Proposed Method for Annotation of Scientific Arguments in Terms of Semantic Relations and Argument Schemes

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

Abstract

This paper presents a proposed method for annotation of scientific arguments in biological/biomedical journal articles. Semantic entities and relations are used to represent the propositional content of arguments in instances of argument schemes. We describe an experiment in which we encoded the arguments in a journal article to identify issues in this approach. Our catalogue of argument schemes and a copy of the annotated article are now publically available.

1 Introduction

This paper presents our current work on semantic annotation of scientific arguments in full-text biological/biomedical journal articles. The goal is to 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 (Lippi and Torroni, 2016). Computational research on scientific discourse has focused on classification of text segments in terms of rhetorical goals (Teufel, 2010), experimental science activities (Liakata et al., 2012), or coherence relations (Prasad et al., 2011). Although some categories of those classification schemes are related to argumentation, those approaches are inadequate for representation of argumentation. Focusing mainly on non-technical literature and social media, argument mining researchers have investigated automatic classification of text segments in terms of argumentation concepts, e.g., premise/ conclusion, support/attack (Peldszus and Stede, 2016; Stab and Gurevych, 2016).

However, the propositional content of scientific arguments does not map neatly to text segments: two distinct arguments may be expressed in overlapping or embedded text; argument premises and conclusions may occur in non-contiguous text segments of varying granularity, and sometimes they may be implicit. As an alternative, we proposed a semantics-informed approach to argument mining in the biological/biomedical sciences as follows (Green, 2018). First, BioNLP tools could be used to derive a partial semantic interpretation of a text; next, argument schemes implemented as logic programs could be used to identify the propositional content of arguments, including implicit conclusions.

Consistent with that approach, we created a publically available catalogue of 15 argument schemes that we identified in journal articles on health effects of genetic variants. The schemes are expressed in terms of domain concepts used in our logic programs, rather than by generic definitions as in, e.g., (Walton et al., 2008). Here we describe an experiment in which we manually encoded the arguments in the "Results/Discussion" section of an article from that domain. The goals were to evaluate the feasibility of the task and to identify issues in the semantic representation of the arguments, as a step towards building a publically available corpus of argument-annotated full-text scientific journal articles. There are currently

no corpora of argument-annotated articles from the natural sciences research literature. The information gained will be used to refine our approach as we build a corpus. The catalogue and annotated article are available at https://github.com/ greennl/BIO-Arg.

2 Method

An article (van de Leemput et al., 2007) was selected from the open-access CRAFT corpus[1] in case the syntactic and concept annotations of that corpus (Bada et al. 2012; Verspoor et al. 2012) might be useful in the future in combination with our argument annotations. We annotated the arguments in the first eight paragraphs of the ten-paragraph "Results/ Discussion" section. (Although not participating in the annotation, a domain expert had previously helped interpret the article.) Encoded in XML, annotations were added using a text editor. (We adopted a lightweight approach to annotation tools due to the experimental nature of this work.) Part of the DTD for argument-related tags is shown in Figure 1.

```
<!ELEMENT  DSEG (content* | entities-props* |
        argument* )*  >
<!ATTLIST   DSEG  ID CDATA #REQUIRED >
<!ELEMENT entities-props (entity* | prop*)* >
<!ATTLIST   entity ID CDATA #REQUIRED >
<!ATTLIST   entity paraphrase CDATA >
<!ELEMENT argument (premise-list, conclusion ) >
<!ATTLIST   argument ID CDATA #REQUIRED >
<!ATTLIST   argument scheme CDATA #REQUIRED >
<!ELEMENT premise-list (premise+ ) >
<!ATTLIST   premise prop CDATA >
<!ATTLIST   premise domain-prop CDATA >
<!ATTLIST   premise paraphrase CDATA >
<!ATTLIST   premise conclusion-of CDATA >
<!ATTLIST   conclusion prop CDATA >
<!ATTLIST   conclusion inferred-prop CDATA >
<!ATTLIST   conclusion paraphrase CDATA >
```

Figure 1: Part of DTD

For illustration, Figure 2 shows an excerpt of the annotated document. (Text of some <content> elements has been elided to save space.)

```
<DSEG ID="Observation" >
<content> During the generation of a line of mice
with knockout of the gene Park7 we noted an early
movement disorder that was inherited
independently of targeting vector transmission. Our
initial observations suggested the affected mice
suffered from an apparently paroxysmal movement
disorder, often induced by touch … At initial
examination, … likened the disorder to episodic
intermittent ataxia … </content>
<entities-props>
 <entity ID="group1"
    paraphrase="the affected mice" />
 <entity ID="pheno1"
    paraphrase="ataxia-like movement disorder" />
<prop>have_pheno(group1, pheno1)</prop>
</entities-props>
</DSEG>
<DSEG ID="Experiment 1">
<content> To map the location of the disease-
causing lesion, we performed genome-wide linkage
analysis … Analysis of these data showed a single
genomic region with significant linkage to disease,
providing a two-point LOD score of 5.13 at marker
20.MMHAP85FLG2 Chromosome 6qE1 …</content>
<entities-props>
<entity ID="geno1"
    paraphrase="homozygous mutation on
    chromosome 6qE1" />
<prop>have_geno(group1,geno1)</prop>
</entities-props>
<argument ID="1" scheme="Agreement">
<premise-list>
  <premise prop="have_pheno(group1, pheno1)" />
  <premise prop="have_geno(group1,geno1)"/>
</premise-list>
<conclusion inferred-prop=
    "cause(geno1, pheno1, group1)"
    paraphrase="A homozygous mutation on
    chromosome 6qE1 may be the  cause of the
    ataxia-like disorder in the affected mice" />
</argument>
</DSEG>
```

Figure 2: Annotated excerpt

The article presents a narrative of scientific discovery: a fortuitous observation followed by a series of experiments, intermediate conclusions, more experiments and final conclusions. To preserve this contextual information we divided the Results section into narrative <DSEG> (discourse segment) elements correspondingly. Fig. 2 illustrates the first two <DSEG>s. Each <DSEG> may contain several types of elements: <content>, <entities-props>, and <argument>. In our proposed annotation approach, the text of an article is enclosed within <content> elements of variable length -- from one to nine sentences in this annotated article.

Immediately following a <content> element, a partial semantic interpretation of that content may be given in an <entities-props> element. This element may contain <entity> tags for entities that have been introduced into the discourse in the preceding <content> element. The ID attribute of an <entity> uniquely identifies it in the discourse. Since an entity may have been introduced earlier, the annotator must determine if mentions are coreferential. In Figure 2 the first <entities-props> element shown describes the preceding <content> as introducing two discourse entities, assigned the IDs *group1* and *pheno1* by the annotator. The paraphrase attribute of <entity> and other elements is used to help the human reader.

An <entities-props> element also may contain propositions, marked <prop>. A proposition consists of a relation name used in the definition of argument schemes in our catalogue and the entity ID of its arguments, e.g., *have_pheno(group1, pheno1)*. A set of six semantic relations is used: *have_geno, have_pheno, have_protein, difference, similar*, and *cause*. Propositions may be negated. Although entities and relations were manually extracted, this is a stop-gap approach until NLP tools can assist in semantic extraction.

After <entities-props> elements are added, any arguments conveyed in the preceding <content> element are added. Argument annotations are not added inside of <content> elements due to the problems noted in the Introduction. In other words, separating annotation of semantic interpretations from the source text, and separating annotation of arguments from semantic interpretations provides the flexibility to overcome those problems. The decision was made to place <argument> elements immediately following the elements conveying them (rather than, e.g., at the end of the document) to preserve the context of the argument, since context informs dialectical structure and may constrain recognition of argument schemes.

A key attribute of an <argument> is the name of the scheme in our catalog of argument schemes. To assist the annotator, the schemes in the catalog are organized in a taxonomy (shown in Figure 3), defined, and accompanied by one or two examples. Most of the schemes involve causation; the causal schemes are differentiated first by whether the conclusion is based upon observation of one group or a comparison of two groups of individuals.

1. Causation
  1.1 One Group
    1.1.1   Agreement Arguments
      1.1.1.1 Agreement
      1.1.1.2 Failed Method of Agreement (effect)
      1.1.1.3 Failed Method of Agreement (cause)
    1.1.2   Eliminate Candidates
    1.1.3   Explanation-Based
      1.1.3.1 Effect to Cause
      1.1.3.2 No Effect to No Cause
      1.1.3.3 Consistent with Predicted Effect
  1.2 Two Group
    1.2.1   Difference
    1.2.2   Analogy (Causal)
    1.2.3   Explanation-Based
      1.2.3.1 Consistent Explanation
      1.2.3.2 Difference Consistent Explanation
2. Other
  2.1 Classification
  2.2 Confirmation

Figure 3: Taxonomy of argument schemes

An <argument> element consists of the <premise-list>, a list of <premise>s, and a <conclusion>. For example, Argument 1 shown in Figure 2 is an instance of the Agreement argument scheme from our catalogue. Its premises are copies of two

<prop> elements, derived from two different <content> elements. Its conclusion, labeled *inferred-prop*, has been inferred by the annotator based upon constraints of the Agreement argument scheme. To paraphrase the first premise, the phenotype[2] of group1 (a certain group of mice) is pheno1 (an ataxia-like disorder). The second premise is that the genotype of group1 is geno1 (a mutation on chromosome 6qE1). According to this argument scheme, one may defeasibly conclude that the cause of pheno1 in group1 is geno1. Note that all of the annotated conclusions are of the type *cause(genotype, phenotype, group)*.

Other aspects of our annotation scheme not illustrated in Figure 2 are illustrated (and underlined) in Figure 4. Implicit premises marked as *domain-prop* are reconstructed by the annotator based on domain knowledge that the reader is assumed to possess and which are required by the argument scheme. In Figure 4, the annotator supplied the domain knowledge that geno2a (a homozygous mutation of Itpr1, a gene on chromosome 6qE1) is similar to geno1 (a homozygous mutation on 6qE1). Some premises may be tagged with an optional *conclusion-of* attribute to indicate when the premise is a conclusion of a preceding argument. In Figure 4, the second premise is the inferred conclusion of argument 1.

```
<argument ID="2" scheme="Analogy">
<premise-list>
<premise prop="have_pheno(group1, pheno1)" />
<premise prop="cause(geno1, pheno1, group1)"
        conclusion-of="ARG 1 " />
<premise prop="have_pheno(group2, pheno2)" />
<premise prop="similar(pheno2,pheno1)" />
 <premise prop="cause(geno2, pheno2, group2)" />
 <premise domain-prop="similar(geno2a,geno1)" />
<conclusion inferred-prop="cause(geno2a, pheno1,
        group1)" />
</argument>
```

Figure 4: Argument with two implicit premises

---

[2] *Phenotype* describes a deleterious effect on an organism. *Genotype* describes a variation at the level of chromosome or gene that may have a deleterious effect.

Due to the preliminary, experimental nature of this annotation effort, it did not seem essential to adopt a particular tag set used by other researchers. However the <entities-props> elements were designed so that they could be automatically transformed into a Prolog knowledge base like the one used in (Green, 2018) for argument mining, and the structure of <argument> elements reflects the structure of logic programming rules used for argument mining in that work. Furthermore, at this stage of our research, we were more concerned with identifying relevant argumentation features to be annotated, rather than XML coding style.

3   Results and Discussion

We annotated 15 arguments in the Results/Discussion section of the article -- instances of seven schemes from our catalogue. In decreasing order, the number of instances of each are as follows: Agreement (4), Difference (4), Analogy (2) Consistent Explanation (2), Failed Method of Agreement (effect) (1), Eliminate Difference (1), Difference Consistent Explanation (1). In addition, we found two fairly domain-specific arguments, e.g. about the proportion of phenotypes predicted by Mendelian genetics, not represented in the catalogue.

In order to annotate the arguments, we also annotated 27 discourse <entity> elements (instances of nine entity types: *human, mouse, chromosome, gene, variant, gene product, gene function, disorder*), and 41 proposition (<prop>) elements. Two (implicit) premises of arguments were marked as *domain-prop* and five premises were conclusions of previous arguments. Nine of the 15 arguments had implicit conclusions.

Practically speaking, manual annotation of discourse entities was the most difficult aspect of the annotation process. It was difficult to keep track of coreferential entity IDs due to the number of <entity> elements. Furthermore, it was sometimes necessary to annotate discourse entities which were indirectly introduced. For example, the text introduced a discourse entity that could be described as a specific mutation of the gene Itpr1, namely Itpr1opt/opt; then a

subsequent argument referred to the related entity, some mutation of Itpr1, i.e., a generalization of Itpr1opt/opt.

We are aware of limitations of this work, due to a lack of resources (annotators and domain experts), and welcome collaboration with other researchers to address them. First, the schemes in the catalogue have not been rigorously evaluated for inter-annotator agreement. However, a previous study (Green, 2017) suggested that some of the schemes, such as Agreement, as well as implicit conclusions of arguments, could be consistently identified. The current catalogue improves upon the materials used in our previous study. Still, work remains to evaluate (and possibly refine) the definitions of the argument schemes in the new catalogue.

Second, as noted earlier we have employed manual annotation of entities and propositions as a stop-gap effort until the articles can be preprocessed by NLP tools. BioNLP is an active area of research and our hope is that in the near future this step can be automated or semi-automated. Also, after annotating this article we became aware of BEL (Fluck et al., 2016), a formal language for describing causal relationships in biology, and are interested in exploring its use for expressing the propositional content of arguments in this domain.

In future work, we would like to analyze arguments in other articles in this subfield as well as in another subfield of genetics, such as evolutionary biology, and extend the present argument scheme catalogue as required. We welcome collaborators to work with us on that as well. The corpus could be used to derive semantic rules for argument mining.

4   Related Work

Most previous computational research on arguments in scientific discourse addressed something different than what we mean by 'argument'. That work is concerned with how an author justifies the publication of his article and positions it with respect to previous claims in his field (Teufel, 2010). It also covers the different functions of text segments in scientific communication, such as reporting the method or results (Liakata et al., 2011). In contrast, we are interested in arguments that present the author's scientific reasoning for validation by other scientists.

There are some correspondences between argument structure and discourse structure induced by certain text coherence relations in models such as Rhetorical Structure Theory (RST) (Mann and Thompson, 1988). However, standard text coherence models are challenged by the existence of arguments with non-contiguous, overlapping, embedded, or implicit components. Also, coherence relation definitions do not encode distinctions among argument schemes. Identification of argument schemes is necessary for evaluating argument acceptability, and for inferring implicit components. In earlier work (Green, 2010), we tried to adapt RST to overcome these problems for the description of arguments in short documents for non-experts about medical conditions. In addition to relaxing text constraints of RST, we annotated the RST analyses with argument schemes. It is not clear though whether this approach could adequately represent the structure of a full-text scientific journal article.

There has been little work addressing argument mining of scientific journals. White et al. (2011) annotated part of the CRAFT corpus with functional labels similar to those of (Liakata et al., 2011) and suggested that patterns of labels might be used to recognize arguments. Mercer's group (2016) is attempting to mine text of biomedical publications as a step towards extracting components of the Toulmin (1998) model of argument. Kirschner et al. (2015) annotated text segments in a corpus of educational research articles. Argument schemes were not annotated. It would be interesting to re-analyze that corpus to compare the types of arguments in it with the types of biological/biomedical arguments identified in our catalogue.

References

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

J. Fluck, S. Madan, S. Ansari, et al. 2016. Training and evaluation corpora for the extraction of causal relationships encoded in biological expression language (BEL). *Database* Vol. 2016. Article ID baw113; doi:10.1093/database/baw133.

N.L. Green. 2010. Representation of argumentation in text with Rhetorical Structure Theory. *Argumentation*, 24(2): 181-196.

N.L. Green. 2017. Manual identification of arguments with implicit conclusions using semantic rules for argument mining. In *Proc. of 4th Argument Mining Workshop*, pages 73-78.

N.L. Green. 2018. Towards mining scientific discourse using argumentation schemes. *Argument and Computation*, 9(2):121-135. DOI 10.3233/AAC-180038.

C. Kirschner, J. Eckle-Kohler, and I. Gurevych. 2015. Linking the thoughts: analysis of argumentative structures in scientific publications. In *Proc. NAACL/HLT*, pages 1-11.

J. van de Leemput, J. Chandran, M. Knight, et al. 2007. Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans. *PLoS Genetics*, 3(6, e108):113-129.

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

M. Lippi and P. Torroni. 2016. Argumentation mining: state of the art and emerging trends. *ACM Transactions on Internet Technology*, 16(2): Article 10.

W. Mann and S. Thompson. 1988. Rhetorical structure theory: Towards a functional theory of text organization. *Text*, 8(3):243-281.

R. Mercer. 2016. Locating and extracting key components of argumentation from scholarly scientific writing. In E. Cabrio, G. Hirst, S. Villata and A. Wynder (eds.), *Natural Language Argumentation Mining: Processing and Reasoning over Textual Arguments*, Dagstuhl Seminar 16161, April 17-22, 2016.

A. Peldszus and M. Stede. 2016. An annotated corpus of argumentative microtexts. In *Proc. 1st European Conference on Argumentation,* Lisbon 2015, v. 2, pages 801-816.

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

C. Stab and I. Gurevych, 2014. Annotating argument components and relations in persuasive essays. In *Proc. COLING 2014*, pages 1501-1510.

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

S. Toulmin. 1998. *The Uses of Argument*. Cambridge University Press, Cambridge, UK.

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, Cambridge, UK.

E. White, K.B. Cohen, and L. Hunter. 2011. The CISP annotation schema uncovers hypotheses and explanations in full-text scientific journal articles. In *Proc. of the 2011 Workshop on Biomedical Natural Language Processing*, ACL-HLT 2011, Portland, OR, USA, June 23-24, 2011, pages 134-135.