

Extracting Imperatives from Wikipedia Article for Deletion Discussions

Fiona Mao

Robert E. Mercer

Department of Computer Science
The University of Western Ontario
London, Ontario, Canada
fiona.wt.mao@gmail.com
mercerc@csd.uwo.ca

Lu Xiao

Faculty of Information and Media Studies
Department of Computer Science
The University of Western Ontario
London, Ontario, Canada
lxiao24@uwo.ca

Abstract

Wikipedia contains millions of articles, collaboratively produced. If an article is controversial, an online “Article for Deletion” (AfD) discussion is held to determine whether the article should be deleted. It is open to any user to participate and make a comment or argue an opinion. Some of these comments and arguments can be counter-arguments, attacks in Dung’s (1995) argumentation terminology. Here, we consider the extraction of one type of attack, the directive speech act formed as an imperative.

1 Introduction

A large group of volunteers participate to make Wikipedia one of the most successful collaborative information repositories. To ensure the quality of the encyclopedia, deletion of articles happens continually. If an article is controversial, an online discussion called “Article for Deletion” (AfD) is held to determine whether the article should be deleted. It is open to any user to participate in the discussion and make a comment or argue an opinion. Some of these comments and arguments can be counter-arguments, attacks in Dung’s (1995) argumentation terminology. A common argumentative attack is a directive speech act suggesting a potential disagreement and a possible way to rectify the matter. Here, we consider the extraction of this type of attack when formed as an imperative.

Researchers are becoming increasingly interested in studying the content of Wikipedia’s Articles for Deletion (AfD) forum. Schneider et al. (2013) investigated the difference in arguments from novices and experienced users. Xiao and Askin (2014) examined the types of rationales in Wikipedia AfD discussions.

2 Speech Acts and Imperatives

A speech act is an utterance that has performative function in communication (Austin, 1975). Of the three types of speech acts, Searle (1976) subcategorized the illocutionary act, the act of expressing the speaker’s intention, into five sub-groups. We are interested here in the Directives sub-group.

Often, a directive can be viewed as an attack (Dung, 1995), albeit an indirect one, e.g., “Could you provide the source to me?”. The user, to whom this directive is made, undercuts (Pollock, 1992) the attack by responding with some sources.

Ervin-Tripp (1976) lists six types of directives one being the imperative. Imperatives express a command. Typically the predicate is an action verb and the subject, often eliminated, is second-person (you). As well, there can be words of politeness and adverbial modifiers of the verb:

- Please do this sort of check in the future.
- Just avoid those sorts of comments and perhaps strike the one above.

Cohortatives (first person plural imperatives) are normally used in suggestions such as, “Let’s have dinner together.” Some directive sentences from AfD discussions are listed below:

- Add the information, and please give us some information so we can judge these sources.
- Let’s avoid compounding the BLP issues caused by the existence of this article, in violation of notability and blp policies, by having it snow-deleted post-haste.
- You must first discuss the matter there, and you need to be specific.
- Perhaps time would be better spent adding more and improving the article rather than just arguing here.
- Instead of complaining, how about finding such content and improving the article?

Viewing the above examples, some users directly suggest or command other users to do something (the first one). Cohortatives include the user (the second example). The third one is obviously commanding someone to discuss the matter first and to be specific. The first three examples are imperatives. Some commands include politeness, as illustrated by the last two examples. Since the form of this kind of utterance varies, it is difficult to define a rule for recognizing it by computer. In this paper, we only detect direct imperatives and leave indirect imperative recognition for future work.

3 Detecting Imperatives

In English, a typical imperative is expressed by using the base form of a verb, normally without a subject. To detect this kind of imperative, we need to analyze the grammatical structure of sentences.

According to our observation, a typical imperative contains a verb in base form without any subject. Therefore, the basic rule for imperative recognition is to find those sentences with a verb (in its base form) as the root in the phrase structure and this particular verb has no subject child in the dependency structure. Another form of imperative is like the sentence: "You must first discuss the matter there, and you need to be specific". We have adapted a modal directive rule suggested by Sinclair et al. (1975): We recognize the use of a personal pronoun or noun (e.g., "you", "we", or a username) followed by a modal verb (e.g., "should", "must", "need") as an imperative. We used keywords to detect this kind of imperative.

4 Evaluation

In this section, we evaluate the performance of our methods to detect imperatives. Two human annotators (undergraduate students at The University of Western Ontario) extracted imperatives from our data. Agreed upon imperatives became our gold standard. Our system had Precision 0.8447, Recall 0.7337, and F-measure 0.7874 on this data.

Most false positives have an implicit subject "I" (e.g., *Agree with most of the rest of this.*), a writing style found in this text genre. Missed imperatives (false negatives) resulted from parsing errors by the parsing tool and sentences with the form of subject + modal verb, but the subject is a noun (person or organization) instead of a pronoun. Our method keyed on pronouns.

5 Related Work

Marsi's (1997) definition of imperative mood is too restrictive for our purposes here. A use of Argumentative Zoning to critique thesis abstracts (Feltrim et al., 2006) gives no details regarding the imperative sentence recognition techniques, and the language of interest is Brazilian Portuguese.

Acknowledgments

This project is partially supported by the Discovery program of The Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- John Langshaw Austin. 1975. *How To Do Things with Words*. Oxford University Press.
- Phan Minh Dung. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357.
- Susan Ervin-Tripp. 1976. Is Sybil there? The structure of some American English directives. *Language in Society*, 5(01):25–66.
- Valéria Feltrim, Simone Teufel, Maria das Graças V. Nunes, and M. Aluísio, Sandra. 2006. Argumentative zoning applied to critiquing novices' scientific abstracts. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, pages 233–246. Springer Netherlands.
- Erwin Marsi. 1997. A reusable syntactic generator for Dutch. In Peter-Arno Coppens, Hans van Halteren, and Lisanne Teunissen, editors, *Computational Linguistics in the Netherlands 1997: Selected papers from the Eighth CLIN Meeting*, pages 205–222. Amsterdam/Atlanta: Rodopi.
- John L. Pollock. 1992. How to reason defeasibly. *Artificial Intelligence*, 57:1–42.
- Jodi Schneider, Krystian Samp, Alexandre Passant, and Stefan Decker. 2013. Arguments about deletion: How experience improves the acceptability of arguments in ad-hoc online task groups. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pages 1069–1080. ACM.
- John R Searle. 1976. A classification of illocutionary acts. *Language in Society*, 5(01):1–23.
- J.M.H. Sinclair and M. Coulthard. 1975. *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.
- Lu Xiao and Nicole Askin. 2014. What influences online deliberation? A Wikipedia study. *J. of the Association for Information Science and Technology*.