig. 2 (lines 1-3). We define topic-focused argument extraction as argument extraction where a user-defined query topic (e.g., “nuclear energy”) is given. The query topic is important for the argument extraction decision because a given sentence may be an argument supporting one topic, but not another. Since we cannot expect that available datasets cover all possible topics, the ability to generalize to unseen topics is an important requirement. Therefore, the better a machine learning model is capable of grasping the context of topic and of potential arguments, the better decisions it can make and the more confident it can be about its decisions. The work introduced recurrent and attention based networks that encode the topic information as an additional input besides the sentence. As context sources we relied on different external sources that provide the context information.

- Shallow **Word Embeddings** [3, 15, 21] are commonly used in natural-language-processing (NLP) applications and encode context information implicitly.
- **Knowledge Graphs** are heterogeneous multi-relational graphs that model information about the world explicitly. Information is represented as triples consisting of subject, predicate and object, where subject and object are entities and predicate stands for the relationship between them. Compared to textual data, knowledge graphs are structured, i.e., each entity and relationship has a distinct meaning, and the information about the modeled world is distilled in form of facts. These facts stem from texts, different databases, or are inserted manually. The reliability of these facts in (proprietary) knowledge graphs can be very high [18].
- Fine-tuning based **Transfer Learning** approaches [7, 23, 24] adapt whole models that were pre-trained on some (auxiliary) task to a new problem. This is different from feature-based approaches which provide pre-trained representations [5, 22] and require task-specific architectures for a new problem.

For the evaluation of our methods we used the UKP Sentential Argument Mining corpus [28]. It consists of more than 25,000 sentences from multiple text genres covering eight controversial topics. We have evaluated all approaches in two different settings. The *in-topic* scenario splits the data into training and test data, which leads to arguments of the same topic to appear in both training and test data. The *cross-topic* scenario aims at evaluating the generalization of the models, i.e., answering the question as to how good the performance of the models is on yet unseen topics and therefore is the more complex task. We further split the experiments in two-classes (Argument or NoArgument) and three-classes (PRO, CON, NON).

#	level	labels	sentences
1	sentence	NON	The opposition to uranium mining and nuclear power within Australia also has been linked with overseas activities .
2		PRO	The industry has shown that it can safely handle , transport and store the radioactive wastes generated by nuclear power .
3		CON	Increasing the amount of waste shipped , particularly in less secure countries , is seen as a significant increase in risk to nuclear terrorism .
4	token	NON	Not many countries have uranium mines and not all the countries have nuclear technology , so they have to hire both things overseas .
5		PRO	80 percent agreed that carbon - free nuclear energy should be expanded as one way to reduce greenhouse gases and prevent global climate change .
6		PRO & CON	Nuclear energy may have horrific consequences if an accident occurs , but it has an enormous capacity for energy production with no carbon emissions .

Fig. 2 Example sentences with annotations for the topic *“nuclear energy”* from sentence- [28] and token-level [31] datasets

Table 1 Sentence-level Macro-F_1 score for 2 classes (argumentative, non-argumentative) and for 3 classes (PRO, CON, NON) for the in-topic and cross-topic setups from our previous publication [9]

	Method	In-Topic	Cross-Topic
2-classes	BiLSTM	0.74	0.66
	BiCLSTM	0.74	0.70
	BiLSTM-KG	0.73	0.68
	CAM-Bert	0.80	0.67
	TACAM-Bert	**0.81**	**0.80**
3-classes	BiLSTM	0.56	0.46
	BiCLSTM	0.53	0.47
	BiLSTM-KG	0.56	0.47
	CAM-Bert	**0.73**	0.61
	TACAM-Bert	0.69	**0.64**

For all tasks we compare the following approaches:

- BiLSTM is the first baseline: a bidirectional LSTM model [13] that does not use topic information at all.
- BiCLSTM is the second baseline: a contextual biderectional LSTM [10]. Topic information is used as an additional input to the gates of an LSTM cell. We use the version from [28] where the topic information is only used at the $i-$ and $c-$gates since this model showed the most promising results in their work.
- BiLSTM-KG is our bidirectional LSTM model using Knowledge Graph embeddings from DBPedia as the context source for the topic.
- CAM-Bert is our fine-tuning based transfer learning approach without topic information.
- TACAM-Bert is our fine-tuning based transfer learning approach with topic information.

Table 1 shows that for the in-topic scenario our models TACAM-Bert and CAM-Bert are able to improve the Macro-F_1 score by 7% for the two-class and by 17% for the three-class classification task by using context information from transfer learning compared to the previous state-of-the-art system BiCLSTM [28]. For the more complex cross-topic task we improve the two-class setup by 10% and for the three-class setup by 17%. Our experimental results show that considering topic and context information from pre-trained models improves upon state-of-the-art argument detection models considerably. The number of parameters of the models and the hyper parameters of the training are reported in the previous publication [9].

3.1.2 Token-Level Models

Our motivation for token-level, i.e., fine-grained, models is that they support more specific selection of argumentative spans within sentences. In addition, the shorter segments are better suited to be extracted and displayed in applications (e.g., argument search engines), which usually present arguments without surrounding context sentences.

We created a new token-level (fine-grained) corpus [31]. Crowdworkers had the task of selecting argumentative spans for a given set of topics and topic related sentences. The sentences were from textual data extracted from Common Crawl[6] for a predefined list of eight topics. The final annotations of five crowdworkers per sentence were merged and a label from the set {PRO, CON, NON} was assigned to each token (word) in the sentence. The final corpus, the AURC (argument unit recognition and classification) corpus, contains 8000 sentences with 4500 being argumentative sentences and a total of 4973 argumentative segments. Examples for token-level annotations of argumentative spans in the AURC corpus are displayed in Fig. 2 in lines 4–6.

The differentiator to previous work and datasets is that there are many sentences in AURC with more than one argumentative segment. An example for a sentence with mixed stance segments can be seen in Fig. 2 in line 6, with a CON and a PRO segment. This kind of fine-grained argumentative data cannot be modeled correctly with a sentence-level approach.

After the corpus creation process, we applied state-of-the-art models in natural language processing to establish strong baselines for this new task of AURC. The proposed baselines were a majority baseline (where all tokens were labeled with the most frequent class), a BiLSTM model (using the FLAIR library [1]) and a BERT model [7] in several configurations (such as base, large and with a CRF-layer). The performance of the models was compared with two different data splits. (i) An in-domain split, where the models were trained, evaluated and tested on the same set of topics. (ii) A cross-domain split, where the models were trained on a subset of the available topics and evaluated and tested on different out-of-domain topics. The second set-up is more challenging, since the models have to generalize the argument span selection for unseen topics. Furthermore, the cross-domain split is also closer to a real world application,

[6] http://commoncrawl.org/2016/02/february-2016-crawl-archive-now-available/.

Table 2 Token-level Macro-F_1 for 2 classes (2-cl: ARG, NON) and for 3 classes (3-cl: PRO, CON, NON) for the in-domain and cross-domain setups from our previous publication [31]

	Set	In-Domain	Cross-Domain
2-classes	dev	0.813	0.797
	test	0.782	0.770
3-classes	dev	0.743	0.615
	test	0.696	0.620

since we typically encounter topics that are not covered in the training set in many practical applications.

An interesting insight from this experiment is that it is also quite challenging for humans to correctly classify argumentative spans. It is probably for this reason that, depending on the evaluation measure, some models performed better than the human annotators. An error analysis provided the following interesting insights: The most common error was incorrect stance classification (especially in the cross-domain setup) compared to good performance for span recognition, for both in-domain and cross-domain. Table 2 shows the results for the best models.

In summary, token-level (i.e., fine-grained) models are close to or better than human performance for known topics. While the cross-domain setup turned out to be challenging, the results for in-domain topics are already useful and can be helpful for many downstream tasks in computational argumentation. Examples include clustering or grouping of similar arguments for the ranking task in argument search engines; and the summarization of argument segments in automated debating systems[7] that generate fluent compositions of extracted argumentative segments. Future work should address annotating sentences for many more topics, cross-domain performance and better representations for linguistic objects of different granularities.

3.2 Same-Side Stance Classification

As the experiments in our previous work ([9], see also Table 1) showed, there is still a huge gap of 16% Macro-F_1 score between the two-class and the three-class cross-topic scenario and of 8% in the in-topic scenario. The reason is that stance detection is a complex task. The Same-Side Stance Classification (SSSC) Challenge[8] addresses this problem. As an illustration consider the PRO argument "religion gives purpose to life". The PRO argument "religion gives moral guidance" is an example for a same-side argument, whereas the CON argument "religion makes people fanatic" is an example for a not-same-side argument.

Given two arguments regarding a certain topic, the SSSC task is to decide whether or not the two arguments have the same stance. This can be exploited for stance classification since the relations bring to bear additional information (information about the network of all arguments) for improved stance classification.

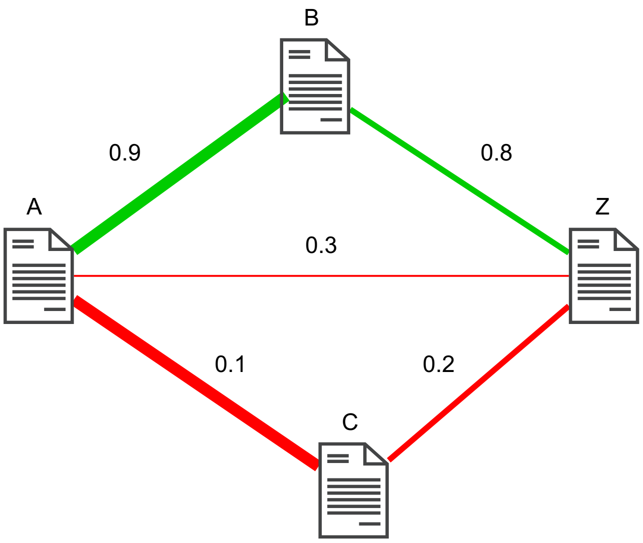

Fig. 3 Example of an argument graph. The nodes are represented as arguments and the edges as the binary SSSC relation. The thickness and the color of the edges represent the confidence and the class. Low confidence values can be interpreted as high confidence values against the relation

Our group participated in the challenge with a pretrained transformer model [7] fine-tuned on the SSSC data. We organized the data as graphs in the following way: we generated one graph per topic where the nodes are arguments and the edges are weighted with the confidence that the SSSC relation holds. If it is already known (e.g., from the training set) that the arguments agree or disagree, the confidence is 0 and 1 accordingly. Otherwise we use the probability predicted by the fine-tuned transformer model. Fig. 1 shows an illustration of the graph.

For each pair of arguments in the test set we computed the confidence of all paths of length k, and greedily selected the edge with the highest confidence for either an agreement or a disagreement between the two arguments. We computed the path score as the product of confidences of the edges on a path. By using the graph structure and the transitivity of the SSSC relation we could improve our Macro-F_1 score from 0.57 by 7 points for the cross-topic scenario.

4 Conclusion

Our ongoing work addresses several of the issues discussed in Sect. 3. Important issues we are addressing are the improvement of stance classification and the annotation for a larger number of topics. For stance classification, it is of

[7] https://www.research.ibm.com/artificial-intelligence/project-debater/.

[8] https://sameside.webis.de/.

interest to incorporate additional information in a multi-task learning setup, e.g., sentiment information and information from knowledge graphs. For annotating more topics, we can use our current models, which are trained on the eight AURC topics with gold labels, for a better sampling of sentences from a corpus such as OpenWebText [11] for new topics.

5 Future Work

This project overview mostly addressed lower-level tasks in computational argumentation. These are very important and essential to solve higher-level tasks that can only be accomplished with this extracted argumentative information on the sentence- and token-level. For the future we see these tasks as building blocks for high-level argumentation applications. One such application is argument validation, i.e., the classification of a sequence of two sentences as a valid vs. invalid link in a reasoning chain. With our improved argument mining techniques and based on our relational framework for stance classification, we would like to exploit graphs for argument validation. Another high-level argumentation application is interpretability of argument mining decisions: users in many applications can benefit from being able to view the rationale for why a particular sentence was selected as argumentative and with a particular stance. Here the human-interpretable information sources that we incorporated into sentence-level mining could be the basis for more effective methods. For future work, we are also considering other demanding tasks which could benefit from our work. One is the clustering or grouping of argumentative sentences or segments; and a second one the summarization of argument segments in automated debating systems that generate fluent compositions of extracted argumentative segments.

Acknowledgements This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project*ReMLAV*, Grant Numbers SCHU 2246/13 and SE 1039/10, as part of the Priority Program *RATIO*(SPP-1999). We are grateful to our collaborators Johannes Daxenberger, Christian Stab, Evgeniy Faerman and Iryna Gurevych.

Funding Open Access funding provided by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Akbik A, Bergmann T, Blythe D, Rasul K, Schweter S, Vollgraf R (2019) Flair: an easy-to-use framework for state-of-the-art nlp. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics (Demonstrations)
2. Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z (2007) DBpedia: a nucleus for a web of open data. In The semantic web. Springer, Berlin, Heidelberg, pp 722–735
3. Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. Trans Assoc Comput Linguist 5:135–146
4. Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J (2008) Freebase: a collaboratively created graph database for structuring human knowledge. In: Proc. of the 2008 ACM SIGMOD Int. Conf. on Management of data, pp 1247–1250
5. Dai AM, Le QV (2015) Semi-supervised sequence learning. In Advances in neural information processing systems, pp 3079–3087
6. Daxenberger J, Eger S, Habernal I, Stab C, Gurevych I (2017) What is the essence of a claim? cross-domain claim identification. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 2055–2066
7. Devlin J, Chang M, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol 1 (Long and Short Papers), pp 4171–4186
8. Freeman JB (2011) Argument structure: representation and theory, Vol 18. Springer Science & Business Media, Dordrecht
9. Fromm M, Faerman E, Seidl T (2019) TACAM: topic and context aware argument mining. In: 2019 IEEE/WIC/ACM Int. Conf. on Web Intelligence WI 2019, Thessaloniki, Greece, October 14–17, 2019, pp 99–106
10. Ghosh S, Vinyals O, Strope B, Roy S, Dean T, Heck LP (2016) Contextual LSTM (CLSTM) models for large scale NLP tasks. CoRR:(abs/1602.06291)
11. Gokaslan A, Cohen V (2019) Openwebtext corpus. http://web.archive.org/web/*/https://skylion007.github.io/OpenWebTextCorpus/
12. Habernal I, Gurevych I (2016) Argumentation mining in user-generated web discourse. Comput Linguist 43(1):125–179
13. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
14. Hua X, Wang L (2017) Understanding and detecting supporting arguments of diverse types. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Vol 2: Short Papers), pp 203–208
15. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint. arXiv:1301.3781
16. Nguyen H, Litman D (2015) Extracting argument and domain words for identifying argument components in texts. In: Proceedings of the 2nd Workshop on Argumentation Mining. Association for Computational Linguistics, Denver, CO, pp 22–28
17. Nguyen H, Litman D (2016) Context-aware argumentative relation mining. In: Long papers, vol 1. Association for Computational Linguistics, Berlin, pp 1127–1137
18. Nickel M, Murphy K, Tresp V, Gabrilovich E (2015) A review of relational machine learning for knowledge graphs. Proc IEEE 104(1):11–33
19. Nickel M, Tresp V, Kriegel H (2011) A three-way model for collective learning on multi-relational data. In: Proc. of the 28th Int.

Conf. on Machine Learning ICML 2011, Bellevue, Washington, USA, June 28–July 2, 2011, pp 809–816

20. Palau RM, Moens MF (2009) Argumentation mining: the detection, classification and structure of arguments in text. In: Proc. of the 12th Int. Conf. on Artificial Intelligence and Law ICAIL '09. ACM, New York, NY, USA, pp 98–107
21. Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Proc. of the 2014 Conf. on empirical methods in natural language processing (EMNLP), pp 1532–1543
22. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: Proc. of NAACL
23. Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/languageunderstandingpaper.pdf
24. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language models are unsupervised multitask learners. OpenAI Blog 1:8
25. Stab C, Daxenberger J, Stahlhut C, Miller T, Schiller B, Tauchmann C, Eger S, Gurevych I (2018) ArgumenText: Searching for arguments in heterogeneous sources. In: Proceedings of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Demonstrations. Association for Computational Linguistics, New Orleans, pp 21–25
26. Stab C, Gurevych I (2014) Identifying argumentative discourse structures in persuasive essays. In: Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing EMNLP 2014, Doha, Qatar, October 25–29, 2014, pp 46–56 (A meeting of SIGDAT, a Special Interest Group of the ACL)
27. Stab C, Gurevych I (2017) Parsing argumentation structures in persuasive essays. Comput Linguist 43(3):619–659
28. Stab C, Miller T, Gurevych I (2018) Cross-topic argument mining from heterogeneous sources using attention-based neural networks. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp 3664–3674
29. Suchanek FM, Kasneci G, Weikum G (2007) Yago: a core of semantic knowledge. In: Proc. of the 16th Int. Conf. on World Wide Web, pp 697–706
30. Toulmin SE (1958) The uses of argument. Cambridge University Press
31. Trautmann D, Daxenberger J, Stab C, Schütze H, Gurevych I (2020) Fine-grained argument unit recognition and classification. In: The Thirty-Fourth AAAI Conf. on Artificial Intelligence AAAI 2020, New York City, NY, USA. AAAI Press. https://aaai.org/Papers/AAAI/2020GB/AAAI-TrautmannD.7498.pdf
32. Wachsmuth H, Potthast M, Al Khatib K, Ajjour Y, Puschmann J, Qu J, Dorsch J, Morari V, Bevendorff J, Stein B (2017) Building an argument search engine for the web. In: Proceedings of the 4th Workshop on Argument Mining. Association for Computational Linguistics, Copenhagen, pp 49–59
33. Walton D (2012) Argument mining by applying argumentation schemes. Stud Log 4(1):2011