Summary of: Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations

Authors: Bao, Wu, Zhang, Chandrasekharan, Jurgens

The authors propose a suite of operationalizations for measuring various prosocial behaviours, and apply the analysis to a large set of Reddit discussions captured in the year 2017. Information sharing is measured through sharing of URLs for typically informative sites, as well as other sites. Gratitude is measured by counting examples from a fixed lexicon, e.g. "thank you". Esteem Enhancement is measured with a combination of a 'politeness regressor', counts for second person pronouns (e.g. you), and measures of the 'scores' given to comments by the community on Reddit. Social Support is measured using data from a previous study (Wang and Jurgens) to fine-tune a pretrained BERT model to predict supportiveness. Fundraising and Donating is measured again by counting the domains in URLs associated with charitable spending. Mentoring was measured by constructing a dataset of positive and negative examples of mentoring, postitive examples being drawn from subreddits with 'advice' in the name (but not 'bad') and negative examples from all other subreddits, training a logistic regression model on unigram and bigram features extracted from these documents. Finally Absence of Antisocial Behaviour is measured with the Perspective API.

The authors synthesise the above metrics with PCA and find that the first prinicipal component is a useful proxy measure of prosociality generally.

The first component of the PCA as well as other text and non-text features are used as inputs to deep network models to train a predictor of a conversation's prosocial trajectory. It's found that the model's predictions outperform human judgements in terms of predicting the trajectory based only on the submission/post content and a first comment (Top Level Comment, TLC). The performance of the model varies significantly depending on the subreddit, i.e. the topic/style/conventions of the conversation, with lighthearted conversations being easier to predict, and more contentious conversations being harder to predict.

References

Wang, Yafei, Yen, John, and Reitter, David (2015) *Pragmatic alignment on social support type in health forum conversations.* In *Proceedings of the 6th Workshop on Cognitive Modeling and Computational Lingustics.* 9-18.