

Manual Identification of Arguments with Implicit Conclusions Using Semantic Rules for Argument Mining

Nancy L. Green

University of North Carolina Greensboro
Greensboro, N.C. 27402, U.S.A.
nlgreen@uncg.edu

Abstract

This paper describes a pilot study to evaluate human analysts' ability to identify the argumentation scheme and premises of an argument having an implicit conclusion. In preparation for the study, argumentation scheme definitions were crafted for genetics research articles. The schemes were defined in semantic terms, following our proposal to use semantic rules to mine arguments in that literature.

1 Introduction

Surface text-level approaches to argument mining in the natural sciences literature face various problems (Green, 2015b). The premises and conclusion of an argument are not necessarily expressed in adjacent phrasal units. Components of different arguments may be interleaved in the text. Even more challenging, some of the premises or the conclusion of an argument may be implicit. For example, the following excerpt can be interpreted as expressing an argument having the implicit, tentative conclusion that a certain mutation within the *Itp1* gene may be the cause of the affected mice's movement disorder: *Our initial observations suggested that the affected mice suffered from an apparently paroxysmal movement disorder ... Sequencing ... revealed a single mutation within Itp1 ...* (Van de Leemput et al., 2007).

Argumentation schemes (Walton et al., 2008) describe acceptable, often defeasible, patterns of reasoning. The schemes place additional

constraints on the relation between an argument's premises and its conclusion than do discourse coherence models such as Rhetorical Structure Theory (Mann and Thompson, 1988) and similar models used to annotate scientific corpora (e.g. Prasad et al., 2011). Recognizing the argumentation scheme underlying an argument is an important task. First, each argument scheme has an associated set of *critical questions*, or potential challenges, so recognizing the scheme can provide information for generating or recognizing challenges. Second, and most relevant to the concerns of this paper, the constraints of the scheme can provide information for inferring an implicit argument component, such as the conclusion of the argument in the above excerpt.

The above problems suggest that a semantics-informed approach to argument mining in this genre would be desirable. We have proposed an approach to argument mining within genetics research articles using semantic argumentation scheme definitions implemented in a logic programming language (Green, 2016). A significant advantage of that approach is that implicit conclusions of arguments can be recognized.

To evaluate such an approach, it would be useful to have a corpus of genetics research articles whose arguments (i.e., argumentation scheme, implicit and explicit premises, and implicit or explicit conclusion) have been identified by human analysts. Note that there is no such corpus currently available, and creating such a corpus will be expensive. To contribute to the creation of such a corpus, we created a draft manual of argumentation scheme definitions in the genetics domain for use by

human analysts, and ran a pilot study to evaluate human analysts' ability to apply the definitions to text containing arguments with *implicit* conclusions. As far as we know, no such study has been performed with text from the natural sciences research literature. The main contribution of this paper is to present the study. However, to motivate our interest in using semantic definitions of argumentation schemes in argument mining, we present additional background in the next section. Section 3 describes the study, and Section 4 outlines plans for future work.

2 Background

This section explains how semantic rules could be used in argument mining as proposed in (Green, 2016) and compares it to current approaches. The first step in our proposal is to preprocess a text to extract semantic entities and relations specific to the domain. For example, to mine articles on genetic variants that may cause human disease, we proposed extracting a small set of semantic predicates, such as those describing an organism's genotype and phenotype. Although automatically identifying entities and relations in biomedical text is very challenging, it is the object of much current research (Cohen and Demner-Fushman, 2014), and we assume that BioNLP tools will be able to automate this in the near future. Current BioNLP relation extraction tools include OpenMutation-Minder (Naderi and Witte, 2012) and DiMeX (Mahmood et al., 2016). The extracted relations would populate a Prolog knowledge base (KB).

Next, Prolog rules implementing argument schemes, such as the following, would be applied to the KB to produce instances of arguments, i.e., including the semantic relations comprising the premises and the conclusion, as well as the name of the underlying argumentation scheme.

```
argument(
  scheme('Method of 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).
```

As a proof-of-concept, the rules were implemented and tested on a manually-created KB. An advantage of this approach is that implicit conclusions of arguments are recognized automatically and can be added to the KB. The added conclusions can then serve as implicit premises of subsequent arguments given in the text.

A semantics-informed approach is in contrast to today's machine learning approaches that use only surface-level text features. Among those approaches, there has been little concern with argument scheme recognition, except for (Feng and Hirst, 2011; Lawrence and Reed, 2016). Saint-Dizier (2012) uses manually-derived rules for argument mining, but the rules are based on syntactic patterns and lexical features. None of these approaches is capable of identifying implicit argument components.

A possible limitation of a semantic rule-based approach is the necessity to first extract semantic relations. However in BioNLP domains, where relation extraction tools are being developed for other purposes and the size of the targeted literature is huge and constantly growing, the benefits may outweigh the cost. Another possible limitation is "scalability", or the cost of manually deriving rules for topics not covered by the current rules. However, the rules are specializations of argumentation schemes that have been previously cited as applicable to the natural sciences in general (Walton et al., 2008; Jenicek and Hitchcock, 2005), so it is plausible that the effort to create rules for other topics in the natural sciences will not be significantly higher than the cost of formulating the current rule set.

3 Pilot Study

3.1 Preparation

The author created a draft document defining argumentation schemes in terms of the domain concepts used in the Prolog implementation of Green's argumentation schemes (2016). Note that in an earlier study (Green, 2015a) we provided definitions of argumentation schemes found in the genetics research literature, but the definitions were abstract and did not refer to domain-

specific concepts used in (Green, 2016). It was decided to redefine the schemes in terms of domain concepts to more closely align with the implementation in (Green, 2016).

A team consisting of the author, a biology doctoral student with research experience in genetics, and a computer science graduate student with an undergraduate degree in philosophy collaborated on identifying the arguments in the Results section of two articles (Van de Leemput et al., 2007; Lloyd et al., 2005) from the CRAFT corpus (Bada et al., 2012; Verspoor et al. 2012). The CRAFT corpus is open access and has already been linguistically annotated. During this process, the argumentation scheme definitions were refined. The goal of the pilot study reported here was to determine if other researchers could apply the resulting definitions with some consistency, and to test the feasibility of this type of study before conducting a larger study.

3.2 Procedure and Materials

Human identification of an argument's premises, conclusion, and scheme in this genre is a very challenging task, requiring some domain knowledge as well as training on argumentation schemes. Thus, it was decided to focus on certain aspects of the problem in this study. The study materials consisted of the draft document of argumentation scheme definitions, and a set of five problems. The problems were constructed to test identification of five different argumentation schemes that we have frequently seen used in this genre and whose definitions are similar to the Prolog rules given in (Green, 2016). The schemes are paraphrased below in a more compact form than that presented to participants. Definitions of domain-specific predicates such as *genotype* and *phenotype* also were included in materials given to participants.

Method of Agreement: If a group has an abnormal genotype G and abnormal phenotype P then G may be the cause of P.

Method of Difference: If a group has an abnormal genotype G and abnormal phenotype P and a second group has neither, then G may be the cause of P.

Analogy: If a group has abnormal genotype G and abnormal phenotype P and G may be the cause of P, and a second group has abnormal genotype G' similar to G and abnormal phenotype P' similar to P, then G' may be the cause of P'.

Consistent with Predicted Effect: If a group has abnormal genotype G and abnormal phenotype P and there is a causal mechanism that predicts that G could cause P, then G may be the cause of P.

Consistent Explanation: If a group has an abnormal genotype G, abnormal gene product Prot, and abnormal phenotype P, and G produces Prot, and Prot may cause P (and thus G may cause P), then if a second group has an abnormal genotype G' similar to G, abnormal gene product Prot' similar to Prot, and abnormal phenotype P' similar to P, then G' may be the cause of P'.

Each problem included a short excerpt containing an argument with an *implicit* conclusion (such as the example given in the Introduction of this paper), three to five sentences of background information on genetics that the intended audience, having domain expertise, would be expected to know, and a paraphrase of the conclusion of the argument. Participants were asked to select (1) the name of the applicable argument scheme from a list of nine scheme names (defined in the other document), and (2) the relevant premises from a list of five possible premises. One reason for designing the problems in this way was that we can envision an application of argument mining as finding an argument for a given conclusion, whether or not it has been stated explicitly. A sample problem is shown in Appendix A.

Invitations to participate in the study were emailed to researchers in biology and (mainly) in computer science, and responses were returned by email. No incentives were offered. Given the difficulty of the task due to the unfamiliarity of the domain to most participants, the lack of training other than receiving the draft document of argumentation scheme definitions, and the

length of time required to participate, we were pleased to receive six responses.

3.3 Results and Discussion

The rows of Table 1 show the number of participants who selected each argumentation scheme (Anlg: Analogy, Agr: Agreement, Diff: Difference, CPr: Consistent with Predicted Effect, CEx: Consistent Explanation), and the diagonal shows the number who selected the correct answer. For example, Analogy was correctly selected by four of the six participants; the other two confused it with Consistent with Predicted Effect and Consistent Explanation. Note that one participant selected both Difference and Consistent Explanation as the answer to the problem whose answer was Difference, thus we scored that as 0.5 for each. However, that participant commented that by selecting Consistent Explanation he actually meant Difference Consistent Explanation, an argumentation scheme defined in the other document but not listed among the choices.

	Anlg	Agr	Diff	CPr	CEx
Anlg	4			1	1
Agr		5			1
Diff			5.5		0.5
CPr				5	1
CEx				1	5

Table 1: Confusion Matrix

The table shows that the two schemes that were incorrectly applied the most times (CPr and CEx) were those with the most complicated definitions involving explicit causal explanations. Nevertheless, the results suggest that with careful revision of the definitions and more training than was provided to the participants, humans will be able to identify these schemes consistently.

Table 2 shows the data for each of the six participants. The first row shows the number of premises marked correctly out of 25 choices in all (five choices for each of the five problems). The average number of correctly marked premises was 21/25 or 84 percent. The second row shows that the number of correctly identified schemes

was on average 4/5 or 80 percent. Participants 3, 4, and 5 had the lowest accuracy.

Partic.	1	2	3	4	5	6
Premises	24	23	16	20	20	25
Schemes	5	5	4	2.5	3	5

Table 2: Participant Data

4 Future Work

The long-term goal of the pilot study was to enable us to document arguments in a corpus of scientific research articles. An earlier proposal of ours (Green, 2015b) was to annotate text spans as argument components. In contrast, our current plan is to semantically annotate the arguments in a two-step process. The first step will be to identify the entities and relations in the text. This could be done manually, or better, using BioNLP tools. For example, the result of this step might be an annotated segment like this: *Sequencing ...* `<entity id="e1"> affected mice from the current study </entity> revealed a single mutation ... <entity id="e9"> Itp1Δ18/Δ18 </entity> <relation id="r1" predicate="genotype" entity1="e1" entity2="e9" />`

The second step would be to manually document the arguments in terms of the entities and relations annotated in the first step, e.g., `<argument scheme="Agreement" premise="genotype(e1,e9)" premise="phenotype(e1, e2)" conclusion="cause(r1, r4)" />`

The documented arguments then could be compared to the arguments mined by the semantic approach proposed in (Green, 2016).

Acknowledgments

The analysis of argument schemes was done with the help of Michael Branon and Bishwa Giri, who were supported by a UNCG 2016 Summer Faculty Excellence Research Grant.

References

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

- K. Cohen and D. Demner-Fushman. 2014. *Biomedical Natural Language Processing*. John Benjamins Publishing Company, Amsterdam.
- V.W. Feng and G. Hirst, 2011. Classifying Arguments by Scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, OR, 987-996.
- N.L. Green. 2015a. Identifying Argumentation Schemes in Genetics Research Articles. In *Proc. Second Workshop on Argumentation Mining*. North American Conference of the Association for Computational Linguistics (NAACL).
- N.L. Green. 2015b. Annotating Evidence-Based Argumentation in Biomedical Text. In *Proc. 2015 IEEE International Conference on Bioinformatics and Biomedicine, 2015 International Workshop on Biomedical and Health Informatics*.
- N.L. Green. 2016. Implementing Argumentation Schemes as Logic Programs. In *Proc. of Computational Models of Natural Argument (CMNA-16)*, CEUR Workshop Vol-1876.
- D. Lloyd, F.W. Halt, L.M. Tarantino, and N. Gekakis. 2005. Diabetes Insipidus in Mice with a Mutation in Aquaporin-2. *PLoS Genetics*, August 2005, Vol. 1, Issue 2, e20, 0171-0178.
- M. Jenicek and D.Hitchcock. 2005. *Logic and Critical Thinking in Medicine*. American Medical Association Press.
- J. Lawrence and C. Reed. 2016. 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.
- A.S. Mahmood, T.J. Wu, R. Mazumder and K. Vijay-Shanker. 2016. DiMeX: A Text Mining System for Mutation-Disease Association Extraction. *PLoS One*.
- W. Mann and S. Thompson. 1988. Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. *Text* 8(3): 243-281.
- N. Naderi and R. Witte. 2012. Automated extraction and semantic analysis of mutation impacts from the biomedical literature. *BMC Genomics*, 13(Suppl 4):510.
- R. Prasad, S. McRoy, N. Frid, A. Joshi, and H. Yu. 2011. The Biomedical Discourse Relation Bank. *BMC Bioinformatics* 2011, 12:188.
- P. Saint-Dizier. 2012. Processing natural language arguments with the <TextCoop> platform. *Argument and Computation* 3(1), March 2012, 49-82.
- 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*, 2007, Volume 3, Issue 6, e108, pp. 113-129.
- K. Verspoor, K.B. Cohen, A. Lanfranchi, et al. 2012. 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* 2012, 13:207.
- D. Walton, C. Reed, and F. Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.

Appendix A. Sample Problem Used in Pilot Study.

2. Excerpt:

"Our initial observations suggested the affected mice suffered from an apparently paroxysmal movement disorder ... At initial examination, a human movement disorder specialist ... likened the disorder to episodic intermittent ataxia ... Sequencing of all exons and intron-exon boundaries of *Itpr1* [gene] in affected mice from the current study revealed a single mutation within *Itpr1*: a novel in-frame deletion of 18 bp within exon 36 (*Itpr1* Δ 18/ Δ 18)."

Extra background information:

- The phrase "*Itpr1* Δ 18/ Δ 18" refers to the *Itpr1* gene mutation found in the affected mice.
- Exons include genetic sequences that code for proteins; introns do not.
- A deletion is a type of mutation in which part of a DNA sequence is lost.

Using information from the excerpt and from the extra background information, what type of argument could you give for the following conclusion? Circle the best answer:

The *Itpr1* Δ 18/ Δ 18 mutation may be the cause of the affected mice's movement disorder.

- Agreement
- Failed Method of Agreement
- Analogy
- Consistent Explanation
- Consistent with Predicted Effect
- Difference
- Failed Method of Difference
- Effect to Cause
- Eliminate Candidates

Circle all and only the argument's premises:

- The affected mice suffered from a movement disorder.
- The movement disorder of the mice was likened to episodic intermittent ataxia in humans.
- All exons and intron-exon boundaries of *Itpr1* were sequenced.
- The affected mice were found to have a single mutation within *Itpr1* (*Itpr1* Δ 18/ Δ 18).
- A deletion is a type of mutation in which part of a DNA sequence is lost.