

Joachim Daiber:

Task: Analysis of politeness

Description:

This task explores the politeness in online exchanges. You are given a corpus with information about a specific type of communication (where did it occur, who was the user. etc, a raw text and a politeness rating). Your task is two-fold:

1. Create a classifier that judges the politeness of a text (0.0 very impolite, 1.0 very polite)
2. Suggest a set of changes that the writer of this text could make to make the text seem more polite.

Both parts should be combined in a demo, which indicates the politeness and makes suggestion to the user. (politeness could be indicated by color and words or phrases that have a great influence on the rating could be highlighted).

Suggested method:

A nice approach to this topic would be build a small politeness treebank, similar to the work on sentiment analysis in: Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank (Socher et al, nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf)

The students could use the sentences from the Stanford politeness treebank, apply a good parser to it and add more detailed annotations on every tree node. Based on this treebank, they could then train a statistical model that factors over tree nodes. Having politeness scores for each tree node would simplify the highlighting and suggestions in the demo.

Data:

- Stanford politness corpus: Short paragraphs are annotated with a politeness score, for example: (free to download, <http://www.mpi-sws.org/~cristian/Politeness.html>)

Source	Text	Politeness
Stack Overflow	I see. Was it talking about that one specifically? :)	0.83689817913255971
Stack Overflow	If the struct is redefined, then everything that uses it's definition needs to be recompiled. Is there a rebuild of all modues that know about the struct?	0.17793777464654717

- Random acts of Pizza dataset: <http://cs.stanford.edu/~althoff/raop-dataset/>

