Critical Review: A Computational Approach to Politeness with Application to Social Factors

Maia Petee
The University of North Texas
maiapetee@my.unt.edu

1 Introduction

This paper, a recent effort from cross-disciplinary authors including Dan Jurafsky (Stanford professor of computer science and linguistics well-known for his role in the development of the first automatic semantic role labeler, as well as known to this author thanks to his introductory texts on natural language processing), is an initial attempt to continue the groundwork done in discourse pragmatics begun in the 1970s largely by Robin Lakoff (1973) in the identification and discussion of social politeness (Lakoff, 1977).

The increasing completeness, approaching perfection, of NLP tasks that look at language on a more granular level has led to a increasing focus on computational discourse pragmatics, and the recent rise of machine learning classification systems that use advanced algorithms to classify statements based on similarity has led to the success of this task. This article (Danescu-Niculescu-Mizil et al., 2013), written in 2013, boasts a classifier that performs at identifying linguistic politeness at what they label a near-human level of accuracy; here, we will take a brief and critical look at the tools and methods used.

2 Summary

Danescu-Niculescu-Mizil et al. proposed a computational framework that could not only identify polite speech acts with a high level of accuracy, but could draw new connections between politeness markers and their context, due to the amount of data and the computational nature of the task. They also took care to train the system to be domain-(register-) independent using a newly annotated corpus of text (specifically, requests of other users) taken from Wikipedia and StackExchange and annotated for a number of politeness markers, such as indirection, deference, impersonalization, and modality.

The new corpus, also annotated for author, will be used as a starting point for exploring politeness and power relations on a large scale. Danescu-Niculescu-Mizil et al. propose that editors and correspondents relatively higher up in the social strata of each site will display their social power partially through the use of fewer politeness markers when making requests of others. They also hypothesize that formerly polite authors who move up in the hierarchy will taper off in the number of politeness markers they use in requests, reflecting their newly powerful status. Using requests further focuses this study on what is termed ”negative politeness,” meaning speaker strategies that minimize the impact of a given request on the interlocutor (Brown et al., 1987). Examples of negative politeness could include using ”Would you mind,” or ”I’m terribly sorry, but,” when making a request (Brown and Levinson, 1978).

Finally, Danescu-Niculescu-Mizil et al. propose to use the relative wealth of their annotated data (a total of over 10 thousand annotated requests) and the power of their machine learning system to uncover new truths about the morphosyntactic positioning of certain politeness discourse markers, such as ”please.”

They used two classifiers on this annotated data; their baseline was a bag-of-words SVM classifier that examined unigram frequencies of words. The second classifier, also an SVM, took a linguistically-informed approach with the use of features known to load to politeness, such as hedges (”suggest”) or greetings (”hey!”). There were a total of 20 politeness features used. Overall, the classifiers performed quite well, with the BOW hovering in the 75-80% range in-domain and around 70% for a cross-domain task, and the linguistically-informed approach yielding on average a 3% - 4% improvement over the BOW.

3 Critical Analysis

A number of potential methodological concerns stood out to me upon reading this study. First among these is the procedure used for obtaining such a large number of annotated politeness requests. There was a sufficient degree of annotation coverage for each request (at least five independent annotations per request, subjected to a mean pairwise correlation to ensure a reasonable degree of inter-annotator agreement), but this large number of requests was — probably for budgetary reasons — sourced from Amazon's Mechanical Turk service. As such, very few if any annotators were likely trained to recognize linguistic politeness, and they were not prompted to use this definition even if they had a mental representation of it. Instead, they were prompted to use a more intuitive and universal definition of politeness, treating two-sentence requests (one sentence for context and lead-up to a request, and the second being the actual request) as though they came from co-workers and rating them for politeness. It is possible that linguistic politeness and annotators' mental representations of social politeness have a high degree of overlap, but due to cultural variation even within the United States, this overlap cannot be uniform. The researchers seemed to be aware of the potential for methodological leak in annotator selection and variation, because they did engage in a number of vetting procedures to ensure high-consistency annotations, such as selecting for similar linguistic background and linguistic precision (using a paraphrasing task (Munro et al., 2010)). Due to the number of annotations needed to create a corpus large enough to draw their desired conclusions, it is probable that compromises like these were necessary, and that annotator variation was controlled for using a number of reasonable measures. Because of the inherent limitations found in every budgetary and time-scale restriction, all that researchers can do is acknowledge those limitations and control for them with available resources.

The requests to be annotated and added to the corpus were taken from free text, and as mentioned before were limited to two sentences: one for context and to capture some lead-up politeness markers, and one containing the primary request. Without seeing examples from the actual raw corpus, one cannot be sure, but it seems as though the inclusion of a third sentence might capture needed content that is relevant to the act of politeness contained in the request. For example, Sentence 1 might contain preamble or politeness strategies such as displaying gratitude for something an interlocutor has done, Sentence 2 might contain a direct request, and Sentence 3 might reiterate gratitude or deference strategies with something like ”Thank you again,” or ”You do great work.” Again, without examining the source personally, the case for a third context sentence cannot be supported or refuted.

One aspect of the study's methodology that I appreciated was the approach taken toward binary annotator perception. While annotators were asked to rank each request for politeness using a Likert scale, the researchers acknowledged that only answers on either far side of the Likert scale were likely associated with a binary mental perception of an utterance being ”polite” or ”impolite.” They tested this intuition by breaking the requests into four politeness quartiles, and measuring for each quartile what percentage of requests all five annotators agreed upon. The blurry mental boundary between mildly impolite and mildly polite was testified to by the fact that full annotator agreement was significantly more common in the first and fourth (i.e. very polite and very impolite) quartiles. The researchers approached this discrepancy intelligently and used it to operationalize one of the primary politeness metrics used in their analysis: percentage of requests per strategy that fall into this top ”polite” quartile.

Finally, the study's successful feature design allowed for a number of hypotheses to be tested in a vastly scaled-up experiment (this time consisting of over 400,000 requests). Due to the structure of Wikipedia communities, in which admin elections were held (and won/lost) and all user identities are transparent, the researchers could perform a longitudinal politeness analysis on individuals who ”came into” power. The structure of Stack Exchange's reputation system also allowed politeness to be cross-referenced with user reputation. In both of these cases, a negative correlation was found between relative social power and politeness used. The nature of these conclusions, while quantitative, was not verified statistically (using a p-value to prove that the null hypothesis can be rejected); despite this, they are still fascinating.

References

Penelope Brown and Stephen C Levinson. 1978. Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction*, pages 56–311. Cambridge University Press.

Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*, volume 4. Cambridge university press.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*.

Robin Lakoff. 1977. What you can do with words: Politeness, pragmatics and performatives. In *Proceedings of the Texas conference on performatives, presuppositions and implicatures*, pages 79–106. ERIC.

Robin Tolmach Lakoff. 1973. *The logic of politeness: Minding your p's and q's*.

Robert Munro, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen, and Harry Tily. 2010. Crowdsourcing and language studies: the new generation of linguistic data. In *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon's Mechanical Turk*, pages 122–130. Association for Computational Linguistics.