Argument Mining for Scholarly Document Processing: Taking Stock and Looking Ahead

Khalid Al-Khatib*
Leipzig University, Germany
khalid.alkhatib@uni-leipzig.de

Tirthankar Ghosal*
Charles University, MFF, ÚFAL, CZ
ghosal@ufal.mff.cuni.cz

Yufang Hou
IBM Research, Ireland
yhou@ie.ibm.com

Anita de Waard
Elsevier, USA
a.dewaard@elsevier.com

Dayne Freitag
SRI International, USA
freitag@ai.sri.com

Abstract

Argument mining targets structures in natural language related to interpretation and persuasion. Most scholarly discourse involves interpreting experimental evidence and attempting to persuade other scientists to adopt the same conclusions, which could benefit from argument mining techniques. However, While various argument mining studies have addressed student essays and news articles, those that target scientific discourse are still scarce. This paper surveys existing work in argument mining of scholarly discourse, and provides an overview of current models, data, tasks, and applications. We identify a number of key challenges confronting argument mining in the scientific domain, and suggest some possible solutions and future directions.

1 Introduction

Scientific papers aim to present verifiable evidence for a series of stated claims, anchoring these claims in experiments, data, and references. However, the interpretation of such objective sources of evidence is often ambiguous and subjective. Thus, much of scientific communication is essentially persuasive and uses an argumentative structure to establish the relevance, validity, and novelty of an author's main claims and conclusions (Pelclova and Weilun, 2018). This argumentation takes the form of a dialogue between the author and her readers, in which new knowledge is proposed and an attempt made to persuade the readers to accept and follow particular claims (Fahy, 2008; Hyland, 2014). However, most current research on automatic document processing ignores this argumentative context and treats statements that are persuasive, tentative, or speculative to be factual. This risks overstating the certainty of claims and hypotheses, and bypasses the rhetorical aspect of scientific discourse (see e.g. (Gross and Chesley, 2012)).

Computational argumentation is a recent and growing field of research concerned with the computational analysis and generation of natural language arguments and argumentative discourses. Over the past decade, this area has attracted researchers seeking to tackle different tasks including argument mining, argument quality assessment, and argument generation (for an overview, see e.g. (Stede et al., 2018)). The most studied task is *argument mining*, i.e., the identification of argumentative units, argument components (e.g., conclusion and premise), and structures of text documents. However, despite a wealth of Natural Language Processing (NLP) research on extracting information from scientific literature—including entity extraction (Augenstein et al., 2017; Hou et al., 2019), relation identification (Luan et al., 2018), question answering (Demner-Fushman and Lin, 2007), and summarization (Erera et al., 2019)—relatively few attempts have been made to model argumentative structures in science.

This paper argues for an increased focus of the NLP community on argument mining in scientific documents. To encourage work at the intersection of Scholarly Discourse Processing and Argument Mining, we provide a brief overview of current work in this field, and discusses the most used models, data, methods, and applications. We discuss a number of challenges in mining the argumentative structure of scientific documents and propose some promising future directions.

2 Argumentation in Scientific Discourses

To support future efforts on argument mining of scientific documents, we present a survey of the literature from 2000 to the present, summarized in Table 1 in the Appendix. To attempt to create a somewhat comprehensive overview, we concentrated on papers published by the NLP commu-

*These authors contributed equally

Proceedings of the Second Workshop on Scholarly Document Processing, pages 56–65

nity[2]. To obtain this list, we used Google Scholar (https://scholar.google.com/) to find papers on "Argumentation Mining on Scientific Papers", "Argumentation Mining on Research Papers", and "Argumentative Zoning on Scientific Papers". We also traced the references of some pivotal papers from the proceedings of Argument Mining workshops[3].

For each paper, we identified the *Domain* of study (i.e., a specific scientific domain, full-text or abstracts), the *Objectives* of the work, and the *Methods* used. Furthermore, the papers can be categorized under four areas of study, discussed, in turn, below.

Corpus Creation and New Annotation Schemes A number of studies propose an annotation scheme for mining argumentative discourse in the science domain. Many of these studies follow the well-known argumentation model of Toulmin (Toulmin, 1958). Toulmin's model targets the structure of an argument, modelling it as a *claim* that is supported by *data* following some *warrants*, which can be supported by *backing*. The model has also two optional components: *qualifiers* and *rebuttals*.

Examples of the studies that adopt Toulmin's model are Green (2014) and Lauscher et al. (2018b). The former proposes the scheme of premise (i.e., data and warrant) and conclusion. The latter's scheme includes background claim, own claim, and data, which is used to annotate 40 publications from computer graphics.

Another model that is often used is that of *argumentation schemes* (Walton et al., 2008). Argumentation schemes target the structure of an argument, where the argument is modeled as a set of propositions, i.e., a conclusion and one or more premises, with a pattern that manifests the logical inference between the conclusion and its premise. Walton et al. (2008) proposed around 60 different schemes including 'argument from cause to effect' and 'argument from example', among others. An example of this approach is Green (2015a), where ten schemes were selected and annotated in a corpus of biomedical genetics articles.

Other studies focus on identifying argumentative discourse roles, especially *argumentative zones* (Teufel and Moens, 2002), assigning roles such as 'aim' and 'background' to large text spans (usually paragraphs). Following this approach, several corpora have been constructed for biomedical papers (Guo et al., 2011), as well as papers in chemistry, computational linguistics (Yang and Li, 2018), and agriculture (Teufel, 2014).

Inspired by the theory of Freeman (2011), some studies annotate the argumentative relations between arguments. For instance, Lauscher et al. (2018a) consider the relations of 'support', 'contradicts', and 'same claim'. Kirschner et al. (2015), in another study, consider the relations of 'support', 'attack', 'detail', and 'sequence', which were annotated in 24 articles belong to the domain of educational and developmental.

Automatic Argument Unit Identification Much work in argument mining focuses on identifying *Argumentative Discourse Units* (ADUs). An ADU is a text span that plays a specific role in an argument. In this way, argument unit identification resembles named entity recognition or discourse segment type identification. Green (2017b) extracted argumentative units from biomedical and biological articles using a semantic rule-based approach. Lauscher et al. (2018a) and Lauscher et al. (2018c) proposed several neural multi-task learning models based on Bi-LSTM to identify premises and conclusions. Other papers propose different approaches to identify argumentative zones, including supervised and weakly-supervised approaches with a rich set of linguistics features (e.g., (Guo et al., 2011)). Identifying the 'claim' unit is tackled in several papers such as Achakulvisut et al. (2019), which employs transfer learning on top of a discourse tagging model using a pre-trained BilSTM-CRF to identify claims in biomedical abstracts. Extracting 'evidence' has been tackled in other studies, e.g. Li et al. (2019) extracted evidence in biomedical publications with sentence-level sequential labelling, using BiLSTM-CRF and attention.

Automatic Argument Structure Identification If unit identification resembles entity recognition, *argument structure identification* is akin to relation extraction: this work aims to find typed relationships between ADUs. This more challenging task has been addressed by relatively few studies: Accuosto and Saggion (2020) extend existing discourse parsing models to address this problem on

[2]In this paper, we focus our research on papers related to argument mining for scholarly document processing and exclude less central topics such as citation analysis: we hope that future scholars can help augment our work with these and similar related approaches

[3]See https://2021.argmining.org/ and links from there for a full list of past workshops

computational linguistics abstracts and identify the argumentative discourses of computational linguistics abstracts using lexical and ELMo embeddings, while Song et al. (2019) analyze the argument structure of information science and biomedical science articles through sequential pattern mining.

Applications To date, much of the application-oriented work has focused on scientific article summarization. An exception is Feltrim and Teufel (2004), which had the goal of developing tools for scientific writing for the computer science domain. Other efforts aim to identify claims and evidence, to enable claim-evidence based representations of collections of documents, such as (de Waard et al., 2009), (Groza et al., 2011) and (Li et al., 2021). The goal here is to allow the reader to traverse the reasoning behind a scientific claim to either experimental evidence in the paper itself, or to reasoning for data provided in cited papers. Recently, Yu et al. (2020) study the problem of correlation-to-causation exaggeration in press releases by comparing claims made in news articles and the corresponding scientific papers.

3 Challenges

In this section, we describe a few challenges that are relevant to argument mining in the scientific literature. Although not only specific for the scientific domain, these are hurdles that need to be faced in future research to allow progress to be made.

Argumentation Modeling As described above, various argument models have been proposed (Stede et al., 2018). The selection of which model fits scientific documents is a crucial and challenging research question.

Most previous studies in argument mining of scientific documents utilize either Toulmin's model or argumentation schemes. However, none of these models seems to be a perfect fit: Toulmin's warrants and rebuttals are not common to scholarly argumentation[4], and none of the other argument schemes take the specific nature of scholarly argumentation into account. Adapting these models for use seems to be an essential step to achieve feasible annotation and identification of argument structures in scholarly discourse.

[4]For example, Lauscher et al. (2018b) conducted an expert annotation of the argumentative structures of a small set of scientific publications based on Toulmin's model. The annotation results show that warrant, backing, qualifier, and rebuttal are not observed in the publications.

Domain Knowledge Science communication encompasses a variety of domains, topics, and methodologies organized into research communities, each following its own standards regarding the structuring of documents and the arguments they contain (Weinstein, 1990). These community conventions present a barrier to understanding for non-specialists and computational models alike. An important open question, therefore, is whether argument mining techniques must be tailored to individual scientific communities, or whether a unified model can be adapted to address domain-specific features of scientific argumentation.

Scientific Document Type Scientific communication involves a variety of document types, including reviews, methods papers, and experimental reports, among others [5]. Each type concentrates on specific aspects of the discussed topic and usually provides particular types of evidence.

Analogous to the previous point, an open question is whether different document types require different models, or whether they can be accommodated by a single representation and modeling approach tailored to different argument structures.

Enthymemes An enthymeme is the implicit (unstated) premise or conclusion in an argument. Because enthymemes are supposedly known by the target audience (or easily constructed using common knowledge), enthymeme are rarely a problem for humans. However, to the extent that shared knowledge is required which is not found in the document, this offers a challenge for argument mining techniques.

As an example, Green (2014) conducted a manual inspection of several arguments in the biomedical genetics research literature, showing that arguments with enthymemes are common there and suggested explicitly providing domain knowledge for reconstructing enthymemes.

Subjective Interpretation A common dilemma in argument mining is that an argumentative text may have multiple valid interpretations of its structure. This is a concern for scientific documents, where the connection between a claim and its evidence can be implicit, i.e., the author leaves this connection to the readers' interpretations.

In particular, experimental papers can follow a line of reasoning that makes e.g. 'biological sense',

[5]For more examples of the types, see https://coling2018.org/index.html%3Fp=156.html

i.e. where a specific experiment follows another experiment to address a potential alternate interpretation of the previous experiment. For a non-biologist, this reasoning is unclear, and the reason for these subsequent results are generally never explicitly stated in the text.

Context-Dependence Context plays a key role in text mining in general and argument mining in particular. Scientific documents are at least as complex as other genres where argument plays a role, such as persuasive essays, to fulfil both the persuasive role and the presentation of objectivity which scientific writing demands (Vazquez Orta and Giner, 2009-11). More specifically, selecting the optimal boundaries of argumentative units in scientific documents is known to be challenging (Green, 2014; Stab et al., 2014). For instance, the distance between a claim and its premise may be particularly wide in scientific discourse, e.g., the claim which is stated in one section can be supported by a premise in a different section.

4 Discussion

In summary, we have provided a brief overview of current work and a summary of issues that need to be addressed to make headway in the automated argument mining for scholarly documents. We hope to have shown that more research is needed in this field to enable better representation of the persuasive aspects of scholarly communication. This can help provide a more realistic representation of how scientific knowledge is obtained, and how authors aim to persuade readers of the validity of claims. In particular, seeing scholarly discourse as a pragmatic discourse, i.e. one that humans undertake with interpersonal, as well as informative goals, can allow richer representations of the knowledge structures underlying scientific progress.

As noted, applications of argument mining in scientific discourse, such as summarization and aids to technical writing, to date have been limited to those that are relatively robust to errors, a partial consequence of the immaturity of the field. In particular, these applications are mostly insensitive to the *factual* content of scientific arguments. Meanwhile, a relatively mature community continues to expand models and methods for information extraction in various scientific domains, usually with no attention to the argumentative context in which the target facts are presented. Because a correct understanding and use of facts is critical to scientific understanding and progress, we see an opportunity for many innovative applications at the intersection of fact and argument. For example, models capable of determining the *salience* of individual facts in a domain could provide the basis for highly precise forms of scientific information retrieval, or even offer forms of automation that assist scientists in maximizing the pertinence of their experiments.

To achieve this vision at scale, the argument mining community must grapple with the problem of increasing scientific domain specialization. It is crucial that we separate the invariant features of scientific argumentation from those that vary with field and specialization, and that we investigate effective methods of cross-domain transfer. To this end, the field should seek consensus regarding how scientific argumentation should be formalized and strive for broad-coverage reference corpora annotated under guidelines optimized for high inter-annotator agreement.

To support these efforts, we suggest a greater collaboration between participants of the scholarly document processing and argument mining domains, with a particular focus on creating shared models and shared and accessible corpora to spur on research. We hope such conversations can commence at this workshop and others, to inspire and unite members of both communities with natural language processing and improve sharing and improving the outputs of science and scholarship.

5 Conclusion

This paper endeavors at promoting the collaboration between the communities of scholarly discourse processing and computational argumentation, arguing for the ultimate importance of more extensive research on *argument mining in scientific documents*. Particularly, we address the current contributions on argument mining for scientific documents by surveying about 40 papers that approach different aspects and tasks such as proposing annotation schemes, creating corpora, and identifying argumentative discourse units as well as argumentative relations in scientific documents. Furthermore, we describe various challenges for mining argumentative structures of scientific documents and suggest some strategic directions in order to accomplish remarkable benefits on a wide range of downstream applications such as scientific writing assistance, scientific articles summarization, and quality assessment.

References

Pablo Accuosto and Horacio Saggion. 2019. Transferring knowledge from discourse to arguments: A case study with scientific abstracts. In *Stein B, Wachsmuth H, editors. Proceedings of the 6th Workshop on Argument Mining; 2019 Aug 1; Florence, Italy. Stroudsburg: Association for Computational Linguistics; 2019. p. 41-51.* ACL (Association for Computational Linguistics).

Pablo Accuosto and Horacio Saggion. 2020. Mining arguments in scientific abstracts with discourse-level embeddings. *Data & Knowledge Engineering*, 129:101840.

Titipat Achakulvisut, Chandra Bhagavatula, Daniel Acuna, and Konrad Kording. 2019. Claim extraction in biomedical publications using deep discourse model and transfer learning. *arXiv preprint arXiv:1907.00962*.

Mohammed Alliheedi, Robert E Mercer, and Robin Cohen. 2019. Annotation of rhetorical moves in biochemistry articles. In *Proceedings of the 6th Workshop on Argument Mining*, pages 113–123.

Isabelle Augenstein, Mrinal Das, Sebastian Riedel, Lakshmi Vikraman, and Andrew McCallum. 2017. SemEval 2017 task 10: ScienceIE - extracting keyphrases and relations from scientific publications. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 546–555, Vancouver, Canada. Association for Computational Linguistics.

Danish Contractor, Yufan Guo, and Anna Korhonen. 2012. Using argumentative zones for extractive summarization of scientific articles. In *Proceedings of COLING 2012*, pages 663–678.

Dina Demner-Fushman and Jimmy Lin. 2007. Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, 33(1):63–103.

Shai Erera, Michal Shmueli-Scheuer, Guy Feigenblat, Ora Peled Nakash, Odellia Boni, Haggai Roitman, Doron Cohen, Bar Weiner, Yosi Mass, Or Rivlin, Guy Lev, Achiya Jerbi, Jonathan Herzig, Yufang Hou, Charles Jochim, Martin Gleize, Francesca Bonin, Francesca Bonin, and David Konopnicki. 2019. A summarization system for scientific documents. 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): System Demonstrations*, pages 211–216, Hong Kong, China. Association for Computational Linguistics.

Kathleen Fahy. 2008. Writing for publication: Argument and evidence. *Women and Birth*, 21(3):113–117.

Syeed Ibn Faiz and Robert E Mercer. 2014. Extracting higher order relations from biomedical text. In *Proceedings of the First Workshop on Argumentation Mining*, pages 100–101.

Valéria D Feltrim and Simone Teufel. 2004. Automatic critiquing of novices' scientific writing using argumentative zoning. In *Proc. AAAI spring symposium exploring affect and attitude in text*.

Valéria D Feltrim, Simone Teufel, Maria Graças V das Nunes, and Sandra M Aluísio. 2006. Argumentative zoning applied to critiquing novices' scientific abstracts. In *Computing Attitude and Affect in Text: Theory and Applications*, pages 233–246. Springer.

J. Freeman. 2011. Argument structure: Representation and theory. In *Argumentation Library*.

Heather Graves, Roger Graves, Robert E Mercer, and Mahzereen Akter. 2014. Titles that announce argumentative claims in biomedical research articles. In *Proceedings of the First Workshop on Argumentation Mining*, pages 98–99.

Nancy Green. 2014. Towards creation of a corpus for argumentation mining the biomedical genetics research literature. In *Proceedings of the first workshop on argumentation mining*, pages 11–18.

Nancy Green. 2015a. Identifying argumentation schemes in genetics research articles. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 12–21.

Nancy Green. 2017a. Manual identification of arguments with implicit conclusions using semantic rules for argument mining. In *Proceedings of the 4th Workshop on Argument Mining*, pages 73–78.

Nancy Green. 2018a. 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.

Nancy L Green. 2015b. Annotating evidence-based argumentation in biomedical text. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 922–929. IEEE.

Nancy L Green. 2017b. Argumentation mining in scientific discourse. In *CMNA@ ICAIL*, pages 7–13.

Nancy L Green. 2018b. Towards mining scientific discourse using argumentation schemes. *Argument & Computation*, 9(2):121–135.

A. Gross and Paula Chesley. 2012. Hedging, stance and voice in medical research articles.

Tudor Groza, Siegfried Handschuh, and Stefan Decker. 2011. Capturing rhetoric and argumentation aspects within scientific publications. In *Journal on data semantics XV*, pages 1–36. Springer.

Yufan Guo, Anna Korhonen, and Thierry Poibeau. 2011. A weakly-supervised approach to argumentative zoning of scientific documents. In *Empirical Methods in Natural language Processing (EMNLP)*.

Yufan Guo, Ilona Silins, Roi Reichart, and Anna Korhonen. 2012. Crab reader: A tool for analysis and visualization of argumentative zones in scientific literature. In *Proceedings of COLING 2012: Demonstration Papers*, pages 183–190.

Yufang Hou, Charles Jochim, Martin Gleize, Francesca Bonin, and Debasis Ganguly. 2019. Identification of tasks, datasets, evaluation metrics, and numeric scores for scientific leaderboards construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5203–5213, Florence, Italy. Association for Computational Linguistics.

Hospice Houngbo and Robert E Mercer. 2014. An automated method to build a corpus of rhetorically-classified sentences in biomedical texts. In *Proceedings of the first workshop on argumentation mining*, pages 19–23.

Ken Hyland. 2014. Introductory chapter:dialogue, community and persuasion in research writing. In Luz Gil-Salom and Carmen Soler-Monreal, editors, *Dialogicity in Written Specialised Genres*, Dialogicity in Written Specialised Genres, pages 1–20. John Benjamins.

Christian Kirschner, Judith Eckle-Kohler, and Iryna Gurevych. 2015. Linking the thoughts: Analysis of argumentation structures in scientific publications. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 1–11.

Anne Lauscher, Goran Glavaš, and Kai Eckert. 2018a. Arguminsci: A tool for analyzing argumentation and rhetorical aspects in scientific writing. Association for Computational Linguistics.

Anne Lauscher, Goran Glavaš, and Simone Paolo Ponzetto. 2018b. An argument-annotated corpus of scientific publications. In *Proceedings of the 5th Workshop on Argument Mining*, pages 40–46.

Anne Lauscher, Goran Glavaš, Simone Paolo Ponzetto, and Kai Eckert. 2018c. Investigating the role of argumentation in the rhetorical analysis of scientific publications with neural multi-task learning models. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3326–3338.

Xiangci Li, Gully Burns, and Nanyun Peng. 2019. Scientific discourse tagging for evidence extraction. *arXiv e-prints*, pages arXiv–1909.

Xiangci Li, Gully Burns, and Nanyun Peng. 2021. Scientific discourse tagging for evidence extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2550–2562, Online. Association for Computational Linguistics.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3219–3232, Brussels, Belgium. Association for Computational Linguistics.

Jana Pelclova and Lu Wei-lun. 2018. Persuasion in public discourse: cognitive and functional perspectives. *Discourse approaches to politics, society, and culture*.

José María González Pinto, Serkan Celik, and Wolf-Tilo Balke. 2019. Learning to rank claim-evidence pairs to assist scientific-based argumentation. In *International Conference on Theory and Practice of Digital Libraries*, pages 41–55. Springer.

Ningyuan Song, Hanghang Cheng, Huimin Zhou, and Xiaoguang Wang. 2019. Argument structure mining in scientific articles: a comparative analysis. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 339–340. IEEE.

Christian Stab, Christian Kirschner, Judith Eckle-Kohler, and Iryna Gurevych. 2014. Argumentation mining in persuasive essays and scientific articles from the discourse structure perspective. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing, Forlì-Cesena, Italy, July 21-25, 2014*, volume 1341 of *CEUR Workshop Proceedings*. CEUR-WS.org.

M. Stede, J. Schneider, and G. Hirst. 2018. *Argumentation Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Manfred Stede, Maite Taboada, and Debopam Das. 2017. Annotation guidelines for rhetorical structure. *Manuscript. University of Potsdam and Simon Fraser University*.

Simone Teufel. 2014. Scientific argumentation detection as limited-domain intention recognition. In *ArgNLP*.

Simone Teufel and Marc Moens. 1999. Discourse-level argumentation in scientific articles: human and automatic annotation. In *Towards Standards and Tools for Discourse Tagging*.

Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational linguistics*, 28(4):409–445.

Stephen E. Toulmin. 1958. *The Uses of Argument*. Cambridge University Press.

Ignacio Vazquez Orta and Diana Giner. 2009-11. Writing with conviction: the use of boosters in modelling persuasion in academic discourses.

Anita de Waard, S Buckingham Shum, Annamaria Carusi, Jack Park, Matthias Samwald, and Ágnes Sándor. 2009. Hypotheses, evidence and relationships: The hyper approach for representing scientific knowledge claims.

Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.

Mark Weinstein. 1990. Towards an account of argumentation in science. *Argumentation*, 4(3):269–298.

An Yang and Sujian Li. 2018. SciDTB: Discourse dependency TreeBank for scientific abstracts. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 444–449, Melbourne, Australia. Association for Computational Linguistics.

Antonio Jimeno Yepes, James G Mork, and Alan R Aronson. 2013. Using the argumentative structure of scientific literature to improve information access. In *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing*, pages 102–110.

Bei Yu, Jun Wang, Lu Guo, and Yingya Li. 2020. Measuring correlation-to-causation exaggeration in press releases. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4860–4872, Barcelona, Spain (Online). International Committee on Computational Linguistics.

A Appendix

Please follow in the next page.

Table 1: Argumentation Mining Literature on Scientific Discourse

Reference	Domain	Objectives	Methods	Additional Contribution
Manual Argument Analysis				
Green (2015b)	Biomedical articles	Analyzed evidence based arguments in four full-text articles on genetic variants that may cause human health problems and created a preliminary catalog of argumentation schemes		
Green (2017a)	Biomedical articles	Evaluate human analysts' ability to identify the argumentation scheme and premises of an argument having an implicit conclusion		
Green (2018b)	Biomedical research articles	Explores how arguments in a research article occur within a narrative of scientific discovery and how they are related to each other		
Green (2018a)	Biomedical Genetics articles	Provide a method for semantic representation of arguments that can be used in empirical studies of scientific discourse as well as to support applications such as argument mining		
Graves et al. (2014)	Biomedical articles	Analyses article title as a potential source of claims and finds that frequency of verbs in titles of experimental research articles has increased over time		
Corpus Creation and New Annotation Schemes				
Green (2014)	Biomedical Genetics articles	Argument annotation scheme: Premise (Data, Warrant) and Conclusion		Theoretical challenges to create an argument corpora
Green (2015a)	Biomedical Genetics articles	Identification of argumentation schemes with specification of ten semantically distinct argumentation schemes		Annotation guidelines for argumentation corpora
Teufel and Moens (1999)	Chemistry, Computational Linguistics	Detect argument zones in scientific articles	Proposed a scheme and annotated 15 argument zone categories for 39 papers (5,374 sentences)	
Kirschner et al. (2015)	Scientific articles (Educational and Developmental Psychology)	New annotation scheme to identify argumentative relations - support, attack, detail, sequence		Study of the annotation strategy across 24 articles, an annotation tool, a new graph-based inter-annotation measure
Lauscher et al. (2018b)	Computer Graphics scientific publications	Proposed a new argument-annotated dataset of scientific publications	Adapted Toulmin's model for argumentative components: Background Claim, Own Claim, Data. Relation between argumentative components: support, contradicts, same claim	Investigation on link between argumentative nature of scientific publications and rhetorical aspects such as discourse categories or citation contexts.
Alliheedi et al. (2019)	Biochemistry articles	Determine rhetorical moves in the argument structure of biomedical articles	Annotated method sections of 105 text files based on a new annotation scheme for identifying the structured representation of knowledge in a set of sentences describing the experimental procedures	
Guo et al. (2012)	Biomedical papers	Introduce a tool for analysis and visualizing argument structure (based on AZ), and also facilitate expert AZ annotation	Used HTML, JavaScript, PHP, XML for the annotation tool; SVM classifier using features from Guo et al. (2011)	Interactive annotation via active learning; CRAB Reader allows user to define AZ schemes; AZ can be performed on each word, sentence, paragraph, document level
Yang and Li (2018)	Scientific abstracts from ACL Anthology	Construct a domain-specific discourse treebank annotated on scientific articles	798 segmented abstracts were labelled by 5 annotators in 6 months. 506 abstracts were annotated more than twice separately by different annotators. In total, SciDTB contains 798 unique abstracts with 63% labelled more than once and 18,978 discourse relations.	Provide several baselines for scientific discourse dependency tree parsing
Automatic Argument Unit Identification				
Green (2017b)	Biomedical, Biological articles	Argumentation extraction	Semantic rule-based approach	Demonstrates the need for a richer model of inter-argument relationships in biomedical/biological research articles.

Reference	Domain	Objectives	Methods	Additional Contribution
Lauscher et al. (2018a)	Computer Graphics scientific publications	A toolkit for rhetorical analysis of argument component identification, discourse role classification, subjective aspect classification, citation context classification, summary relevance classification	Token-level sequence labelling, sentence-level classification using Bi-lSTM	Command-line tool, RESTful API, web application
Lauscher et al. (2018c)	Computer Graphics scientific publications	Proposed two neural multi-task learning (MTL) models for argumentative analysis based on the tasks in (Lauscher et al., 2018a)	Bi-LSTM based simple MTL model for sentence-level classification, hierarchical MTL for sequence labelling	Adapted Toulmin's model for argumentative components: Background Claim, Own Claim, Data. Relation between argumentative components: support, contradicts, same claim
Teufel (2014)	Chemistry, Computational Linguistics, Agriculture	Views scientific argumentation detection as limited-domain intent recognition	Model based on recognition of 28 rhetorical moves in text	
Guo et al. (2011)	Biomedical abstracts	Investigating a weakly-supervised approach for AZ detection when a limited amount of training data is available	Features like location, word bi-gram, verb, verb cues, PoS, grammatical relations, subj/obj, voice are used with ASVM, ASSVM, TSVM, SSCRF	Conclusion that location of AZs are super important, directions to facilitate easy porting of AZ schemes to new NLP tasks and domains
Li et al. (2019)	Biomedical publications	Automatic evidence extraction using scientific discourse tagging based on classification by de Waard et al. (2009)	sentence-level sequential labelling using BiLSTM-CRF + Attention	Leveraging scientific discourse tagging for evidence fragment detection
Achakulvisut et al. (2019)	Biomedical abstracts	Automated claim extraction	Neural discourse tagging model based on a pre-trained BilSTM+CRF followed by transfer learning and fine tuning on a expert annotated dataset	New dataset of 1,500 expert-annotated biomedical abstracts indicating whether the sentence presents a scientific claim.
Houngbo and Mercer (2014)	Biomedical articles	Identify the components of IMRaD rhetorical structure in biomedical papers	Applied a few heuristics to construct a corpus and used machine learning techniques (Naive Bayes and SVM) to classify sentences into Method,Result or Conclusion	
Pinto et al. (2019)	Biomedical papers	Claim-evidence matching as a learning to rank problem where goal is to find evidence in the form of a paper to make a natural language claim appear credible; to assist scientific argumentation	Rhetoric Classification Task and Claim-Evidence Rank Task using NB-BoW, SVM-BoW, CNN on data from a Wikipedia dump with word2vec trained on PubMed Central UMLS, SemMedDB databases	Augmenting "prestige" meta-data features for a paper improved performance, to rank claim-evidence pairs, a model should account for other semantic properties beyond simple content-matching
Faiz and Mercer (2014)	Biomedical papers	Extraction of connections or "higher order relations" between *biomedical relations* (relationship between biomedical entities). The higher order relation conveys a causal sense, which indicates that the latter relation causes the earlier one.	In the first stage, the authors use a discourse relation parser to extract the explicit discourse relations from text. In the second stage, the authors analyze each extracted explicit discourse relation to determine whether it can produce a higher order relation.	Pilot evaluation on AIMed corpus for protein-protein interaction prediction: identify the full argument extent which contain the biomedical entities
Yepes et al. (2013)	MEDLINE/PubMed abstracts	An evaluation of several learning algorithms to label abstract text with argumentative labels, based on structured abstracts available in MEDLINE/PubMed	Naive Bayes, SVM, Logistic Legression, CRF, AdaBoostM1 as classifiers for the argumentation labels on abstract text. In addition to textual features, the position of the sentence or paragraph from the beginning of the abstract is used	A data set to compare and evaluate GeneRIF indexing approaches. The sentence annotation are: Expression, Function, Isolation, Non-GeneRIF, Other, Reference, and Structure on MEDLINE articles.
Automatic Argument Structure Identification				
Stab et al. (2014)	Scientific articles	Identification of argumentation structures	Argument unit identification and relation extraction	An evaluation dataset of 20 scientific full-texts annotated with argument relations 'support', 'attack', 'sequence'
Feltrim et al. (2006)	Brazilian PhD Theses	A system to detect argumentative structures in text	The annotation scheme has the following rhetorical categories: Background, Gap, Purpose, Methodology, Results, Conclusion and Outline. A Naive Bayes classifier to identify the argumentative units	Porting of Argumentative Zoning (AZ) from English to Portuguese. A pilot system to demonstrate the effectiveness of AZ for a critiquing tool to support academic writing
Accuosto and Saggion (2020)	Computational linguistics abstracts	Argument unit identification and relation extraction	Explore two transfer learning approaches in which discourse parsing is used as an auxiliary task when training argument mining models	Propose a new annotation schema and use it to augment a corpus of computational linguistics abstracts that had previously been annotated with discourse units and relations
Song et al. (2019)	Information Science and Biomedical articles	Apply sequential pattern mining to analyse the common argument structure in two scientific domains (Information science and biomedical science)		

Reference	Domain	Objectives	Methods	Additional Contribution
Applications				
Accuosto and Saggion (2019)	Computational Linguistics abstracts	Leverage existing discourse parsing RST annotations (Stede et al., 2017) to identify argumentative components and relations	Transfer learning to improve the performance of argument mining tasks trained with a small corpus of 60 abstracts by leveraging the discourse annotations available in the full SciDTB () corpus; sequence labelling task with dependency-based word embeddings, contextualized ElMo, RST encodings, GloVe	Enrich a subset of SciDTB with additional layer of argumentation, EDUs as minimal span for annotation, pilot task to predict acceptance/rejection using automatically identified argumentative components and relations
Contractor et al. (2012)	Biomedical papers	Leveraging on AZ features for extractive summarization of scientific articles	Used AZ categories as features in final sentence selection process + additionally used verbs, tf-idf, citation and reference occurrences, locative features for classification to generate initial set of candidate sentences. Then performed k-Means cluatering to group similar sentences and select the centroid from each group to generate the summary (redundancy elimination)	Demonstrated the efficacy of weakly-supervised AZ classifier for less training data by Guo et al. (2011) for scientific article summary extraction
Teufel and Moens (2002)	Computational Linguistics papers	Summarize scientific articles by concentrating on the rhetorical status of statements in an article	Developed an algorithm to select content from articles and classify them into rhetorical categories which integrate argumentation structure in scientific papers	
Feltrim and Teufel (2004)	Brazilian PhD Theses in Computer Science	Integrated Argumentative Zoning into an automatic Critiquing Tool for Scientific Writing in Portuguese (SciPo)	Implemented a set of 7 features, derived from the 16 used by (Teufel and Moens, 2002), Naive Bayes as the classifier	Port the feature detection stage of AZ from English to Portuguese, a human annotation experiment to verify the reproducibility of the annotation scheme, intrinsic evaluation of AZ-part of SciPo
Groza et al. (2011)	Production and Manufacturing, Biomedical, Law/Legal	The authors present SALT (Semantically Annotated LATEX), a semantic authoring framework that enables the externalization of the argumentation and rhetoric captured in scientific publication's content.	The annotation framework is a layered organization of three ontologies: the Document Ontology - capturing the linear structure of the publication, the Rhetorical Ontology - modeling the rhetorical and argumentation, and the Annotation Ontology - linking the rhetoric and argumentation to the publication's structure and content.	A LATEX and MS-Word plugin for semantic annotation of scientific publications as per SALT scheme
de Waard et al. (2009)		Proposal to extract knowledge from articles to allow the construction of a system where a specific scientific claim is connected, through trails of meaningful relationships, to experimental evidence. To improve access to collections of scientific papers represented as networks of collection of claims that have a defined epistemic value, with links to experimental evidence and argumentative relationships to other statements and evidence. The authors coin this conceptual approach 'Hypotheses, Evidence and Relationships' (HypER).		
Yu et al. (2020)	PubMed papers and news articles	Study exaggeration in press releases	Developed a new corpus and trained models that can identify causal claims in the main statements in a press release. By comparing the claims made in a press release with the corresponding claims in the original research paper, the authors found that 22% of press releases made exaggerated causal claims from correlational findings in observational studies.	
Li et al. (2021)	Biomedical papers	demonstrate the benefit of leveraging scientific discourse tags for downstream tasks such as claim-extraction and evidence fragment detection	Develop a sentence-level sequence tagging model to label discourse types for each sentence in a paragraph	