Argumentation Mining in Scientific Discourse

Nancy L. Green
University of North Carolina Greensboro, Greensboro, NC, USA
nlgreen@uncg.edu

Abstract

The dominant approach to argumentation mining has been to treat argumentation scheme detection as a machine learning problem based upon superficial text features, and to treat the relationships between arguments as support or attack. However, applications such as accurately representing and summarizing argumentation in scientific research articles require a deeper understanding of the text and a richer model of relationships between arguments. This paper presents a semantic rule-based approach to extracting individual arguments, and demonstrates the need for a richer model of inter-argument relationships in biomedical/biological research articles.

1   Introduction

The dominant approach to argumentation (or argument) mining [e.g., Green et al., 2014; Cardie et al., 2015; Reed et al., 2016] has been to treat it as a machine learning problem based upon superficial text features, enabling researchers to adopt methods that have been applied successfully to other natural language processing tasks. This approach has been useful for applications such as identifying reasons given for opinions in social media, or automatic assessment of student essay quality. However some applications, such as accurately summarizing argumentation in scientific research articles, require a deeper understanding of the text.

There are a number of problems with mining arguments in scientific documents at the text level rather than at the semantic level [Green, 2015a; 2015b]. Argument components may not occur in contiguity. In fact, the content of an argument may be widely separated or the content of two arguments may be interleaved at the text level. Furthermore scientific text often contains enthymemes, i.e. arguments with implicit premises or an implicit conclusion. Interpretation of enthymemes may require use of the preceding discourse context (including inferred conclusions of other arguments), presumed shared knowledge of the author and

audience, as well as constraints of the underlying argumentation scheme [Green, 2010].

Although human-level understanding of natural language text is currently beyond the state of the art, we contend that an inference-based approach is feasible for applications requiring a deeper analysis of argumentation. In [Green, 2016] we proposed an approach to mining individual arguments in biomedical research articles using argumentation schemes implemented as logic programs. The schemes are formulated in terms of semantic predicates that could be obtained from a text by use of BioNLP (biomedical/biological natural language processing) tools. This semantic approach to mining avoids the various problems faced by purely feature-based approaches, e.g., that argument components may be conveyed through non-contiguous or overlapping text segments of varying granularity, the sparcity of discourse cues marking argument components, and the occurrence of enthymemes.

In this paper, we build on our previous proposal by considering the role of discourse structure in mining argumentation in scientific texts. In section 2 we summarize our previous proposal to mining individual arguments. In section 3 we discuss the relationship of the individual arguments to other aspects of discourse structure, and how the arguments are related to each other, i.e., the argumentation structure of the discourse.

2   Mining Arguments

This section summarizes our proposed approach to mining individual arguments described in [Green, 2016], using argumentation schemes implemented as logic programs written in Prolog [Bratko, 2001]. Argumentation schemes are abstract descriptions of acceptable, possibly defeasible, arguments used in conversation as well as in formal genres such as legal and scientific text [Walton et al., 2008]. To provide examples of argumentation schemes in open-access text, we analyzed arguments in the Results section of a biomedical research article [van de Leemput et al., 2007] in the CRAFT corpus [CRAFT]. The CRAFT corpus has been annotated by other researchers for purposes of biomedical text mining [Verspoor et al., 2012; Bada et al., 2012], but not for argument mining. The seven argumentation schemes presented

in [Green, 2016] were implemented in terms of domain-specific semantic predicates that could in theory be automatically extracted by BioNLP tools. (Results of a preliminary study of human analysts' ability to apply the argumentation schemes consistently will be reported in the future.) We expect that these rules, while domain-specific, are applicable to the large body of research articles on genetic variants with effects on human health.

To illustrate a few of the schemes, first, the *Method of Agreement* scheme can be paraphrased as follows.

Premises:
- A group of individuals G have atypical phenotype P
- All of the individuals in G have atypical genotype M.

Conclusion: M may be the cause of P (in G).

In the rules, *genotype* describes a variation (mutation) at the level of chromosome, region on the chromosome, or gene that may have a deleterious effect (or effects), and *phenotype* describes the effect(s). This scheme can be seen as a specialization of a more general scheme, related to Mill's Method of Agreement [Jenicek and Hitchcock, 2005].

Another scheme, related to Mill's *Method of Difference* [Jenicek and Hitchcock, 2005], can be paraphrased as follows.

Premises:
- A group of individuals G have atypical phenotype P
- All of the individuals in G have atypical genotype M.
- A group of individuals Control do not have P.
- None of the individuals in Control have M.

Conclusion: M may be the cause of P (in G).

The following scheme can be seen as a specialization of *Argument by Analogy*, e.g. as described in [Walton et al., 2008].

Premises:
- Phenotype P1 of group G1 is similar to phenotype P2 of group G2
- Genotype M1 of group G1 may be the cause of P1.
- Genotype M2 of group G2 is similar to genotype M1.

Conclusion: M2 may be the cause of P2 (in G2).

To use the implemented rules for argument mining, i.e., to extract individual arguments, it is assumed that, firstly, BioNLP tools would be applied to a source text to create a knowledge base (KB). Named entity recognition tools such as ABNER [Settles, 2005] or MutationFinder [Caporaso et al., 2007] could be used to recognize expressions referring to semantic class names such as genes, mutations, proteins, and phenotypes. Domain-specific relations in the argumentation schemes such as *have_phenotype* and *have_genotype* could be extracted from the text using relation extraction tools such as OpenMutationMinder [Naderi and Witte, 2012] and DiMeX [Mahmood et al., 2016]. Also, a certain amount of domain knowledge would be required, e.g., for the relations *similar* and *difference*, which could be acquired from a domain ontology or domain experts. After a KB has been created, the argument scheme rules would be applied to the KB to recognize the premises, conclusion, and argumentation scheme of each argument in the text. The rules were tested by manually creating a KB and then applying the rules to the KB.

To illustrate, the implementation of the argumentation scheme for Method of Agreement is as follows.

```
arg(
  scheme('Agreement'),
  premise(have_phenotype(G, P)),
  premise(have_genotype(G, M)),
  conclusion(cause(M, P)))
:-
group(G),
have_phenotype(G, P),
have_genotype(G, M).
```

Applying this rule to a KB containing the facts:

```
group(mice1).
have_phenotype(mice1, ataxia).
have_genotype(mice1, 'Itpr1 opt/opt').
```

would derive an argument whose scheme is identified as Agreement, whose premises are the above listed facts, and whose conclusion is that the Itpr1 opt/opt variant may be the cause of their ataxia. Note that the rules are formulated in such a way that even implicit conclusions of arguments can be recognized, given the premises and an argumentation scheme rule. The inferred conclusion itself can be added to the KB, enabling it to be used as a premise in subsequent arguments. Also note that the conclusions of the schemes are not asserted with complete certainty. The corresponding arguments in the source

text range in force from 'plausible hypothesis' to 'fairly certain conclusion'. For details of the Prolog implementation of the seven schemes, see [Green, 2016].

This approach is in contrast to machine-learning approaches to argumentation scheme recognition that use only superficial text features such as keywords, parts of speech, and clause length [Feng and Hirst, 2011; Lawrence and Reed, 2016]. In addition, some of those approaches, e.g. [Feng and Hirst], assume that clauses of a text are labeled as premise or conclusion before argumentation scheme recognition begins. On the other hand, similar to our approach, [Saint-Dizier, 2012] uses manually-derived rules encoded in a logic programming language for automatic identification of arguments giving reasons for a conclusion in instructional texts or opinion texts. However, the rules are based on syntactic patterns and lexical features.

3 The Results Narrative

Having proposed an approach to mining individual arguments, the next step of our research is to investigate how the arguments are related to other aspects of discourse structure, and how the arguments are related to each other. Previous computation-oriented investigations of discourse in the natural sciences have addressed automatic classification of text segments, e.g., discourse coherence relations in corpora such as BioDRB [Prasad et al., 2011] and BioCause [Mihaila et al., 2013], argumentative zones [Teufel, 2010], and activities in a scientific investigation (CoreSC) [Liakata, 2012]. None of those annotation schemes treat arguments in the sense described in the previous section.

The Results section of the article whose arguments were analyzed in [Green, 2016] reports on a logical and temporal sequence of experiments. Arguments are given in the context of this narrative, i.e. the report of the scientific investigation. Figure 1 shows our ad hoc analysis of the narrative using descriptive terms similar to those of the argumentative zone and CoreSC systems. The relevant content of the article has been paraphrased in the figure.

The Results section begins, in its first paragraph, with a description of the fortuitous discovery of an inherited disorder in mice bred in the authors' lab. Then the authors describe a sequence of three experiments intended to reveal the genetic variant responsible for that mouse disorder. Figure 1 describes each individual experiment in terms of its Goal, (use of) Previous Research or Background Knowledge, Hypothesis, Method, Result, and/or Conclusion. (To avoid confusion with argument components, the first letter of terms describing parts of an experiment, e.g. Conclusion, will be capitalized, and labels of argument components will be italicized.)

In the report of Experiment 1, the *conclusion* of Argument 1 must be inferred to understand the Conclusion of the experiment. Also, it can be seen that the *conclusion* of Argument 1 is a *premise* needed in the argument (Argument 2) for the Hypothesis of Experiment 2 in the next paragraph. Note that Experiment 2 contains two arguments, one (Argument 2) for the Hypothesis, and one (Argument 3) for the Conclusion. The *conclusions* of Argument 2 and Argument 3 are not identical, i.e., the *conclusion* of Argument 3 is more specific than that of Argument 2. The *conclusion* of Argument 3 can be challenged by a critical question of the Method of Agreement scheme, i.e., whether the putative cause of the disorder is causally plausible. Experiment 3 contains Argument 4, whose *conclusion* addresses that critical question. Note that, in general, critical questions associated with argumentation schemes provide ways in which arguments may be challenged [Walton et al., 2008].

Paragraph 4 states a new Goal: to discover any related genetic variants causing a similar disorder in humans. The article then goes on to describe another sequence of experiments towards that goal.

In Experiment 4, Background Knowledge and the *conclusion* of Argument 4 (in Experiment 3) are used (in Argument 5) to argue for the Hypothesis of the experiment. Based on the Result of Experiment 4, the implicit Conclusion (*conclusion* of Argument 6) that a deletion in ITPR1-SUMF1 may be the cause, is broader than the original Hypothesis, that a deletion in ITPR1 may be the cause. However, the *conclusion* is also narrower in the sense that it is restricted to particular individuals in the AUS1 family.

In Experiments 5 and 6, the respective *conclusions* of Argument 7 and Argument 8 agree with the *conclusion* of Argument 6. Experiment 7 provides support (Argument 9) for a related *conclusion*. The *conclusions* of arguments in Experiments 5-7 are used as *premises* to argue for a more general *conclusion* in Argument 10. Then, using results of Previous Research, the authors argue (Argument 11) against part of the *conclusion* of Argument 10. The *conclusion* of Argument 11 is a *premise* of Argument 12, whose *conclusion* is a refinement of the *conclusion* of Argument 10. Finally, the *conclusion* of Argument 13 addresses a critical question of Argument 12, a relationship like that of Argument 4 to Argument 3.

As for the first of the above research questions (how arguments are related to other aspects of discourse structure), the above analysis raises some interesting possibilities. It could be that some

argumentation schemes are more commonly used for certain scientific purposes and not others. For example, Argument by Analogy is used only in arguments for Hypotheses in the article that was analyzed. This could be verified by statistical analyses of corpora, and if true, could augment the semantic method we have proposed for extracting individual arguments. Another possibility is that the location in the narrative could be used as a constraint on argument scheme recognition. Although it can be seen in Figure 1 that argument content is not in one-to-one correspondence with paragraph or Experiment boundaries, in cases where multiple rules match, a heuristic strategy of preferring local content might be applied.

As for the second of the above research questions (how the arguments are related to each other), the analysis in Figure 1 shows that a richer model is needed than support-attack relationships represented in current argument mining approaches, e.g., [Cabrio and Villata, 2012], [Peldszus and Stede 2016], [Stab and Gurevych, 2014]. The relationships between conclusions of the arguments in Figure 1 is summarized in Table 1.

Table 1. Dialectical structure of Results section.

Conclusion of argument #	Relationship to	Conclusion of argument #
2	Refines	1
3	Refines	2
4	Responds to CQ of	3
5	Analogous to	3
6	Broadens from del. of ITPR1 to ITPR1-SUMF1 & restricts from humans to AUS1	5
7	Agrees with	6
8	Agrees with	6
9	Extends to more individuals	6
10	Extends	9 and 6
11	Attacks part of and extends to humans	6
12	Refines	11
13	Responds to CQ of	12

It would be misleading to reduce these relationships to pro or con some claim. As described in research on formal dialogue games for use by software agents, the discovery dialogue [McBurney and Parsons, 2001] has a key feature in common with the scientific article that we analyzed. In a discovery dialogue, the goal is not to try to prove or disprove a given claim, but to discover something not previously known.

4 Discussion

Previous argumentation mining research has not addressed the natural sciences. However, argumentation is an important feature of scientific discourse. This paper proposes a semantic approach to automatic recognition of premises, conclusion, and argumentation scheme of arguments in scientific text. Argumentation schemes are implemented as logic programs. The logic programs would be used with a knowledge base that could be constructed from a text in a large part automatically using existing language processing tools (as described in section 2). The logic programs can be used, not only to recognize fully explicit arguments in the text, but also arguments with implicit conclusions. This is important because often the conclusions are implicit and may function as implicit premises of subsequent arguments in the text. Although the argumentation schemes have been implemented using domain-specific predicates, they are specializations of more general schemes applicable to other qualitative causal domains in the natural sciences.

Also, as a step towards automatic recognition of the structure of argumentation in scientific discourse, we present a discourse analysis of part of a scientific text and discuss the relationship of the individual arguments to other aspects of the discourse structure, and how the arguments are related to each other, i.e., the argumentation structure of the discourse. It is shown that a richer model is needed than support-attack relationships.

Acknowledgments

References

[Bada et al., 2012] M. Bada, M. Eckert, D. Evans, et al. Concept Annotation in the CRAFT corpus. *BMC Bioinformatics*, 13:161, 2012.

[Bratko, 2001] Ivan Bratko. *Prolog Programming for Artificial Intelligence*. 3rd edition. Addison-Wesley, Harlow, England, 2001.

[Cabrio and Villata, 2012] Cabrio, E. and Villata, S. 2012. Generating Abstract Arguments: A Natural Language Approach. In Verheij, B., Szeider, S., and Woltran, S. (eds.) *Computational Models of Argument: Proceedings of COMMA 2012*. Amsterdam, IOS Press, 454-61.

[Caporaso et al. 2007] J. G. Caporaso, W.A. Baumgartner Jr., D. A. Randolph, K. B. Cohen, and L. Hunter. MutationFinder: A high-performance system for extracting point mutation mentions from text. *Bioinformatics*, 23:1862-1865.

[Cardie et al., 2015] C. Cardie et al. (Eds.) *Second Workshop on Argumentation Mining*. North American Conference of the Association for Computational Linguistics, Denver, 2015.

[CRAFT] http://bionlp-corpora.sourceforge.net/ CRAFT/ index.shtml.

[Feng and Hirst, 2011] V.W. Feng and G. Hirst. Classifying Arguments by Scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics,* pages 987-996, Portland, OR, 2011.

[Green, 2010] N. Green. Representation of Argumentation in Text with Rhetorical Structure Theory. *Argumentation* 24(2): 181-196.

[Green, 2015a] N. Green. Identifying Argumentation Schemes in Genetics Research Articles. In *Proceedings of the Second Workshop on Argumentation Mining*, North American Conference of the Association for Computational Linguistics (NAACL), Denver, CO, 2015.

[Green, 2015b] N. Green. Annotating Evidence-Based Argumentation in Biomedical Text. In *Proc. 2015 Int. Workshop on Biomedical and Health Informatics, IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM 2015), Washington, D.C, Nov. 9-12, 2015*. IEEE Computer Society Press.

[Green, 2016] N. Green. Implementing Argumentation Schemes as Logic Programs. *Workshop on Computational Models of Natural Argument (CMNA 2016)*. International Joint Conference on Artificial Intelligence, New York.

[Green et al., 2014] N. Green, et al. (Eds.) *First Workshop on Argumentation Mining*. Association for Computational Linguistics, Baltimore, MD., 2014.

[Jenicek and Hitchcock, 2005] M. Jenicek and D. Hitchcock. *Logic and Critical Thinking in Medicine*. American Medical Association Press, 2005.

[Lawrence and Reed, 2016] Lawrence, J. and Reed, C. Argument Mining Using Argumentation Scheme Structures. In Baroni, P. et al. (eds.) *Computational Models of Argument: Proceedings of COMMA 2016*. Amsterdam, IOS Press, 379-90.

[van de Leemput et al., 2007] J. van de Leemput, J. Chandran, M. Knight, et al. Deletion at *ITPR1* Underlies Ataxia in Mice and Spinocerebellar Ataxia 15 in Humans. *PLoS Genetics*, 3(6) e108:1076-1082, 2007.

[Liakata, M. et al., 2012] M. Liakata. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics* 28(7), 2012.

[Mahmood et. al., 2016] Mahmood A.S., Wu, T.J., Mazumder, R. and Vijay-Shanker, K. (2016) DiMeX: A Text Mining System for Mutation-Disease Association Extraction. *PLoS One*.

[McBurney and Parsons, 2001] P. McBurney and S. Parsons. Chance Discovery Using Dialectical Argumentation. In *New Frontiers in Artificial Intelligence*, T. Terano et al. (Eds.). Lecture Notes in Artificial Intelligence, v.2253, Berlin, Springer Verlag, 414-424.

[Naderi and Witte, 2012] N. Naderi and R. Witte. Automated extraction and semantic analysis of mutation impacts from the biomedical literature. *BMC Genomics*, 13(Suppl 4):510, 2012.

[Peldszus and Stede, 2016] A. Peldszus, M. Stede. Rhetorical structure and argumentation structure in monologue text. In: Proceedings of the 3rd Workshop on Argumentation Mining, ACL Berlin, 2016.

[Prasad et al., 2011] R.Prasad, S. McRoy, N. Frid, A. Joshi, and H. Yu. The Biomedical Discourse Relation Bank. *BMC Bioinformatics*, 12:188, 2011.

[Reed et al., 2016] C.Reed et al. (Eds.) *Third Workshop on Argumentation Mining*. Association for Computational Linguistics, Berlin, 2016.

[Saint-Dizier, P, 2012] P. Saint-Dizier. Process natural language arguments with the <TextCoop> platform. *Argument and Computation* 3(1), March 2012, 49-82.

[Settles, 2005] B. Settles. ABNER: an open source tool for automatically tagging genes, proteins, and other entity names in text. *Bioinformatics* 21(14):3191-3192, 2005.

[Stab and Gurevych, 2014] C. Stab and I. Gurevych, Annotating Argument Components and Relations in Persuasive Essays. In *Proc. COLING 2014*, pp. 1501-1510.

[Teufel, 2010] S. Teufel, S. *The Structure of Scientific Articles: Applications to Citation Indexing and Summarization.* Stanford, CA, CSLI Publications.

[Verspoor et al., 2012] K. Verspoor, K.B. Cohen, A. Lanfranchi, et al. A Corpus of Full-text Journal Articles is a Robust Evaluation Tool for Revealing Differences in Performance of Biomedical Natural Language Processing Tools. *BMC Bioinformatics* 13:207, 2012.

[Walton et al., 2008] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes.* Cambridge University Press, 2008.

Appendix. Figure 1. Analysis of discourse structure and argumentation.

¶ 1

Report Observation: Some mice bred in the authors' lab are affected with an autosomal recessive disorder resembling a kind of ataxia. *(premise, Argument 1)*

¶ 2

Report Experiment 1:

    Method: Linkage analysis

    Result: The affected mice have a lesion on chromosome 6qE1 *(premise, Argument 1)*

    Conclusion (implicit): A genetic variant on 6qE1 may be the cause of their disorder. *(conclusion, Argument 1 - Method of Agreement; premise, Argument 2)*

¶ 3

Report Experiment 2:

    Previous Research: A certain deletion on 6qE1 in gene Itpr1, the Itpr1 opt/opt variant, is known to cause a similar disorder in mice. *(premise, Argument 2)*

    Hypothesis: A deletion in Itpr1 may be the cause the disorder in the lab's mice. *(conclusion, Argument 2 – Analogy)*

    Method: Sequence Itpr1

    Result: The lab's affected mice have a deletion in Itpr1: the Itpr1 Δ18/Δ18 variant. *(premise, Argument 3)*

    Conclusion: The Itpr1 Δ18/Δ18 variant may be the cause of their disorder. *(conclusion, Argument 3 – Method of Agreement)*

Report Experiment 3:

    Previous Research: Cerebellar Purkinjee cells of Itpr1 opt/opt mice, who have ataxia, have decreased Itpr1 expression *(premise, Argument 4)*

    Method: Measure expression of Itpr1 in cerebellar Purkinjee cells of mice with the Itpr1 Δ18/Δ18 variant.

    Result: Decreased level of Itrpr1 expression found. *(premise, Argument 4)*

    Conclusion (implicit): There is a plausible explanation at the molecular level of how deletions in Itpr1 may cause ataxia-like disorders in mice. *(conclusion, Argument 4 – Consistent Explanation; premise, Argument 5)*

¶ 4

Report Goal: Discover cause of cognate human disorders, such as spinocerebellar ataxia 15 (SCA15), where no causal mutation has been identified.

¶ 5

Report Experiment 4:

Background knowledge (implicit): The ITPR1 gene in humans is functionally similar to Itpr1 in mice. *(premise, Argument 5)*

Hypothesis: A deletion in ITPR1 is a cause of SCA15 in humans. *(conclusion, Argument 5 – Analogy)*

Method: Sequence DNA from three AUS1 family members with SCA15.

Result: The three family members had a deletion in ITPR1-SUMF1. *(premise, Argument 6; premise, Argument 7)*

Conclusion (implicit): A deletion in ITPR1-SUMF1 may be the cause of ataxia in the three AUS1 family members. *(conclusion, Argument 6 – Method of Agreement)*

Report Experiment 5:

Goal: Determine if the ITPR1-SUMF1 deletion is a benign polymorphism.

Method: Compare to ITPR1 and SUMF1 in two control groups.

Results: No deletion found in ITPR1 or SUMF1 in the control groups. *(premise, Argument 7)*

Conclusion: A deletion in ITPR1-SUMF1 may be the cause of ataxia in the three affected AUS1 family members. *(conclusion, Argument 7 – Method of Difference)*

¶ 6

Report Experiment 6:

Method: Fine-map the breakpoints of the deletion in the affected AUS1 family members and in the controls.

Result: Deletion of the first three of the nine exons of SUMF1 and the first 10 of the 58 exons of ITPR1 in the affected family members only. *(premise, Argument 8)*

Conclusion: A deletion in ITPR1-SUMF1 may be the cause of ataxia in the three AUS1 family members. *(conclusion, Argument 8 – Method of Difference)*

¶ 7

Report Experiment 7:

Method: Analyzed two additional families (H33 and H27) with an inherited cerebellar ataxia similar to that described in the AUS1 family.

Result: The affected H33 and H27 family members have a deletion at the SCA15 locus from SUMF1 through IPTR1; the unaffected family members do not. *(premise, Argument 9)*

Conclusion (implicit): A deletion in ITPR1-SUMF1 may be the cause of the ataxia disorder in the affected H33 and H27 family members. *(conclusion, Argument 9 – Method of Difference)*

Report Conclusion:

In three families [AUS1, H33, H27] cerebellar ataxia segregated with a deletion in SUMF1-ITPR1, not observed in controls. (*premise, Argument 10*)

The deletion in ITPR1-SUMF1 is the cause of SCA15 in those families. (*conclusion, Argument 10 – Method of Difference; premise, Argument 12*)

¶ 8

Report Previous Research:

Homozygous mutation of SUMF1 results in autosomal recessive multiple sulfatase deficiency…

No co-occurrence of ataxia has been described in heterozygous parents [i.e. who have one copy of a SUMF1 mutation] of patients with multiple sulfatase deficiency (*premise Argument 11*).

It is improbable that the deletion of SUMF1 … itself causes or contributes to

SCA15 (*conclusion Argument 11 – Failed Method of Agreement; premise Argument 12*).

[Therefore, deletion in ITPR1 is the likely cause of ataxia in the three families (*implicit conclusion, Argument 12 – Eliminate Difference*)]

Mutation of ITPR1 is biologically plausible as a cause of ataxia (*conclusion, Argument 13 – Consistent Explanation*