Disciplinary Differences in Citation Opinion Expressions

Bei Yu, Syracuse University
Feifei Zhang, Syracuse University

Abstract
This study examines academic opinion expressions in citation context. We first developed an annotation schema to annotate three aspects of each academic opinion expressed in a citation statement: rhetorical purpose, content aspect, and opinion polarity. We then annotated two samples: a natural science sample consisting of biomedical journal articles, and an engineering sample consisting of conference papers in the natural language processing field. A comparison of the annotations on the two samples showed disciplinary differences in citation opinion expressions. The result contributes to the understanding of academic opinion expressions in citation context and the development of automated citation opinion analysis tools to assist researchers' literature search and navigation.
Keywords: Citation Analysis; Opinion Mining; Natural Language Processing; Biomedicine
Citation: Yu, B., Zhang, F. (2015). Disciplinary Difference in Citation Opinion Expressions. In *iConference 2015 Proceedings*.

Acknowledgements: This project is partly supported by IMLS grant RE-07-14-0015. Thanks to Kyra Nay and Ehsan Sabaghian for discussions and annotations. Thanks to Colin Connors for proofreading this document.
Research Data: In case you want to publish research data please contact the editor.
Contact: byu@syr.edu, ffzhang09@syr.edu

1 Introduction

Researchers express their evaluative opinions toward peer work in their citation statements. For example, accordance was expressed in the statement, "*The clearer visibility of this reduction at 24 months when compared with 12 months* ***is consistent with*** *the findings of other interventions [26] and suggests that this positive behaviour change is being strengthened over time*". In another example, a negative critique was expressed in the statement, "***Although*** *the short term results for effectiveness of these new generation devices are* ***encouraging****, [30 31 32] their long term safety profile in people with and without diabetes is still* ***ill defined***".

Researchers also retrieve and summarize these academic opinions following citation links in order for comprehensive literature review. However, this task becomes increasingly challenging with the rapid expansion of academic literature. Current full-text bibliographic databases provide citation links, but no further citation context analysis is offered to help researchers find the most relevant citations and analyze the embedded opinions.

To address this problem, a number of studies have tried to develop intelligent tools to automate citation opinion analysis (e.g. Garzone & Mercer, 2000; Teufel et al., 2006; Dong & Schäfer, 2011; Athar & Teufel, 2012; Jochim & Schutz, 2012; Abu-Jbara et al., 2013). Most of the work has focused on publications in the Natural Language Processing (NLP) discipline. However, Hyland (1999) found that citation opinion expressions are discipline-dependent. For example, negative citations are more rare in natural sciences than in social sciences. Therefore, identifying disciplinary differences is an important task for developing cross-discipline citation opinion analysis tools.

This study addresses the research question, "From the perspective of citation opinion expression, what are the disciplinary differences between publications in science and engineering?" We first developed a schema to annotate three dimensions of each academic opinion expressed in a citation statement: rhetorical purpose (e.g. comparison, critique, use, or information), content aspect (e.g. goal, method, data, or claim), and opinion polarity (positive, negative, or neutral). We then annotated two samples: a natural science sample consisting of biomedical journal articles, and an engineering sample consisting of conference papers in the NLP field. After the annotation we compared the disciplinary differences of citation opinions based on their frequency distributions along the three dimensions and interpreted the implications.

2 Method

In this study, we focus on comparing the differences between natural science and engineering disciplines. Biomedical literature is chosen to represent natural science and NLP for engineering discipline. Journals are the main publication venues in biomedicine. Sample biomedical articles were chosen from the top journals *Lancet* and *The BMJ* (formerly known as the *British Medical Journal*). In comparison, the sample

NLP papers were selected from conference proceedings, the main venues for NLP research. To date we have annotated 49 biomedical articles and 20 NLP articles. We are in the process of annotating more articles, especially in the domain of NLP.

The annotation schema includes three dimensions: aspect, purpose, and polarity. Citation aspect refers to the content aspect of the cited work that is discussed in the citation, including research goal, method, data, claim, and general background. Citation purpose interprets the reason why the citation is needed, which includes four categories: comparison, critique, use, and information. Any citation purpose can be combined with any citation aspect, for example, "comparison:method". The polarity dimension is closely related to the other two dimensions. Comparisons and critiques are explicitly expressed opinions with evaluative outcomes. For comparisons, the polarity outcomes include accordant vs. discordant claims, advantageous vs. disadvantageous methods/data, and similar vs. different goals/methods/data. For critiques, the polarity outcomes can be positive, negative, or a combination of the two categories (mixed category). The polarity for the "use" cases is usually neutral, e.g. "*We used the method developed by [1]*"; however, it can be interpreted as an implicitly positive opinion toward [1]. The polarity for the "information" citations is usually neutral as well.

Each article was annotated by two graduate students. All annotations were conducted using GATE (Cunningham et al., 2011).

3 Result

The inter-coder reliability test results showed substantial agreement on citation purpose with Cohen's *Kappa* at .64 for the biomedical sample and .68 for NLP. The polarity annotation is at a similar agreement level in that once the purpose is agreed upon, the polarity is always agreed upon too. The agreement on citation aspect is lower with moderate Cohen's *Kappa* at .45 for the biomedical sample and .38 for NLP. Because none of the student annotators had backgrounds in either biomedicine or NLP, this result suggests that domain knowledge may affect the identification of citation aspect, but not purpose or polarity.

The disagreements were resolved in group discussions with guidance from a senior NLP researcher. In the resolved annotations, the category distributions are reported in Tables 1-4.

Discipline	Claim	Method	Data	Goal	General	Total
BM	38.8	30.8	6.2	.8	23.5	100
NLP	13.2	66.0	10.4	5.2	5.2	100

Table 1. Citation Aspects (in percentage)

Discipline	Comparison	Critique	Use	Info	Total
BM	15.2	9	22.2	53.6	100
NLP	22.4	10	19.2	48.4	100

Table 2. Citation Purposes (in percentage)

Discipline	Acc	Disc	Acc/Disc	Adv	Disadv	Adv/Disadv	Sim	Diff	Sim/Diff	Total
BM	35.8	26.8	4.3	12.5	3.4	.6	6.5	8.6	1.5	100
NLP	0	8.9	1.8	14.3	1.8	0	16.1	44.6	12.5	100

Table 3. The Polarity Outcomes of Comparison (in percentage)

Discipline	Positive	Negative	Mixed	Total
BM	22.8	72.5	4.8	100
NLP	56.0	40.0	4.0	100

Table 4. The Polarity Outcomes of Critique (in percentage)

Table 1 demonstrates several disciplinary differences: biomedical articles cited many more claims (38.8% vs. 13.2%) and general backgrounds (23.5% vs. 5.2%), but many fewer methods (30.8% vs. 66.0%), data (6.2% vs. 10.4%), and goals (.8% vs. 5.2%) than NLP papers. These differences reflect the characteristic of biomedical research, which is usually based on hypothesis testing, and the main goal of

NLP research, which is developing new computational approaches to help computers understand human language.

Table 2 shows similar citation function distributions in the two disciplines. Comparing biomedical to NLP articles, the former used fewer comparisons, more "use" and "info" citations, and a similar percentage of critiques; however, all differences are within 1-7 percentage points.

Table 3 shows two stark differences in the polarity outcomes of comparisons. First, "accordance" and "discordance" occurred much less in NLP, because they are tags to describe the comparison of two claims. As discussed above that claims are cited much less often in NLP. Second, "discordance" and "difference" outnumbered "accordance" and "similarity" to great extent in NLP (8.9% vs. 0% and 44.6% vs. 16.1%), compared to the relatively balanced ratio in biomedicine (26.8% vs. 35.8% and 8.6% vs. 6.5%). This is a particularly interesting finding. A possible explanation is that method innovation is the most important contribution that NLP research can claim, and thus researchers need to describe the innovation in their methods compared to existing methods. In comparison, different biomedical researchers may conduct similar experiments (e.g. clinical trials of a drug) to compare their findings with previous works, and both accordant and discordant findings can be published.

Table 4 shows clear differences in the polarity outcomes of critiques. The majority of NLP critiques are positive (56.0%). In comparison, more than two-thirds of biomedical critiques are negative (72.5%). An examination of the citation context shows that NLP researchers often acknowledged existing studies positively before describing the different research goals or methods of their own, resulting in avoiding negative critiques of previous methods.

4 Conclusion

This study found a number of disciplinary differences in citation opinion expression in natural science publications (represented by biomedical articles) and engineering publications (represented by NLP articles). The two disciplines demonstrate some similar distributions in citation purposes, but exemplify many differences in citation aspect and polarity. The differences can be interpreted as that scientific articles focus on hypothesis testing, and both accordant and discordant findings are valued; in contrast, engineering papers focus on developing new methods/tools, and authors tend to balance the need for claiming innovation and the need for paying homage for prior work by maintaining positive tones in literature review. The findings contribute to the understanding of researchers' citation behavior and may be useful for developing automated citation opinion analysis tools. The limitations of current work, including the relatively small NLP sample and annotators' lack of domain knowledge, will be addressed in future work.

References

Athar, A. & Teufel, S. (2012). Context-enhanced citation sentiment detection. Proceedings of NAACL-HLT 2012, 597-601.

Abu-Jbara, Amjad, Jefferson Ezra, and Dragomir Radev. (2013). Purpose and polarity of citation: Towards nlp-based bibliometrics. Proceedings of NAACL-HLT 2013, 596-606.

Cunningham, H., et al. Text Processing with GATE (Version 6). University of Sheffield Department of Computer Science. 15 April 2011

Dong, C., and Schäfer, U. (2011). Ensemble-style self-training on citation classification. Proceedings of IJCNLP 2011, 623-631.

Garzone, M. & Mercer, R.E. (2000). Towards an automated citation classifier. Advances in Artificial Intelligence, Lecture Notes in Computer Science 1822, 337-346.

Hyland, K. (1999). Disciplinary discourses: Writer Stance in Research Articles', in C. Candlin and K. Hyland (eds) Writing: Texts: Processes and Practices, 99–121.London: Longman

Jochim, C. and Schutz, H. (2012). Towards a generic and flexible citation classifier based on a faceted classification schema. Proceedings of COLING 2012, 1343-1358, 2012.

Teufel, S. Siddharthan, A., & Tidhar, D. 2006. An annotation scheme for citation function. Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, 80-87, Sydney, July 2006

Table of Tables