Modeling Arguments in Scientific Papers to Support Pharmacists
Jodi Schneider, Carol Collins, Lisa E Hines, John R Horn, Richard Boyce

To cite this version:

HAL Id: hal-01076327
https://hal.archives-ouvertes.fr/hal-01076327
Submitted on 21 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The scientific literature is a well-known source of arguments, which are embedded in natural language publications. Ontologies and knowledge bases have become increasingly valuable sources of knowledge synthesized from an ever-increasing body of scientific publications [Renear and Palmer2009]. We present a motivating scenario for using argumentation within a scientific knowledge base.

The Drug Interaction Knowledge Base (DIKB) [Boyce et al.2009] is a hand-constructed evidence-base of safety issues that may occur when two drugs are co-prescribed and taken together. It collects evidence about pharmacokinetic drug interactions for over 60 drugs. This sort of evidence can be used to make clinical decisions on whether it is safe to prescribe two drugs together. This is a complex decision-making environment: decisions are made under uncertainty and reasonable people may come to different conclusions, in effect disagreeing with each others’ decisions [Boyce et al.2009]. Using formal argumentation structures could make the assumptions and logical steps of reasoning open to examination, making the resulting knowledge base more robust and trustworthy.

Currently, the DIKB (Version 1.2) has no way to explicitly model the data and methods in papers, which may need to be reviewed as new knowledge comes to light. The SWAN Discourse ontology it uses only models evidence in the form of citations.
to an entire paper, with swanco:citesAsSupportingEvidence and swanco:citesAsRefutingEvidence¹ [Ciccarese et al.2008]. Fortunately, a new model is now available for explicitly modeling data and methods.

Recently, [Clark et al.2014] proposed a new semantic model for claims, evidence, arguments and annotation in biomedical communications. This model, the Micropublications ontology,² is inspired by the Toulmin model of argumentation (and Verheij’s formalization thereof [Verheij2009]) in order to organize citable claims from scientific papers into networks of agreement and disagreement. A minimal micropublication is a claim with attribution—as in earlier models such as nanopublications [Groth et al.2010]. Unlike those models, Micropublications enable explicitly representing the warrant, i.e. the materials and methods that were used to collect the data. The data collected is considered meaningful due to a subfield’s approval of the materials and methods. This approval is the backing of claims made in everyday science (e.g. ‘paradigm’ science in the terms of [Kuhn2012]). With this model, truth-bearing Statements (possibly with Qualifiers) from biomedical communications are explicitly connected to the Data, Material, and Methods. Attribution also has a place (for instance justifying use of a method by citing it in the literature). A subargument can be represented as a micropublication for a related claim, however far back we wish to extend the network.

In ongoing work, we are focusing on creating micropublications through manual annotation. Currently we are engaging with pharmaceutical curators to envision the most appropriate way to record evidence for the next generation of the DIKB knowledge base. Argumentation will play a key part in this model, especially for auditing the available evidence as the state of knowledge changes. In particular, with the Micropublications model, “methods and materials later found to be flawed might be easily traced to claims based upon them” [Clark et al.2014]. This motivates us to experiment with Micropublications as the underlying model, using a Web-based

¹http://swan-ontology.googlecode.com/svn/tags/1.2/swan.owl
²http://purl.org/mp/
annotation tool (Domeo\textsuperscript{3}) to help a panel of three drug interaction experts create the formal markup of a Micropublication. To determine what evidence should be modeled, we are drawing on new guidelines for systematically evaluating drug-drug interaction evidence [Scheife et al.2014]. The knowledge representations that result from this work will enable people to audit the evidence from scientific papers.

Subsequent automation in constructing such knowledge representations is also conceivable. Previous work has automatically categorized scientific discourse, often using rhetorical zoning approaches [Teufel1999], and a human annotated corpus can be used as the basis for predicting patterns on new examples, with supervised machine learning. Machine learning systems such as SAPIENTA [Liakata et al.2012] have achieved reasonable accuracy in recognizing Hypothesis, Motivation, Goal, Object, Background, Method, Experiment, Model, Observation, Result and Conclusion in highly-structured texts from chemistry and biochemistry. Previous work studied the Conclusion category and found that SAPIENTA’s automatically mined efficacy/effectiveness claims compared favorably to manual annotation from the same abstracts [Boyce et al.2013]. Systematically evaluating drug-drug interaction evidence is a difficult problem, requiring human oversight. For identifying and recording complex evidence, mixed-initiative systems (e.g. [De Liddo et al.2012]) have been suggested: machine annotation can draw the attention of human curators to the zones most likely to contain the data and methods that provide evidence about a harmful drug-drug interaction. Our goal is to transform natural language papers, with the manual work of expert curators, into elaborated claim-argument networks.

In the future we plan to test what sort of machine annotation can aid human curators in extracting the evidence recorded in papers and whether expressing the current evidence on pharmacokinetic drug interactions as micropublications facilitates searching and updating knowledge bases.

\textsuperscript{3}http://swan.mindinformatics.org/
Acknowledgements

Thanks to the reviewers for helpful comments. This work was carried out during the tenure of an ERCIM “Alain Bensoussan” Fellowship Programme. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 246016. Funding is also provided by a grant from the United States National Library of Medicine: “Addressing gaps in clinically useful evidence on drug-drug interactions” (1R01LM011838-01).

References


ArgDiaP 2014: From Real Data to Argument Mining


