Text Prediction in Conversational Dialogue Processing

Zachary Munro, Felix Gervits, Bradley Oosterveld, Matthias Scheutz

Human Robot Interaction Laboratory, Tufts University, Medford

Overview

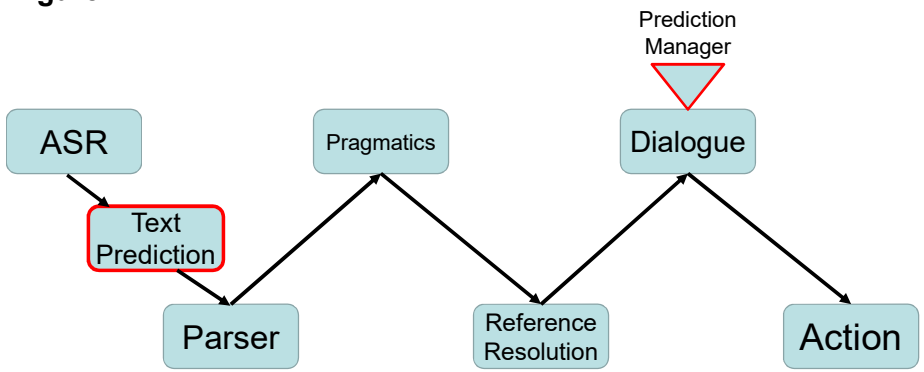
Research Goal:

Enable our Natural Language Processing pipeline to incrementally form predictions based on input and respond precisely after an utterance.

The DIARC Natural Language Processing pipeline that is developed by the Tufts Human Robot Interaction Laboratory is a distributed integrated affect, reflection, cognition architecture. It is structured as shown in figure 1.

Our plan was to add a new component to the pipeline that would predict the full text of someone's utterance before they had said it, allowing a robot to preemptively prepare a response, and respond either normally or by interrupting.

Figure 1



Method

Use a bigram prediction model to construct an utterance based on probabilities from a corpus of data containing utterances regarding household tasks. Bias model weights using context from surroundings such as proximity.

Algorithm

1. **Get first word** of utterance
2. **Find** most likely **second word**
3. **Repeat** until an entire "coherent" utterance is created
4. **Send utterance** through NLP pipeline to the Prediction Manager
5. **Prepare** a response/action
6. **Weigh** the likelihood of our prediction with the cost of the action we'd take
7. **Act/Respond** timely
8. **Use new state** of the world to **bias the weights** in our predictive model

Pipeline

A spoken utterance travel through the pipeline as follows:

- **ASR:** speech is converted to text
- **Text Prediction:** partial utterances are built up into complete phrases using bigram prediction algorithm.
- **Parser:** phrases converted to semantics
- **Pragmatics:** Semantics are converted to recognizable sentence types
- **Reference Resolution:** references to objects are resolved to known IDs
- **Dialogue:** Natural Language Understanding and Generation take place as well as turn taking management
- **Action:** Action scripts are run so that robot can respond accordingly to utterance

Future Testing and Research

Testing:

- Against a non-incremental system to see if there is a performance improvement
- See if we can hit different types of overlap if wanted
- See how people respond to being interrupted by a robot in different contexts

Use cases:

- Speech therapy robots
- Military robots serving as assistants in complex tasks
- Robots aiding in tasks for people with disabilities
- Improving speed of dialogue processing in personal AI assistants on phones or household devices

References

1. Stivers, T., et al. "Universals and Cultural Variation in Turn-Taking in Conversation." *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, 2009, pp. 10587–10592., doi:10.1073/pnas.090361**6106.**
2. Gervits, M., Felix and Scheutz. (2018). Pardon the Interruption: Managing Turn-Taking through Overlap Resolution in Embodied Artificial Agents. In Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue (pp. 99–109). Melbourne, Australia: Association for ComputationalLinguistics