

Automated opinion detection analysis of online conversations

Yuki M. Asano

University of Oxford

yuki@robots.ox.ac.uk

Niccolo Pescetelli

MIT MediaLab

niccolop@mit.edu

Jonas Haslbeck

University of Amsterdam

jonashaslbeck@gmail.com

1 Introduction

People’s ability to communicate complex concepts through language sets them apart from other animals. We are able to effectively read other people’s *views beliefs* and opinions through their conversational language, mapping the spoken words to a lower dimensional space of concepts. Attempts to replicate this natural language understanding capability has lead to the development and application of artificial Natural Language Processing (NLP). Two applications are topic modeling and sentiment analysis, topic modeling maps text into topics whereas sentiment analysis maps text into an affective continuum (typically; bad vs good). However, people’s opinions are multifaceted, topic-specific, and highly sensitive to context. The limitations of these two dominant methods have left a large part of communication untouched, for example the sentences *Climate change is a hoax.* and *Climate change is upon us.* express opposite opinions, yet have the same sentiment and topic. In this work, we develop a model that fills the missing third of the sentiment-topic-opinion Venn-diagram to enable a more complete automated understanding of human language.

2 Method

Data: We have chosen ‘abortion legalisation’ as the topic to analyse. This issue has been historically presented within two polarised camps within public debate: Pro-life (anti-abortion) and Pro-choice (pro-abortion). Both ‘camps’ carry specific arguments, making it a good starting-point to train our model. As there is no dataset which includes sentences with their associated opinions, we collect the training data for our model from the sub-reddit pages *pro-life* and *pro-choice*, yielding each about 50k sentences with noisy labels. By substituting synonyms we quadruple our dataset to

200k sentences and achieve better performance on the held-out test-set of classification accuracy of around 72%.

We test our model also on abortion related *Change my View* discussions of reddit (CMVs) and also on a real (fictional) debate between representatives from each camp for which we implemented an additional speech-to-text system.

Model: We use a Bi-directional Gated Recurrent Unit (GRU) with learned word-embeddings followed by a fully connected layer to classify the opinion of a two concurrent sentences. Our architecture parameters are:

3 Initial Results

We have analysed CMV threads and find diverse behaviors with some *qualitatively* showing highly correlated opinion expression, although no formal test have been made. A screen-capture of our speech-to-text system is available [here](#).

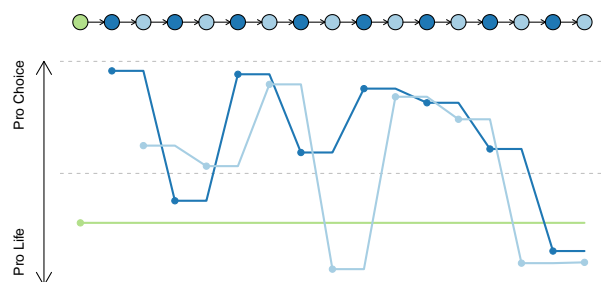


Figure 1: Example CMV conversation chain with two users in a dialogue.

4 Conclusion

We have shown the feasibility of classifying opinion in an automated way using machine learning. Our method can be extended to analyze political debates and polarization or serve as a tool for journalists for promoting a healthy online debate.