Troy Kelley, Nidhi Sinha & Hana La

6 April 2021

Natural Language Processing

Adam Meyers

Final Project Proposal

For this project, we are interested in creating an annotated dataset for online-dating conversations. We will use standardized annotation guidelines and tags to ensure a final product which is as consistent as possible.

Much like Olshefki’s annotation work with Esports, there are currently no publicly available datasets which primarily focus on dialogue within the realm of online-dating despite its growth and influence on present-day relationships, so we wish to provide this resource to those in the future who might have a need for a dataset of this kind (Olshefski 2015). We will be sourcing this data from several online websites where screenshots of user’s conversations are widely available and shared, such as from reddit.com/r/tinder. We will also be exploring other resources from similar online forum type websites and human sources, as well as comparing and discussing the best manner of annotation of this corpus.

For the task of annotation and corpus building, we only plan on doing this ourselves, as we suspect the availability of willing and free persons within our sphere of influence will not be very large, or existent at all. We recognize that the final size of the dataset may be smaller than what might typically be usable for larger NLP systems using it for training. We do not see this as a large problem for the actual completion of the task, as our main focus is creating a corpus, and learning more about how non-formal dialogue is mapped to formal linguistic actions.

We plan on tagging the dialogue based on two main categories: interaction groups and utterance categorization. We will define several interaction groups as subsets of a larger one. We will first group the same agent’s consecutive messages into a ‘session’. Within each session are one or more *messages* that the agent has sent. And within each message is an utterance or sentence. We break down communications in this way since messaging dialogue is never exactly one-to-one in each agent’s utterances. The instigator of this conversation will be labelled as agent A and the other person will be labelled as agent B.

Secondly, we will tag specific utterances with corresponding action tags, relating to the general type of action this utterance performs within the context of the dialogue. We will use a subset of tags as described in Stolcke et al, 2000 that are most relevant to our context of purely written dialogue. These tagging approaches are directly influenced by the work done in Ivanovic (2005), where they used the same schema to mark instant messaging dialogue on customer support pages. In this work they also attempt to resynchronize messages so that question-response type pairs are formally matched with each other, however we will not be taking this into account and will only be tagging the aforementioned categories.

In addition to these standard action tags, we will add more action tags to the possible set, such as “unconventional opening”. We will also label sentences for their general sentiment as understood by the annotator. Sentences can contain multiple sentiments within different categories: general, comedic tone, unsolicited-information-question, and unsolicited-information-response. We find it a particularly interesting part of our research to include these tags as part of the annotation, since these are largely unique to this form of dialogue, influenced by the annotation work done in Napoles, et. al. (2017), and will be useful in sentiment analysis or other research done specific to online dating conversations. To maintain accuracy to the original dialogue, we will not perform any spelling corrections, and any emoticons (emojis) used will be transcribed with its CLDR short name listed in the unicode table of emojis (<https://unicode.org/emoji/charts/full-emoji-list.html>) with the prefix of “EMOJI\_”.

We will aim to annotate as many unique conversations as possible within our time frame. However in order to provide some type of evaluation metric, our team will also annotate a small subset of the same conversations and compare results between them. The evaluation metrics we will use between a conversation annotated by different annotators will be the Kappa score, or inter-annotator agreement. A larger kappa score will indicate more agreement between annotators, and meaning the annotators are ‘good’ or the dataset is easily agreed upon. For a conversation with a good enough kappa score, we will use one of the annotations as an ‘answer key’ to compare against other annotations of the same sentence. In this way we will calculate the precision, recall and f-measure of the “better” annotated dialogues against the other annotation (Rosenberg & Binkowski 2004). The threshold for a good kappa score is a hyper-parameter of our evaluation system which will need to be looked into and tweaked to get the most accurate results.

As an example of our annotation methods, below is a short annotation of a conversation: BEG\_SES\_A

BEG\_MES

BEG\_SENT

EMOJI\_waving\_hand CONVENTIONAL\_OPENING positive

END\_SENT

END\_MES

END\_SES\_A

BEG\_SES\_B

BEG\_MES

BEG\_SENT

hey CONVENTIONAL\_OPENING neutral

END\_SENT

END\_MES

END\_SES\_B

BEG\_SES\_A

BEG\_MES

BEG\_SENT

hey CONVENTIONAL\_OPENING neutral

END\_SENT

END\_MES

END\_SES\_A

BEG\_SES\_B

BEG\_MES

BEG\_SENT

what’s up CONVENTIONAL\_OPENING positive

END\_SENT

END\_MES

END\_SES\_B

BEG\_SES\_A

BEG\_MES

BEG\_SENT

i will delete my account :) STATEMENT negative

END\_SENT

END\_MES

BEG\_MES

BEG\_SENT

take care :) CONVENTIONAL\_CLOSING positive

END\_SENT

END\_MES

END\_SES\_A

retrieved from: https://iwastesomuchtime.com/PrimaryPetDevote

We have found several relevant research papers that discuss similar annotation projects as mentioned throughout the proposal. We believe it to be important not to just determine with what words an agent is responding, but what those words mean, and how they are relevant to the dialogue at hand. This is why we choose to tag utterances with the Stolcke et al. (2000) tags as standard that is used within the NLP community. It will be our continued investigation to learn more about relevant problems and solutions to annotation tasks in this domain.

In order to complete this project, we will be using a Github repository to collaborate with each other. We will also have meetings twice a week, or more as needed, to work on the project simultaneously. During these meetings, we can discuss our general plans and thoughts for the progress of the project, as well as specific details about annotation guidelines, retrieval methods, and evaluation metrics. We all plan on contributing to each part of the project equally and as the project develops, we will take lead in certain categories as we see fit.

References

Ivanovic, Edward. 2005. *Dialogue Act Tagging for Instant Messaging Chat Sessions*. Association

for Computational Linguistics.

Napoles, Courtney, Joel Tetreault, Enrica Rosato, Brian Provenzale & Aasish Pappu. 2017.

*Finding Good Conversations Online: The Yahoo News Annotated Comments Corpus*. Proceedings of the 11th Linguistic Annotation Workshop, pages 13–23. Association for Computational Linguistics.

Olshefski, Emily Grace. 2015. *Game-Changing Event Definition and Detection in an eSports*

*Corpus*. Proceedings of the 3rd Workshop on EVENTS at the NAACL-HLT 2015, pages

77–81. Association for Computational Linguistics.

Rosenberg, Andrew and Ed Binkowski. 2004. *Augmenting the kappa statistic to determine*

*Interannotator reliability for multiply labeled data points*. Association for Computational Linguistics.

Stolcke, Andreas, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel

Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema & Marie Meteer. 2000. *Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech*. Association for Computational Linguistics.