Argumentation Mining SOTA

By a lazy guy

2021

Contents

1 Introduction

A simple document where we collect all interesting papers about Argumentation Mining (AM). Feel free to add new papers, small summaries and/or your thoughts about them!

2 Datasets

Datasets are categorized based on the tasks that can be formulated.

2.1 Argument Quality

- **IBM-ArgQ-6.3kArgs** [77]: contains 6.3k arguments along with a real score $\in [0, 1]$ concerning its quality. Quality here means: "would you recommend this argument or not?".
- **IBM-ArgQ-14kPairs** [77]: contains 14k pairs of arguments with a binary label that tells which of the two arguments is of higher quality: "would be preferred by people".
- **IBM-ArgQ-Rank30kArgs** [26]: 30k arguments from 71 topics annotated with stance score and quality score.
- **Discussion Tracker Corpus** [56]: 29 multi-party discussions of English literature transcribed from 985 minutes of audio. Text is annotated with argument moves (claim, evidence and explanation), specificity (low, medium, high) and collaboration (disagreement or agreement). There's a total of 3261 student turns, 2128 of them are argumentative.
- **UKPConvArg1 Corpus** [28]: 16k pairs of arguments concerning 32 topics.
- **Classroom Discussion Corpus** [48]: 2047 argument moves from 73 student discussions (claim, evidence and warrant) with specificity annotations (argument quality).
- **IBM-EviConv** [23]: 5.697 evidence pairs with annotation about which evidence is the most convincing
- **GAQCorpus** [41]: Large theory-based corpus that annotates arguments based on: cogency, effectiveness and reasonableness. The corpus is comprised of 3 different corpora: CQA (Yahoo) with 2.088 arguments, Debates (CMV and IAC V2) with 1.337 and 766 arguments respectively and Reviews (Yelp) with 1.104 arguments.

2.2 Argument Similarity

- **ASPECT** [64]: 3.595 sentence pairs overs 28 controversial topics. Pairs are annotated using three different levels of similarity from very similar arguments to totally unrelated.

2.3 Stance Classification

- **Claim Polarity Dataset**: [6]: contains 2.394 claims across several topics.
- **IBM Evidence Sentences** [19]: 35.211 sentences from Wikipedia on 321 topics annotated for their stance w.r.t. topic and whether they are evidence or not.
- **Internet Argument Corpus** [1]: 3 corpora: 4forums (414k posts), ConvinceMe (65k posts) and CreateDebate (3k posts). Each post is annotated for stance (pro/con) and stuff like emotion, sarcasm and agreement scores.

2.4 Argument Summarization

- **ArgKP** [8]: 24k pairs of argument - key point labeled as matching/non-matching. Key points are generated by experts.
- **DebateSum: A large-scale argument mining and summarization dataset** [68]: 187.386 unique documents organized into triples: argument, evidence and summary (along with document meta-data).

2.5 Argument Link Prediction

- **Argumentative Microtext** [53]: 112 argumentative microtexts encompassing argument structure, discourse relations and argument schemes.
- **CrowdSourced Argumentative Microtext** [73]: Extension of Argumentative Microtexts Corpus via crowdsourcing. 171 argumentative microtexts with argument component and argument relation annotations: 932 argumentative units.
- **AbstRCT** [50]: 500 abstracts of medical components with 4.198 arguments (major claim, claim, premise). Attack relations are given as well.
- **Argumentative Microtexts** [59]: 112 argumentative short texts with attack/support relations.
- **Italian Short Argumentative Texts** [54]: italian translation of argumentative microtexts dataset.
- **Discussion Tracker Corpus** [56]: 29 multi-party discussions of English literature transcribed from 985 minutes of audio. Text is annotated with argument moves (claim, evidence and explanation), specificity (low, medium, high) and collaboration (disagreement or agreement). There's a total of 3261 student turns, 2128 of them are argumentative.
- **Relation Argumentation Corpus** [14]: 854 sentence pairs with relation annotations (attack, support, non-relation).
- **CMV Subreddit Corpus** [15]: 380 turns of dialogue for 2.756 sentences: 2.741 argumentative propositions out of which 1.205 are claims and 1.536 are premises with additional 799 non-argumentative propositions.

2.6 Argument Detection

- **Sentential Argument Mining Corpus** [76]: 25k sentences covering 8 topics. Sentences are labelled as pro/con argument or non-argumentative.
- **AbstRCT** [50]: 500 abstracts of medical components with 4.198 arguments (major claim, claim, premise). Attack relations are given as well.
- **LN55k** [25]: 55.024 manually curated claims for 192 topics (from 400 million newspaper articles provided by LexisNexis)
- **IBM Evidence Sentences** [19]: 35.211 sentences from Wikipedia on 321 topics annotated for their stance w.r.t. topic and whether they are evidence or not.
- **Discussion Tracker Corpus** [56]: 29 multi-party discussions of English literature transcribed from 985 minutes of audio. Text is annotated with argument moves (claim, evidence and explanation), specificity (low, medium, high) and collaboration (disagreement or agreement). There's a total of 3261 student turns, 2128 of them are argumentative.
- **IBM Speech Claim Mining 19** [42]: 400 speeches with annotated claims for a total of 4.882 claims. Each motion has two speeches for rebuttal.
- **Webis-Debate-16 Corpus** [5]: 28.689 text segments concerning 14 topics from idebate.org, 23.880 argumentative and 4.809 non-argumentative. Corpus is created using distant supervision heuristic and templates.
- **Persuasive Essays Corpus** [75]: 402 essays annotated with argumentative components (major claim, claim, premise) and with attack/support relations. Overall there are 751 major claims, 1.506 claims and 3.832 premises.
- **News Editorial Corpus**: 300 editorial annotated with 6 different types of ADUs: Common Ground, Assumption, Testimony, Statistics, Anecdote, Other.
- **CrowdSourced Argumentative Microtext** [73]: Extension of Argumentative Microtexts Corpus via crowdsourcing. 171 argumentative microtexts with argument component and argument relation annotations: 932 argumentative units.
- **Argument Online Civic Discussion Corpus** [51]: Japanese dataset comprised of 399 threads (1.327 posts) with argument components and argument relation annotations.
- **Dr. Inventor Corpus** [40]: argument component and argument relation annotations on scientific publications. Components: 2.751 background claims, 5.445 own claim, 4.093 data. Relations: 5.790 supports, 696 contradicts, 44 semantically same.
- **Classroom Discussion Corpus** [48]: 2047 argument moves from 73 student discussions (claim, evidence and warrant) with specificity annotations (argument quality).
- **A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics** [2]: Wikipedia articles over 33 topcis with 2.683 arguments (claim and evidence).
- **On the Discoursive Structure of Computer Graphics Research Papers** [21]: 10.789 sentences concerning scientific documents on computer graphics categorized 5 different categories: Approach, Background, Challenge, Future Work, Outcome.

- **DebateSum: A large-scale argument mining and summarization dataset** [68]: 187.386 unique documents organized into triples: argument, evidence and summary (along with document meta-data).
- **Argument Annotated User Generated Discourse Corpus** [29]: 340 user comments, forum posts, blogs and newspaper articles annotated for argument units: premise, claim, rebuttal, refutation and backing.
- **AURC-8** [81]: 8.000 annotated sentences: 3.500 non-argumentative, 4.500 argumentative divided into: 1.951 PRO arguments, 1799 con arguments and 750 that may be both since they contain multiple argumentative units.
- **USElecDeb60to16 Corpus** [31]: 39 political debates from last 50 years US presidential campaigns with 29k argument components (premises and claims).

2.7 Argument Knowledge Graph Construction

- **End-to-End Argumentation Knowledge Graph Construction** [37]: 16.429 annotations of 4.740 claims. Annotations included positive and negative effect relations found in claims, along with their associated concept instances within the external KB.

2.8 Argument Reasoning Comprehension

- **UKP SemEval18** [30]: 1.970 instances divided into three sets based on the year of the debated. 2011-2015 (train set with 1.210 instances); 2016 (val set with 316 instances); 2017 (test set with 444 instances).

2.9 Argument Framing

- **Webis-argument-framing-19** [3]: 465 topics with 1.623 frames and 12.326 arguments.

2.10 Argument Aspect Detection

- **UKP Sentential Aspect Argument Mining Corpus** [69]: 5.032 arguments annotated with aspects.
- **UKP ABAM Corpus** [80]: extension of UKP Sentential Argument Mining Corpus with the following statistics: 4.396 sentences, 4.792 segments, 12.040 aspects (4.525 unique).

2.11 Debate Topic Expansion

- **Debate Topic Expansion Dataset** [10]: 3k pairs of debate concepts and their expansion candidates.

2.12 Argument Invention

- **Argument Invention Corpus** [11]: 689 motions associated to 37 CoPAs (class of principled arguments)

2.13 Argument Generation

- **CMV Corpus** [35]: 26.761 threads from Reddit subcommunity Change My View (CMV).
- **Style Annotated CMV Corpus** [36]: CMV corpus with style annotations (claim, premise and functional)

3 Annotation Schemes

Papers:

- **Towards Feasible Guidelines for the Annotation of Argument Schemes** [52]: offers a hierarchical and finite taxonomy of argument schemes as well as systematic linguistically-informed criteria to distinguish various types of argument schemes. Schemes are based on the Argumentum Model of Topics.
- **Towards Assessing Argumentation Annontation - A First Step** [47]: Using Walton's argument schemes for annotating arguments in Swedish political texts. They describe some annotation schemes.
- **Proposed Method for Annotation of Scientific Arguments in Terms of Semantic Relations and Argument Schemes** [24]: 15 annotation schemes for scientific argument annotation in biological/biomedical journal articles.
- **Annotating Claims in the Vaccination Debate** [79]: 3 different annotation schemes for claim detection in the Vaccination Debate Corpus. First: Major Claim, Claim, Premise. Second: Claims, Grounds, Backing, Rebuttal and Refutation. Third: Claim, Not-Argumentative
- **Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum** [33]: two-tiered annotation schem to label claims and premises and their semantic types in an online persuasive forum, Change My View. Premises are annotated with the three types of persuasive modes: ethos, logos, pathos. Claims are labeled as interpretation, evaluation, agreement and disagreement.
- **Argumentation Synthesis following Rhetorical Strategies** [82]: Annotation attempt of Argumentative Microtext corpus following rhetorical strategies. Arguments are selected, arranged following a given rhetorical strategy and then presented with a chosen writing style.
- **A Bayesian Approach for Sequence Tagging with Crowds** [72]: Bayesian approach for aggregating sequence labels of different croudsource workers.

4 Tasks

4.1 Argument Quality

Argument Quality is the task of predicting the quality score (real score) of an argument or determine which argument (among a given set - pair, triplet, etc..) is of higher quality. Lastly, argument quality may also refer to a general purpose ranking system.

Papers:

- **Automatic argument quality assessment - new datasets and methods** [77]: BERT-based classification of arguments quality scores and pair ranking.
- **Finding Convincing Arguments Using Scalable Bayesian Preference Learning** [71]: Gaussian process for preference learning concerning argument ranking when gold labels are not available or subdued to annotator noise.
- **Incorporating Topic Aspects for Online Comment Convincingness Evaluation** [27]: argument convincingness via bi-lstm + gnn model that also incorporates topic information to determine argument-to-argument edges.
- **Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM** [28]: classification of which argument is more convincing in a given pair and global argument ranking.
- **Assessing Convincingness of Arguments in Online Debates with Limited Number of Features** [16]: feature selection for argument convincingness prediction on the UKPConvArg1 corpus.
- **Are You Convinced? Choosing the More Convincing Evidence with a Siamese Network** [23]: Siamese LSTM to determine which evidence in a pair is the most convincing one on the IBM-EviConv corpus.

4.2 Argument Ranking

Usually arguments are ranked based on their quality (which is synonym for persuasiveness).

Papers:

- **Ranking Passages for Argument Convincingness** [61]: roughly 30k passages concerning 3.234 topics.

4.3 Stance Classification

In AM it generally means to predict the stance of an argument towards a given topic. Thus, the input is comprised of two sentences/texts: an argument and a topic.

Papers:

- **Improving Claim Stance Classification with Lexical Knowledge Expansion and Context Utilization** [7]: linear SVM classifier that uses sentiment polarity scores (from another trained classifier) and claim contextual hand-made features to predict the stance of the claim towards the given topic.
- **Multilingual Argument Mining: Datasets and Analysis** [78]: evidence detection and stance classification via BERT models on a multi-lingual argument dataset.

4.4 Argument Detection

In this subsection we both consider sequence tagging formulation as well as the sentence level classification.

Papers:

- **Unsupervised corpus-wide claim detection** [44]: Hand-made queries based on a claim lexicon to retrieve potential claim sentences from the web. The claim is then extracted from the given pattern.
- **Transformer-based Argumentation Mining for Healthcare Applications** [50]: BERT-based argument component classification (sequence tagging) and argument relation prediction.
- **Multilingual Argument Mining: Datasets and Analysis** [78]: evidence detection and stance classification via BERT models on a multi-lingual argument dataset.
- **Lexicon Guided Attentive Neural Network Model for Argument Mining** [46]: argument detection on Sentential Argument Mining Corpus where an attentive bi-LSTM network uses an attention mechanism that is extend via external lexicons: claim, sentiment, emotion lexicons and WordNet. In particular, lexicons are used to define an additional input (from the given one) which is then embedded and concatenated to the original input before attention. Not so clear how this additional input is defined...
- **Discussion Tracker Corpus** [56]: CNN based neural network single and joint classification of: argument component detection, argument quality and argument relation.
- **Argumentative Evidences Classification and Argument Scheme Detection Using Tree Kernels** [45]: Evidence classification SVM with tree kernels on two datasets.
- **Cross-Domain Mining of Argumentative Text through Distant Supervision** [5]: feature-based classifier trained on Webis-Debate-16 Corpus on argumentative sentence detection and cross-domain test on Persuasive Essays corpus.
- **Unlocking Transfer Learning in Argumentation Mining: A Domain-Independent Modelling Approach** [83]: BERT-based classification of argumentative components on multiple argumentative datasets: Persuasive Essays Corpus, AuracariaDB, UGC Web Discourse and Blog Comments Corpus.

- **Neural End-to-End Learning for Computational Argumentation Mining** [18]: argument detection and argument link prediction as sequential tagging problems via Tree-LSTMs.
- **End-to-End Argumentation Mining in Student Essays** [60]: ILP for argument detection and relation prediction in Persuasive Essays Corpus.
- **End-to-End Argument Mining for Discussion Threads Based on Parallel Constrained Pointer Architecture** [51]: Joint learning of argument sentence detection and link prediction via bi-LSTM and constrained pointer networks on a proprietary dataset.
- **Argument Component Classification for Classroom Discussions** [48]: argument component classification for transcribed spoken classroom discussions, with the goal of automatically classifying student utterances into: claims, evidence and warrants.
- **Unit Segmentation of Argumentative Texts** [4]: Feature enhanced LSTM + CRF model for argument component detection formulated as a sequential tagging problem. Datasets: Persuasive Essays corpus, Webis-Debate-16-Corpus and Argument Annotated User Generated Web Discours corpus.
- **Fine-Grained Argument Unit Recognition and Classification** [81]: BERT-based argument component detection along with stance detection on AURC-8 dataset.
- **Argumentation Mining on Essays at Multi Scales** [85]: BERT based classification specialized for each argument type on Persuasive Essays corpus.
- **GRASP: Rich Patterns for Argumentation Mining** [70]: feature-based iterative algorithm that expands terms and filters them via information mutual gain (w.r.t. labels).

4.5 Argument Aspect Detection

Argument aspect detection also considers sub-tasks such as: Aspect Term Extraction (ATE) and Nested Segmentation (NS).

Papers:

- **Aspect-Based Argument Mining** [80]: BERT based ATE and NS formulated as a sequential tagging problem.

4.6 Argument Summarization

Papers:

- **From Arguments to Key Points: Towards Automatic Argument Summarization** [8]: BERT-based classification is explored to predict the corresponding key point to each given argument.

- **Quantitative Argument Summarization and Beyond: Cross-Domain Key Point Analysis** [9]: automatic extraction of key points via score-based models for argument summarization and general analysis of key point

4.7 Argument Knowledge Graph Construction

Papers:

- **End-to-End Argumentation Knowledge Graph Construction** [37]: automatic construction of argument knowledge graph. The task is divided in: relation detection, relation type classification, consequence classification, concept identification

4.8 Argument Generation

Papers:

- **The workweek is the best time to start a family - A study of GPT-2 Based Claim Generation** [25]: claim generation based on GPT-2 models and exploration of generated content by measuring claim veracity, plausibility and stance using manual and automatic assessment.
- **Computational Argumentation Synthesis as a Language Modeling Task** [20]: Given a topic-stance pair and a set of pre-defined ADUs, the pipeline selects a set of ADUs and arranges them according to a rhetorical strategy.
- **Aspect-Controlled Neural Argument Generation** [69]: conditioned BERT model (CTRL) for argument generation conditioned on topic and given aspect.
- **Fixed That for You: Generating Contrastive Claims with Semantic Edits** [32]: generation of contrastive claims from a Reddit post dataset via a Seq2Seq approach. The model takes into account terms to replace via a counter variable and might use edit values in input as well.
- **Neural Argument Generation Augmented with Externally Retrieved Evidence** [35]: generation of constrastive/counter arguments via a 2-step seq2seq pipeline. Firstly, the model generates a set of talking point phrases as an intermediate representation. Secondly, a separate decoder produces the final arguments based on the input and on the keyphrases.
- **Sentence-Level Content Planning and Style Specification for Neural Text Generation** [36]: multi-purpose 2-step generation pipeline experimented for: argument generation, summarization and abstract generation. In the first step, a set of keyphrases is selected from a memory and a style specification is selected as well (claim, premise, functional). Secondly, the decoder, produces the a sentence based on input and step 1 results.

- **Argument Generation with Retrieval, Planning and Realization** [34]: counter argument generation via a new framework called CANDELA. The framework is a 2-step pipeline. At each step, a set of keyphrases is selected from a memory M (with notion of previously selected keyphrases) and encodes each input sentences. Secondly, a decoder generates arguments based on encoded inputs and selected keyphrases. The decoder chooses between two types of style: argumentative content sentence and argumentative filler sentence (annotation is done via hand-made rules).
- **Automatic Claim Negation: Why, How and When** [12]: rule based claim negation algorithm along with feature-based plausibility classification.

4.9 Argument Reasoning Comprehension

Given an argument with a claim and a premise, the goal is to choose the correct implicit warrant from two options. Both warrants are plausible and lexically close, but lead to contradicting claims.

Papers:

- **The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants** [30]: intra-attention bi-LSTM based classification w/o additional context.
- **Frame- and Entity-Based Knowledge for Common-Sense Argumentative Reasoning** [13]: classification is enhanced with external knowledge: FrameNet's Event Knowledge and Wikidata's Fact Knowledge.

4.10 Argument Framing

It is the task of splitting a set of arguments into a set of non-overlapping frames. A frame is a set of arguments that focus on the same aspect.

Papers:

- **Modeling Frames in Argumentation** [3]: fully unsupervised approach to cluster arguments into non-overlapping frames. The method firstly clusters arguments based on topics, then removes topic features from each cluster (BoW approach) and lastly clusters parsed arguments into clusters.

4.11 Argument Link Prediction

Papers:

- **Dissecting Content and Context in Argumentative Relation Analysis** [57]: shows that link prediction is usually biased w.r.t. context information. In particular, context information per se is sufficient to solve the problem. To alleviate this issue, argumentative units are separated from their context such that the system is forced to rely on ADUs content. This work analyses Stab and Gurevych features employed on Persuasive Essays.
- **Neural End-to-End Learning for Computational Argumentation Mining** [18]: argument detection and argument link prediction as sequential tagging problems via Tree-LSTMs.
- **End-to-End Argumentation Mining in Student Essays** [60]: ILP for argument detection and relation prediction in Persuasive Essays Corpus.
- **Context-aware Argumentative Relation Mining** [55]: Relation prediction in Persuasive Essays Corpus is enhanced with topic information or context information within a parametric window. All feature are hand-made.
- **End-to-End Argument Mining for Discussion Threads Based on Parallel Constrained Pointer Architecture** [51]: Joint learning of argument sentence detection and link prediction via bi-LSTM and constrained pointer networks on a proprietary dataset.
- **Here's My Point: Joint Pointer Architecture for Argument Mining** [62]: joint link prediction and argument component detection with pointer networks on Persuasive Essays and Argumentative Microtext corpora.
- **Cross-Lingual Argumentative Relation Identification: from English to Portuguese** [67]: argumentative link prediction on english and portoguese datasets with transfer learning techniques.

4.12 Argumentation as a Support Tool

Papers:

- **Gradual Argumentation Evaluation for Stance Aggregation in Automated Fake News Detection** [38]: Stance classification for fake news detection by applying stance aggregation strategy from argumentation mining.
- **Predicting the Usefulness of Amazon Reviews Using Off-The-Shelf Argumentation Mining** [58]: enrich a classifier with features of argumentation mining (MARGOT prediction scores).

4.13 Argument Template Instantiation

Automatically identifying reasoning patterns of arguments using predefined templates.

Papers:

- **Feasible Annotation Scheme for Capturing Policy Argument Reasoning using Argument Templates** [65]: propose the task of Argument Template Instantiation by providing a set of argument templates (AT), a small dataset of instances from Argument Microtext Corpus and formulate the task as a structured prediction task.

4.14 Dialogue Agent

Papers:

- **Dave the debater: a retrieval-based and generative argumentative dialogue agent** [43]: retrieval and generative dialogue agent. The retrieval is based on a siamese LSTM called Manhattan LSTM that retrieves the most similar sentence to the user query in the given corpus. The retrieval system is trained on a similarity dataset (Quora Kaggle question pairs dataset). The generative process is a hierarchical LSTM that reads previous words/sentences in the conversation and then proceeds generating a response.

4.15 Debate Topic Expansion

Find topics related to the given one.

Papers:

- **From Surrogacy to Adoption; From Bitcoin to Cryptocurrency: Debate Topic Expansion** [10]: classification of consistent topics via logistic regression and LSTMs.

4.16 Argument Invention

Association of topics to general topics, known as first principles or, more formally, classes of principled arguments (CoPAs).

Papers:

- **Argument Invention from First Principle** [11]: motion to CoPA matching

5 Models

- **BERT**: [77], [8], [9], [50], [15]
- **XLNet**: [9]
- **RoBERTa**: [9], [50]

- **ALBERT**: [9]
- **BioBERT**: [50]
- **SciBERT**: [50]
- **CTRL**: [69]
- **Linear SVM**: [7]
- **LSTM (w/o attention)**: [30], [28], [51], [39], [48], [67]
- **Seq2Seq**: [32], [35], [34], [36]
- **Feature-based Classifier**: [75], [55], [28], [48], [21], [16], [12]
- **Tree-LSTM**: [18]
- **GNN**: [27]
- **Pointer Network**: [51], [62]
- **CNN**: [48]

6 General Argumentation

In this section, we collect papers that discuss about argumentation (whether computational, natural or a comparison between them).

Papers:

- **How Natural is Argument In Natural Language?** [17]: Discuss the differences of argumentation (dialogical scenario) in natural and formal settings. For example, disagreement is usually avoided or delayed.
- **Neural-Symbolic Argumentation Mining: An Argument in Favor of Deep Learning and Reasoning** [22]: Probabilistic symbolic approach using DeepProbLog that also combines subsymbolic methods (deep learning).

7 Tools

- **ArgumenText**: static web crawler (over 400 million pages) to retrieve relevant texts. Texts are then annotated at sentence level via a classifier (BERT or modified LSTM that takes into account topic information) and are eventually clustered based on their similarity/affinity using a model trained on a argument similarity dataset.

- **AL: An Adaptive Learning Support System for Argumentation Skills** [84]: An online tool that provides feedback about argument convincness and its structure. In particular, it tells if some claim is missing an evidence, if its readable (readibility score), coherent and persuasive (scores).
- **ArgMine: A Framework for Argumentation Mining** [66]: a tool for annotating and detecting argumentative components in Portogues texts.
- **ArguminSci: A Tool for Analyzing Argumentation and Rhetorical Aspects in Scientific Writing** [39]: tool like MARGOT that is specific to scientific text and that allows for parallel analysis of four other rhetorical aspects of scientific writing: discourse role classification, subjective aspect classification, summary relevance classification and citation context identification. Models (bi-LSTMs) are trained on Dr. Inventor Corpus.
- **ArgumenText: Searching for Arguments in Heterogeneous Sources** [74]: argument retrieval systems capable of retrieving sentential arguments for any given controversial topic. Document retrieval conditioned on the given topic is done via Elasticsearch tool. Argument components are then retrieved along with stance information.

8 Surveys

Papers:

- **The Argument Web: an Online Ecosystem of Tools, Systems and Services for Argumentation** [63]: a survey of argumentation mining with particular attention on how to digest such information for user interfacing.
- **The evolution of argumentation mining: From models to social media and emerging tools** [49]: a survey on the challenges of applying argumentation to social media data

References

[1] Rob Abbott, Brian Ecker, Pranav Anand, and Marilyn Walker. Internet argument corpus 2.0: An SQL schema for dialogic social media and the corpora to go with it. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4445–4452, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).

[2] Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics. In *Proceedings of the First Workshop on Argumentation Mining*, pages 64–68, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[3] Yamen Ajjour, Milad Alshomary, Henning Wachsmuth, and Benno Stein. Modeling frames in argumentation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*

(EMNLP-IJCNLP), pages 2922–2932, Hong Kong, China, November 2019. Association for Computational Linguistics.

[4] Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein. Unit segmentation of argumentative texts. In *Proceedings of the 4th Workshop on Argument Mining*, pages 118–128, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[5] Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein. Cross-domain mining of argumentative text through distant supervision. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1395–1404, San Diego, California, June 2016. Association for Computational Linguistics.

[6] Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. Stance classification of context-dependent claims. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 251–261, Valencia, Spain, April 2017. Association for Computational Linguistics.

[7] Roy Bar-Haim, Lilach Edelstein, Charles Jochim, and Noam Slonim. Improving claim stance classification with lexical knowledge expansion and context utilization. In *Proceedings of the 4th Workshop on Argument Mining*, pages 32–38, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[8] Roy Bar-Haim, Lilach Eden, Roni Friedman, Yoav Kantor, Dan Lahav, and Noam Slonim. From arguments to key points: Towards automatic argument summarization, 2020.

[9] Roy Bar-Haim, Yoav Kantor, Lilach Eden, Roni Friedman, Dan Lahav, and Noam Slonim. Quantitative argument summarization and beyond: Cross-domain key point analysis, 2020.

[10] Roy Bar-Haim, Dalia Krieger, Orith Toledo-Ronen, Lilach Edelstein, Yonatan Bilu, Alon Halfon, Yoav Katz, Amir Menczel, Ranit Aharonov, and Noam Slonim. From surrogacy to adoption; from bitcoin to cryptocurrency: Debate topic expansion. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 977–990, Florence, Italy, July 2019. Association for Computational Linguistics.

[11] Yonatan Bilu, Ariel Gera, Daniel Hershcovich, Benjamin Sznajder, Dan Lahav, Guy Moshkowich, Anael Malet, Assaf Gavron, and Noam Slonim. Argument invention from first principles. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1013–1026, Florence, Italy, July 2019. Association for Computational Linguistics.

[12] Yonatan Bilu, Daniel Hershcovich, and Noam Slonim. Automatic claim negation: Why, how and when. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 84–93, Denver, CO, June 2015. Association for Computational Linguistics.

[13] Teresa Botschen, Daniil Sorokin, and Iryna Gurevych. Frame- and entity-based knowledge for common-sense argumentative reasoning. In *Proceedings of the 5th Workshop on Argument Mining*, pages 90–96, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[14] Lucas Carstens and Francesca Toni. Towards relation based argumentation mining. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 29–34, Denver, CO, June 2015. Association for Computational Linguistics.

[15] Tuhin Chakrabarty, Christopher Hidey, Smaranda Muresan, Kathy McKeown, and Alyssa Hwang. AMPERSAND: Argument mining for PERSuAsive oNline discussions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2933–2943, Hong Kong, China, November 2019. Association for Computational Linguistics.

[16] Lisa Andreevna Chalaguine and Claudia Schulz. Assessing convincingness of arguments in online debates with limited number of features. In *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 75–83, Valencia, Spain, April 2017. Association for Computational Linguistics.

[17] S. Concannon, P. Healey, and Matthew Purver. How natural is argument in natural dialogue? In *CMNA@IJCAI*, 2016.

[18] Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. Neural end-to-end learning for computational argumentation mining. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11–22, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[19] L. Ein-Dor, Eyal Shnarch, Lena Dankin, Alon Halfon, B. Sznajder, Ariel Gera, C. Alzate, Martin Gleize, Leshem Choshen, Yufang Hou, Yonatan Bilu, Ranit Aharonov, and N. Slonim. Corpus wide argument mining - a working solution. *ArXiv*, abs/1911.10763, 2020.

[20] Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, Manfred Stede, and Benno Stein. Computational argumentation synthesis as a language modeling task. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 54–64, Tokyo, Japan, October–November 2019. Association for Computational Linguistics.

[21] Beatríz Fisas, Horacio Saggion, and Francesco Ronzano. On the discoursive structure of computer graphics research papers. In *LAW@NAACL-HLT*, 2015.

[22] Andrea Galassi, Kristian Kersting, Marco Lippi, Xiaoting Shao, and Paolo Torroni. Neural-symbolic argumentation mining: An argument in favor of deep learning and reasoning. *Frontiers in Big Data*, 2, Jan 2020.

[23] Martin Gleize, Eyal Shnarch, Leshem Choshen, Lena Dankin, Guy Moshkowich, Ranit Aharonov, and Noam Slonim. Are you convinced? choosing the more convincing evidence with a Siamese network. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 967–976, Florence, Italy, July 2019. Association for Computational Linguistics.

[24] Nancy Green. Proposed method for annotation of scientific arguments in terms of semantic relations and argument schemes. In *Proceedings of the 5th Workshop on Argument Mining*, pages 105–110, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[25] Shai Gretz, Yonatan Bilu, Edo Cohen-Karlik, and Noam Slonim. The workweek is the best time to start a family – a study of GPT-2 based claim generation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 528–544, Online, November 2020. Association for Computational Linguistics.

[26] Shai Gretz, Roni Friedman, Edo Cohen-Karlik, Assaf Toledo, Dan Lahav, Ranit Aharonov, and Noam Slonim. A large-scale dataset for argument quality ranking: Construction and analysis, 2019.

[27] Yunfan Gu, Zhongyu Wei, Maoran Xu, Hao Fu, Yang Liu, and Xuanjing Huang. Incorporating topic aspects for online comment convincingness evaluation. In *Proceedings of the 5th Workshop on Argument Mining*, pages 97–104, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[28] Ivan Habernal and Iryna Gurevych. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional LSTM. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1589–1599, Berlin, Germany, August 2016. Association for Computational Linguistics.

[29] Ivan Habernal and Iryna Gurevych. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43(1):125–179, April 2017.

[30] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1930–1940, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[31] Shohreh Haddadan, Elena Cabrio, and Serena Villata. Yes, we can! mining arguments in 50 years of US presidential campaign debates. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4684–4690, Florence, Italy, July 2019. Association for Computational Linguistics.

[32] Christopher Hidey and Kathy McKeown. Fixed that for you: Generating contrastive claims with semantic edits. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1756–1767, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[33] Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[34] Xinyu Hua, Zhe Hu, and Lu Wang. Argument generation with retrieval, planning, and realization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2661–2672, Florence, Italy, July 2019. Association for Computational Linguistics.

[35] Xinyu Hua and Lu Wang. Neural argument generation augmented with externally retrieved evidence. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 219–230, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[36] Xinyu Hua and Lu Wang. Sentence-level content planning and style specification for neural text generation, 2019.

[37] Khalid Al Khatib, Yufang Hou, Henning Wachsmuth, Charles Jochim, F. Bonin, and B. Stein. End-to-end argumentation knowledge graph construction. In *AAAI*, 2020.

[38] Neema Kotonya and Francesca Toni. Gradual argumentation evaluation for stance aggregation in automated fake news detection. In *Proceedings of the 6th Workshop on Argument Mining*, pages 156–166, Florence, Italy, August 2019. Association for Computational Linguistics.

[39] Anne Lauscher, Goran Glavaš, and Kai Eckert. ArguminSci: A tool for analyzing argumentation and rhetorical aspects in scientific writing. In *Proceedings of the 5th Workshop on Argument Mining*, pages 22–28, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[40] Anne Lauscher, Goran Glavaš, and Simone Paolo Ponzetto. An argument-annotated corpus of scientific publications. In *Proceedings of the 5th Workshop on Argument Mining*, pages 40–46, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[41] Anne Lauscher, Lily Ng, Courtney Napoles, and Joel Tetreault. Rhetoric, logic, and dialectic: Advancing theory-based argument quality assessment in natural language processing, 2020.

[42] Tamar Lavee, Matan Orbach, Lili Kotlerman, Yoav Kantor, Shai Gretz, Lena Dankin, Michal Jacovi, Yonatan Bilu, Ranit Aharonov, and Noam Slonim. Towards effective rebuttal: Listening comprehension using corpus-wide claim mining. In *Proceedings of the 6th Workshop on Argument Mining*, pages 58–66, Florence, Italy, August 2019. Association for Computational Linguistics.

[43] Dieu Thu Le, Cam-Tu Nguyen, and Kim Anh Nguyen. Dave the debater: a retrieval-based and generative argumentative dialogue agent. In *Proceedings of the 5th Workshop on Argument Mining*, pages 121–130, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[44] Ran Levy, Shai Gretz, Benjamin Sznajder, Shay Hummel, Ranit Aharonov, and Noam Slonim. Unsupervised corpus–wide claim detection. In *Proceedings of the 4th Workshop on Argument Mining*, pages 79–84, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[45] Davide Liga. Argumentative evidences classification and argument scheme detection using tree kernels. In *Proceedings of the 6th Workshop on Argument Mining*, pages 92–97, Florence, Italy, August 2019. Association for Computational Linguistics.

[46] Jian-Fu Lin, Kuo Yu Huang, Hen-Hsen Huang, and Hsin-Hsi Chen. Lexicon guided attentive neural network model for argument mining. In *Proceedings of the 6th Workshop on Argument Mining*, pages 67–73, Florence, Italy, August 2019. Association for Computational Linguistics.

[47] Anna Lindahl, Lars Borin, and Jacobo Rouces. Towards assessing argumentation annotation - a first step. In *Proceedings of the 6th Workshop on Argument Mining*, pages 177–186, Florence, Italy, August 2019. Association for Computational Linguistics.

[48] Luca Lugini and Diane Litman. Argument component classification for classroom discussions. In *Proceedings of the 5th Workshop on Argument Mining*, pages 57–67, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[49] Anastasios Lytos, T. Lagkas, P. Sarigiannidis, and Kalina Bontcheva. The evolution of argumentation mining: From models to social media and emerging tools. *Inf. Process. Manag.*, 56, 2019.

[50] Tobias Mayer, Elena Cabrio, and Serena Villata. Transformer-based Argument Mining for Healthcare Applications. In *ECAI 2020 - 24th European Conference on Artificial Intelligence*, Santiago de Compostela / Online, Spain, August 2020.

[51] Gaku Morio and Katsuhide Fujita. End-to-end argument mining for discussion threads based on parallel constrained pointer architecture. In *Proceedings of the 5th Workshop on Argument Mining*, pages 11–21, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[52] Elena Musi, Debanjan Ghosh, and Smaranda Muresan. Towards feasible guidelines for the annotation of argument schemes. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 82–93, Berlin, Germany, August 2016. Association for Computational Linguistics.

[53] Elena Musi, Manfred Stede, Leonard Kriese, Smaranda Muresan, and Andrea Rocci. A multilayer annotated corpus of argumentative text: From argument schemes to discourse relations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).

[54] Ivan Namor, Pietro Totis, Samuele Garda, and Manfred Stede. Mining italian short argumentative texts. In *CLiC-it*, 2019.

[55] Huy Nguyen and Diane Litman. Context-aware argumentative relation mining. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1127–1137, Berlin, Germany, August 2016. Association for Computational Linguistics.

[56] Christopher Olshefski, Luca Lugini, Ravneet Singh, Diane Litman, and Amanda Godley. The discussion tracker corpus of collaborative argumentation. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1033–1043, Marseille, France, May 2020. European Language Resources Association.

[57] Juri Opitz and Anette Frank. Dissecting content and context in argumentative relation analysis. In *Proceedings of the 6th Workshop on Argument Mining*, pages 25–34, Florence, Italy, August 2019. Association for Computational Linguistics.

[58] Marco Passon, Marco Lippi, Giuseppe Serra, and Carlo Tasso. Predicting the usefulness of Amazon reviews using off-the-shelf argumentation mining. In *Proceedings of the 5th Workshop on Argument Mining*, pages 35–39, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[59] Andreas Peldszus and Manfred Stede. An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon*, volume 2, pages 801–815, 2015.

[60] Isaac Persing and Vincent Ng. End-to-end argumentation mining in student essays. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1384–1394, San Diego, California, June 2016. Association for Computational Linguistics.

[61] Peter Potash, Adam Ferguson, and Timothy J. Hazen. Ranking passages for argument convincingness. In *Proceedings of the 6th Workshop on Argument Mining*, pages 146–155, Florence, Italy, August 2019. Association for Computational Linguistics.

[62] Peter Potash, Alexey Romanov, and Anna Rumshisky. Here's my point: Joint pointer architecture for argument mining. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1364–1373, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[63] C. Reed, Katarzyna Budzynska, R. Duthie, M. Janier, Barbara Konat, John Lawrence, A. Pease, and M. Snaith. The argument web: an online ecosystem of tools, systems and services for argumentation. *Philosophy & Technology*, 30:137 – 160, 2017.

[64] Nils Reimers, Benjamin Schiller, Tilman Beck, Johannes Daxenberger, Christian Stab, and Iryna Gurevych. Classification and clustering of arguments with contextualized word embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 567–578, Florence, Italy, July 2019. Association for Computational Linguistics.

[65] Paul Reisert, Naoya Inoue, Tatsuki Kuribayashi, and Kentaro Inui. Feasible annotation scheme for capturing policy argument reasoning using argument templates. In *Proceedings of the 5th Workshop on Argument Mining*, pages 79–89, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[66] Gil Rocha, Henrique Lopes Cardoso, and Jorge Teixeira. ArgMine: A framework for argumentation mining. In *Proceedings of the International Conference on Computational Processing of the Portuguese Language*, Tomar, Portugal, 2016.

[67] Gil Rocha, Christian Stab, Henrique Lopes Cardoso, and Iryna Gurevych. Cross-lingual argumentative relation identification: from English to Portuguese. In *Proceedings of the 5th Workshop on Argument Mining*, pages 144–154, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[68] Allen Roush and Arvind Balaji. DebateSum: A large-scale argument mining and summarization dataset. In *Proceedings of the 7th Workshop on Argument Mining*, pages 1–7, Online, December 2020. Association for Computational Linguistics.

[69] Benjamin Schiller, Johannes Daxenberger, and Iryna Gurevych. Aspect-controlled neural argument generation, 2020.

[70] Eyal Shnarch, Ran Levy, Vikas Raykar, and Noam Slonim. GRASP: Rich patterns for argumentation mining. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1345–1350, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[71] Edwin Simpson and Iryna Gurevych. Finding convincing arguments using scalable Bayesian preference learning. *Transactions of the Association for Computational Linguistics*, 6:357–371, 2018.

[72] Edwin Simpson and Iryna Gurevych. A bayesian approach for sequence tagging with crowds, 2019.

[73] Maria Skeppstedt, Andreas Peldszus, and Manfred Stede. More or less controlled elicitation of argumentative text: Enlarging a microtext corpus via crowdsourcing. In *Proceedings of the 5th Workshop on Argument Mining*, pages 155–163, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[74] Christian Stab, Johannes Daxenberger, Chris Stahlhut, Tristan Miller, Benjamin Schiller, Christopher Tauchmann, Steffen Eger, and Iryna Gurevych. ArgumenText: Searching for arguments in heterogeneous sources. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 21–25, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[75] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3):619–659, September 2017.

[76] Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. Cross-topic argument mining from heterogeneous sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[77] Assaf Toledo, Shai Gretz, Edo Cohen-Karlik, Roni Friedman, Elad Venezian, Dan Lahav, Michal Jacovi, Ranit Aharonov, and Noam Slonim. Automatic argument quality assessment - new datasets and methods. *CoRR*, abs/1909.01007, 2019.

[78] Orith Toledo-Ronen, Matan Orbach, Yonatan Bilu, Artem Spector, and Noam Slonim. Multilingual argument mining: Datasets and analysis, 2020.

[79] Benedetta Torsi and Roser Morante. Annotating claims in the vaccination debate. In *Proceedings of the 5th Workshop on Argument Mining*, pages 47–56, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[80] Dietrich Trautmann. Aspect-based argument mining. In *Proceedings of the 7th Workshop on Argument Mining*, pages 41–52, Online, December 2020. Association for Computational Linguistics.

[81] Dietrich Trautmann, Johannes Daxenberger, Christian Stab, Hinrich Schütze, and Iryna Gurevych. Fine-grained argument unit recognition and classification, 2019.

[82] Henning Wachsmuth, Manfred Stede, Roxanne El Baff, Khalid Al-Khatib, Maria Skeppstedt, and Benno Stein. Argumentation synthesis following rhetorical strategies. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3753–3765, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[83] Thiemo Wambsganss, Nikolaos Molyndris, and M. Söllner. Unlocking transfer learning in argumentation mining: A domain-independent modelling approach. In *Wirtschaftsinformatik*, 2020.

[84] Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. Al: An adaptive learning support system for argumentation skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[85] Hao Wang, Zhen Huang, Yong Dou, and Yu Hong. Argumentation mining on essays at multi scales. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5480–5493, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.