ACL Paper Summary

The paper discussed in this essay is “Improving Dialog Systems for Negotiation with Personality Modeling” by Runzhe Yang from Princeton University, Jingxiao Chen from Shanghai Jiao Tong University and Karthik Narasimhan from Princeton University as well. There are already many dialog systems for negotiation out there in the world, but the authors of this paper took it one step further and created a new framework for negotiation dialog. Recently, there are neural models that are able to show that useful dialog strategies can be learned but the current models are unable to model the mental state of the other agents involved in the negotiation which makes it difficult for these strategies to be specifically tailored for multiple different types of opponents. Yang and their colleagues propose a new framework for coming up with dialog strategies by taking inspiration from the cognitive science concept of Theory of Mind. Theory of Mind is the ability humans have to be able to understand how others think and to be able to predict how they will act. When humans negotiate with others, they innately predict how their opponent will react based off how they speak using Theory of Mind, so Yang and colleagues wanted to infuse that thinking into a dialog framework since there wasn’t one that did that already.

In this paper the authors also discuss prior related work. They talk about the Rational Speech Acts (RSA) models and speaker-follower models. The recent RSA models assume the optimality of the listener in the other’s mind and considers speakers and listeners in constrained settings while what Yang and colleagues propose doesn’t. Another similar model is the ToMnet model. This model uses a multi-agent reinforcement learning environment to deduce the identity from past events and predict the mental state of the agent from the current state. There are also dialog systems that already include some Theory of Mind in them. For example, Theory of Mind has been used in modeling personality types and predicting responses for building user simulators to train reinforcement learning dialog systems. Most of these have only been used in domain specific areas such as navigation settings.

This paper has many new and unique contributions to it. First of all, the authors’ proposed models for negotiation take utterances into account which helps to infer the type of opponent by capturing things like emotion which can’t be recorded in the standard dialog state transitions. Next, the proposed model is also capable of working with both cooperative and competitive opponents where previous Theory of Mind dialog systems only worked with cooperative agents. Another unique contribution from this paper is that the model has real-time inference and can change its dependence on the Theory of Mind predictions based on a parameter. These give the model more flexibility and the means to be more accurate.

To test their model, Yang and their colleagues created a task where there are two agents, a buyer and a seller. The buyer wants to buy the item at the lowest price they can while the seller wants to sell the item at the highest price. The two agents take turns in negotiating until they reach an agreement or disagreement. The actual negotiation process is formulated as a multi-agent Markov Decision Process and includes a parser that converts the utterances to dialog act, a manager that decides the responding dialog act, and a generator that take the dialog act and produces a natural language response from it. The Theory of Mind model is specifically a first-order model. That’s where the goal is to predict how what we say will affect our opponent and the authors make two types, implicit or explicit personality modeling. The first-order Theory of Mind with explicit personality modeling predicts the opponent’s personality where the implicit model doesn’t. The implicit model directly learns from the opponent’s dialog without the extra supervision needed to make a prediction of the personality. These models were then tested with cooperative, competitive, and some in between opponents in a negotiation setting to see how they would do. The evaluation metrics used were the agreement rate, which is the percentage of the dialogs that could reach an agreement, the objective utility, which is basically how far each agent was from their goal price, the deal fairness, and the dialog length. The results of these tests showed that the authors Theory of Mind model consistently had the highest agreement rates against all the opponent agents by about 20% when compared to the baselines. It was even found that against the competitive opponents, the authors model was able to achieve more than 30% increase in deal utility. They found that the explicit Theory of Mind model performed better than the implicit model because it has more information about the opponent type during the training.

Runzhe Yang is cited 282 times, Jingxaio Chen is cited 8 times, and Karthik Narasimhan is cited 7569 times. The author’s work is important because it described an innovative way to approach dialog systems by adding in techniques related to psychology and how humans behave. This approach of taking inspiration from the idea of Theory of Mind shows that there are ways to implement how our brains work into technology and see that there can be benefits and advances from it.

Works Cited

* Runzhe Yang, Jingxiao Chen, and Karthik Narasimhan. 2021. [Improving Dialog Systems for Negotiation with Personality Modeling](https://aclanthology.org/2021.acl-long.56). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 681–693, Online. Association for Computational Linguistics.