

Hidden Gems in the Wikipedia Discussions: The Wikipedians' Rationales

Lu Xiao

Faculty of Information & Media Studies, Department of Computer Science,
The University of Western Ontario, London, ON, CANADA, N6A 3B7
lxiao24@uwo.ca

Abstract

This paper presents a series of completed and ongoing studies that are aimed at understanding the role of the Wikipedians' rationales in the Wikipedia discussions. We define a rationale as one's justification of her view-point and suggestion. Our studies demonstrate the potential of leveraging the Wikipedians' rationales in the discussions as resources for future decision-makings and as resources for eliciting knowledge about the community's norms, practices and policies. Viewed as the rich digital traces in these environments, we consider them to be beneficial for the community members, such as helping newcomers familiarize themselves on the commonly accepted justificatory reasoning styles. We call for more re-search attention to the discussion content from this rationale study perspective.

Introduction

This research program is interested in closing a research gap - the role of the Wikipedia participants' rationales in the discussion activities and in the Wikipedia organization. The interested rationales are those statements that the Wikipedians have provided in the discussions to justify their viewpoints. In our studies of small and large group activities, we found that the rationales provided by the others can make an impact on the individual (Xiao, 2012; Xiao & Carroll, 2013; Xiao & Carroll, 2015; Xiao, 2014).

As a first step to fill this gap, we have been focusing on one type of Wikipedia discussions - the article for deletion (AfD) discussions. We have conducted a few studies about the rationales in the AfD discussion context. We are still yet at the preliminary stage of our long-term research agenda to understand the effects of the Wikipedian's rationales in the organization. In this paper, we briefly review the completed and ongoing studies to illustrate the potential of these rationale.

Our AfD Studies

To study the rationales in Wikipedia's AfD discussions, we first analyzed a sample of the AfD discussions to understand the types of rationales used in the argument, the factors that contributed to the outcome of the discussion such as the article topic, the votes other than keep and delete, and the unanimity (or lack thereof) (Xiao & Askin, 2014). We then conducted an online questionnaire study to reach out those who have experiences of participating in the AfD discussions to probe their perceived effects of the others' rationales in the discussions (Xiao & Askin, 2015).

The amount of AfD discussion content in Wikipedia is enormous for the traditional content analysis approach through human annotation. Recognizing this challenge, we developed a tool that detects and extracts representative rationales of a discussion using the semantic similarity technique and our sentiment polarity analysis algorithm (Mao, Xiao, & Mercer, 2014). We are in the process of evaluating this tool by comparing its output with a human annotated set of representative rationales for a given sample. We hope that by leveraging the computational techniques like this we will be able to analyze a larger dataset than our 2014 study to probe the types of important rationales and how they influence the final decisions.

Our analysis of the AfD discussions suggest that in these discussions experts often are directive in their interactions. They sometimes suggest ways of proceeding, and sometimes monitor and flag inappropriate arguments. To understand these directive rationales, we developed a computational technique that identifies and extracts imperatives from the discussions. Imperatives are one type of directive speech act which often express suggestions, or requests that ask someone to do or not to do something (Mao, Mercer, & Xiao, 2014).

To obtain a more comprehensive understanding about the AfD discussions, we developed a database tool (Xiao &

Javarnmardi, 2016). With this tool, we to parse and organize two years of the AfD discussions, and are developing visualization tools to present the use of different policies.

Future Work

Our studies of the Wikipedians' rationales in the AfD discussions illustrate that these rationales not only contribute to the decision-making process of the current article but also contain knowledge that may be reused in the other articles' discussions and that reveal the community's norms, practices, and the common policies that are critical in the quality monitoring and control. We envision the development of tools to help discover such knowledge, e.g., by automatically detecting the community liked/disliked argumentative strategies, examining the patterns in a user's AfD rationales across different article discussions and over time, or exploring the development of the Wikipedia community's reasoning styles and examining how they are similar or different from the other online communities'. Viewing these AfD discussions as a collection of rationales that reflect how people offer their justifications, one can leverage this resource for a variety of studies related to the online participants' reasoning behavior.

A major advantage of studying rationales in the AfD discussions is that these discussions, by default, consist of the participants' opinions and rationales and they follow a common structure: an opinion (e.g., keep, delete, redirect) is followed by a rationale. In other Wikipedia discussions (e.g., Talk page discussions), the discussions are more informal and like any online discussions. In such context, detecting the rationales is a challenging task that needs to be completed in order to explore the effects of the rationales in these environments. We have been exploring a computational approach for this task (Khazaei & Xiao, 2015; Khazaei, Xiao, & Mercer, 2015). Being able to detect the rationales from the online discussions will bring a lot of potential to explore the role of the rationales. Are the Wikipedia discussions that contain more rationales tend to have better outcome? Are there differences in terms of the participants' justificatory reasoning styles when they are engaged in the article talk page discussions versus in the committee discussions?

As a community, Wikipedia offers a virtual space for people to freely articulate and share their rationales. In the author's view, these rationales function as superglues that tie different parts together in the organization, from the policies to the actual practices, from the article editing to the committee forming and functioning, etc. It is the goal of this paper to bring more research attention to these rationales, the "hidden treasures" in this virtual space, and leverage the full potential they bring to the community.

Acknowledgement

This work is supported by the Natural Sciences and Engineering Research Council of Canada, Discovery program.

References

- Khazaei, T., Xiao, L., and Mercer, R. 2015. Identification and Disambiguation of Lexical Cues of Rhetorical Relations across Different Text Genres, *Workshop "Computational Semantics: Linking Lexical, Sentential and Discourse-level Models"(LSDSem) at 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, <http://www.emnlp2015.org/proceedings/LSDSem/pdf/LSDSem06.pdf>
- Khazaei, T., and Xiao, L. 2015. Corpus-based Analysis of Rhetorical Relations: A Study of Lexical Cues, In *Proceedings of the IEEE International Conference on Semantic Computing (ICSC)*, 417-423
- Mao, W. T., Mercer, R., and Xiao, L. 2014. Extracting Imperatives from Wikipedia Article for Deletion Discussions, *First Workshop on Argumentation Mining at the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, <http://acl2014.org/acl2014/W14-21/pdf/W14-2117.pdf>
- Mao, W. T., Xiao, L., and Mercer, R. 2014. Using Text Similarity and Sentiment Analysis to Identify Representative Rationales in Large-Scale Online Deliberations, *5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis at the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, <http://acl2014.org/acl2014/W14-26/pdf/W14-2624.pdf>
- Xiao, L. 2012. The Effects of a Shared Free Form Rationale Space in Collaborative Learning Activities, *Journal of Systems and Software*, 86(7), 1727 – 1737
- Xiao, L., and Carroll, J. M. 2013. The Effects of Rationale Awareness on Individual Reflection: Processes in Virtual Group Activities, *International Journal of e-Collaboration*, 9(2), 78 – 95
- Xiao, L., and Carroll, J. M. 2015. Shared practices in articulating and sharing rationale: An empirical study. *International Journal of e-Collaboration*, 11(4), Article 2, 11-39.
- Xiao, L. 2014. Effects of rationale awareness in online ideation crowdsourcing tasks, *Journal of the American Society for Information Science and Technology*, 65, 1707-1720,
- Xiao, L., and Askin, N. 2014. What Influences Online Deliberation? A Wikipedia Study, *Journal of the American Society for Information Science and Technology*, 65: 898–910
- Xiao, L., and Askin, N. 2015. Rationale Sharing in Large-Scale Online Deliberations, *Proceedings of iConference*, <http://hdl.handle.net/2142/73741>
- Xiao, L., Javarnmardi, A. (2016), A Relational Database for Analyzing Wikipedia's Article for Deletion Deliberations, *2016 Chinese CHI Symposium (May 8, San Jose, CA, USA)*, to appear