The Role of Conversational Structure in Dialogue Act Prediction

Charles Threlkeld[1], Karin Knudson[2], & JP de Ruiter[3]

[1]Department of Computer Science & Cognitive Science
[2]Data Intensive Studies Center
[3]Departments of Computer Science & Psychology

ABSTRACT

Underlying the words we use in speech are the dialogue acts they perform—the social implications conversation. For example, "I am married" Is a statement about my marital status. On the other hand, "you are married" could either be declaring two people married or it could be asking about the marital status of someone. The difference is in the dialogue act of the utterance. Correctly construing the dialogue act is a difficult task because it does not strictly follow from the words or the linguistics, but from the contextual information surrounding the utterance. In this work, we are quantifying aspects of the conversational context that contribute to dialogue act construal. We examine the conversational context, like the function of the previous and next utterance, the timing context, like the silence duration before or after a turn, the speaker context of whether an utterance is the first or last of its turn, and the sentence type of an utterance to find which aspects of conversational organization are useful in dialogue act detection. We are building from linguistic and sociological theory to create extensible and explainable models of dialogue act prediction and classification.

INTRODUCTION

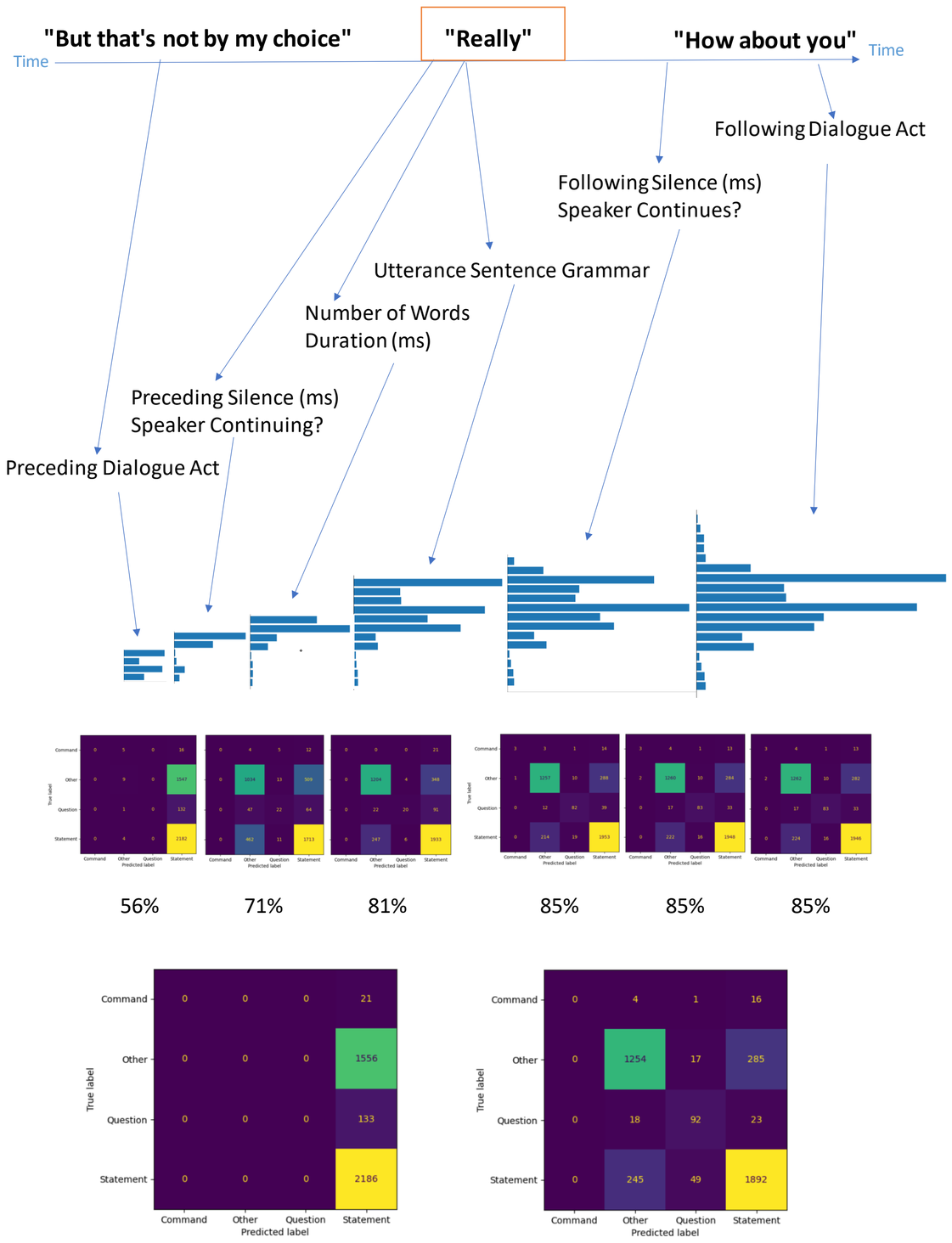
Much of the previous work in dialogue act detection relies on opaque, black box neural networks. They perform very well, but are computationally expensive, do not operate incrementally, and do not allow explanation or extension. In contrast, our work roots itself in linguistic theory, allowing for disciplined extension. Previous work (see below, current authors in press) sets standards for direct and indirect speech acts based on grammatical sentence-type and the Literal Force Hypothesis-inspired schema of dialogue acts.

In the current work, we are using a large corpus of randomly paired people having a conversation. We are using a rich but small set of annotations to try to both predict the upcoming dialogue act and label the dialogue act after it has occurred.

This work is crucial for Spoken Dialogue Systems, that need to recognize the social actions to which they are responding, and for shaping their own utterances. For example, one study found that participants order food extremely rarely with the "direct" command, and any agent must recognize and "indirect" request such as "I'd like some water" (Williams et al. 2018).

METHODOLOGY AND PRELIMINARY RESULTS

We are doing a simple parse of the data for only a few features, making our model more explainable, easier to modularize, and easier to compute. We are working in discrete timesteps, each of which adds information to the overall model. Our naïve baseline assumes no knowledge of the turn and is the metric for predicting dialogue acts. The grammatical baseline shows how well a simple mapping for sentence type to dialogue act performs and is the metric for post-hoc classification.

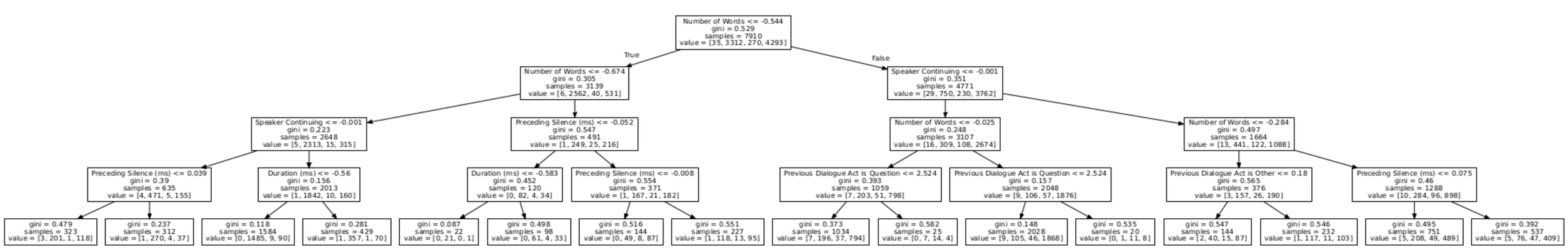


Naïve Baseline: 55%

Grammatical Baseline: 82%

Above, we show factor importance comparisons for each timestep of conversation given by random forest models. Next, we have heatmap confusion matrices for logistic regressions and comparisons for baselines which guess the dialogue act only by most frequent (left) or sentence type (right). Sentence type is only available after the utterance is given, so it is only comparable to the rightmost three model heatmaps.

One goal of this research is to create a dialogue act prediction model. When implemented in a spoken dialogue system, a dialogue act prediction model will allow turn planning much earlier in the turn, leading to more fluid, natural conversational style. Below is a decision tree based on the predictive model shown (third from the left above). We are planning to implement this model in a spoken dialogue system in the next few months.



NEXT STEPS

1. Correcting for an imbalanced training set
Our current models are having trouble because the number of Statement and Other dialogue acts greatly outweighs the number of Questions and Commands. We are looking into methods of resampling which can help our models overcome the sparse data we have.

3. Unsupervised Learning Of Dialogue Acts
We plan on performing some clustering withing the Other and Statement categories to find systematic differentiation patterns. We expect to replicate findings like the Pregnant Pause where no-answers are slower than yes-answers (Bögels et al. 2015).

2. A More Fine-Grained Dialogue Act Schema
A key point of our model is to be extensible. The small dialogue act schema is motivated by linguistic theory and previous work but can be extended. We have collapsed tags from the original corpus that we could separate into more refined categories. We anticipate that the unsupervised learning approach above will converge with the aspects of the extended schema to give us an explainable, empirically validated dialogue act set.

CONCLUSION

Our research so far has several findings: 1) features of sentence structure and conversational organization predict dialogue acts and, 2) conversations are dominated by statements and dialogue acts that are not sentences, and 3) despite the dialogue act being a construal of the words spoken, the context following an utterance is not helpful to classification. Importantly, none of these results rely on the words within the utterances, and so can work with and extend previous work that does rely on semantic structure.

Our utterances are predictable. We can see that our best models are correct 85% of the time, but the models that have no syntactic or semantic information are still correct 81% of the time. So, without knowing any of the content of a turn, we are still much better than chance at determining the dialogue act of an utterance.

Conversations are dominated by Statements and Other dialogue acts. Much work in dialogue systems and linguistic theory differentiates interrogative or imperative sentences into sub-types, but there is less work on subsets of declarative sentences and a finely-grained derivative dialogue act system. The original Switchboard corpus tags are of some help, but still dominated by Statement Non-Opinion and Statement Opinion. For a conversational agent, we will need to do more work to figure out how these two Dialogue Acts fit together and how they can be shaped internally to differentiate them.

Finally, it was a surprising find that the following context does not add meaningfully to the dialogue act recognition. Previous work by the authors has shown that the following silence duration is sensitive to the dialogue act, but not the sentence type. Our models were not able to operationalize this previous finding. Further, theory about dialogue acts says that they are construals of sentences, and the way the utterance is treated determines its dialogue act. Our models show that the dialogue act is predictable to the extent that the construal evidence does not alter the original classification.

Data Science and Artificial Intelligence – Challenges in Convergent Research

DISC/TIAI Symposium
April 28–29th 2022
Joyce Cumming Center

ACKNOWLEDGEMENTS

This research is made possible by a SEED grant from DISC, as well as the collaboration of DISC associates.

Tufts University | Data Intensive Studies Center