Masters Thesis

Part 3 of the document.

Post Meeting 4 Notes

Tasks:

- Find open subtitles, SRT files (we should have some, check Leverage and other Movies and Shows in our library)
- Run similarity algorithm on sample traces of labeled data
- Consider automated process to expedite our labeling

Next meeting is 19 Oct (Mon) @ 13:00

Labeling Process Notes

We took the SRT files from the Graham Norton Show, which is a sort of late-night host talk show where celebrity guests are invited. It resembles casual conversation with a direction so it's actually very good to use. However, the challenge is that multiple people are always talking, and the subtitles also contain audience reaction, though the guests do interact with audience members as a third collective.
We immediately notice that our labels are a little vague again, but this is to be expected. Here we should note how we're condensing labels, though we should honestly add new ones to be more accurate.
Additionally, in the talkshow, there are places where the closed questions immediately are answered and grouped with the question, so we should be careful to just list those as respond.agree, unless there is a separate question before, then to which we leave it as a question tag… but we should maybe also think about a better solution for this.

Labels that we may need:

- misc - to indicate applause and other things like actions/gestures, instead of using use.social.convention or x.
- give.statement - instead of using give.opinion, since there's a big differential here between a value neutral statement rather than a statement that has an opinion.

- recall - we should have something to indicate that a speech contains quoted speech, which is unique. Or maybe put it under a category of "give.recall"
- relax.atmosphere - we can expand this definition to include statements that are meant as a joke/banter (since the goal is to impact the mood, not necessarily involved with using social conventions) and maybe expand to just exclamations too?

We may also need to consider sub-conversations and how we can make them generic enough to catch multiple speakers in a conversation, i.e: catch trends without specific labeling on who the speaker is. This may be difficult, and perhaps more inaccurate, but something we should think about.

We definitely have to alter the labels and create a more accurate labeling process. Expanding the genre makes it more apparent. Perhaps we need to think about creating a good set of labels that can fit most conversation in different mediums, and from labeling and process determination, we can discover some sub-conversations, and those can help us determine the archetype of conversation (interview show, extended documentary, language exchange, etc) it becomes more apparent.

Post Meeting 5 Notes:

Tasks:

- ~~Finish sub cost~~
- ~~Finish indel cost~~
- Finish clustering
- Have examples confirming the validation of the functions (breaking each step into a separate part might help with this)

Next meeting is Monday, 26 Oct

Post Meeting 6 Notes:

Tasks:

- More formal explanation of the functions
- Move beyond the paper

Next Meeting is Monday, 2 Nov

We've fixed some of the issues with the substitution and indel scores, and they should be for the most part correct (maybe with exception to certain edge cases). Additionally, we've also been able to replicate the table-algorithm for calculating similarity (it's the same algorithm for edit distance, just the opposite). Additionally, we have to account for the fact that traces can have different lengths, so we enforce that the shorter trace be the base comparison.
In the paper, they suggest that insertion is the complement of deletion, but the implementation for our case gets affected when we implement this because of the many cases where there are cases of labels having a severely low insertion score, causing deletion to increase the similarity score too much. For our current purposes, we have listed deletion as a unit 1 similarity cost.

Now, we're moving onwards first to see how we can further add on to the concept of similarity by adding a notion of "sub-conversations." Which are larger trends in the conversation trace than just co-occurrence and context, which seem like intuitively should be tighter and shorter distance similarity measures, whereas a larger presence of similar "archetypes" of conversation should still contribute to similarity. (i.e: for us, our graham norton show traces shouldn't be that similar to the blackpink documentary, but in a sense, they are still interview based traces, and there should be some attention given to that).

So, for our current situation, our similarity metric is displaying that trace1 and trace2 (the two Graham Norton traces) are similar (positive score) whereas trace1 and trace3 (the blackpink documentary) are dissimilar (negative score). However, as we mentioned, there should be some notion where these two are similar, so perhaps we can look to define other metrics to display similarity. (i.e: WHERE are these two traces similar? For us, it would be the interview format, as well as the

idea that speakers display a high amount of recall). We can start building by first defining these trends in the abstract?

Let's maybe start with brainstorming labels that we would choose to define what we intuitively would call a sub-conversation?

> Monologue - maybe something like a long string of give.statements, with some give.opinions inside? (marked by a respond.agree/respond.deny, since a monologue is broken up by the interjection of another person, in this case of speakers not being labeled, we can assume that a respond label is essentially interjection. Things like questions can still be rhetorical)
>
> Making a deeper connection - this would be something along the lines of recall until question or recall until give.statement/opinion or recall until respond.agree/deny for a chain sequence of recalls. The idea is that one person asks a question that is followed by a longer sequence of recalling, which usually indicates sharing a memory, which in polite conversation is how people tend to form a deeper social connection.

Then we can move onto considering how we would discover this longer trend of labels, maybe something like LTL? Or a variation of LTL to check a longer sequence?
Additionally, there is the issue of determining the speaker. Is there a need for us to determine the speaker currently? Because if we separate speakers as labels, then we'll only drive the alignment for similarity of two traces further apart, but if we don't it might be harder to apply notions of longer sequences in an intuitive way? I guess we can say statistically, long sequences of give.statements/give.opinions can be considered a monologue because it's unlikely two speakers interchange give.statements or give.opinions without some sort of interjection. Additionally, the same goes with long sequences of recalls. I guess you could call it human nature. Can we think about something like, recall until question or recall until give.statement/give.opinion/respond.agree/respond.deny. This very linear notion

would fit the bill for something like a question about someone's background, which is a likely candidate to pop up in LE and conversation as a format.

I feel like even looking for sequences for just these two to begin with would already be a difficult task. We can probably just start from there. Finding a long sequence using LTL or some variation of logic is algorithmic, isn't it?