Summary on Selected ACL Paper

The paper that I selected was “MISC: A Mixed Strategy-Aware Model Integrating COMET for Emotional Support Conversation” by the following authors in order: Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan (although its worth noting that Quan Tu and Yanran Li equally contributed according to the footnotes). This paper is affiliated with the Gaoling School of Artificial Intelligence, Renmin University of China; Xiaomi AI Lab; and the Beijing Academy of Artificial Intelligence.

The problem that the paper addresses is the inadequacy of current AI models to provide meaningful responses to those that need emotional support. They claim that current methods are too course-grained, only understanding the seeker’s emotions on a static, conversational level rather than really understanding the emotions on a fine-grained level. They also claim that current methods, because they only understand the seeker’s emotions on a shallow level, provide responses that are more empathetic rather than trying to help de-escalate the seeker’s distress and provide them with meaningful help.

Prior work addressed in this paper includes a categorization of the three types of emotion-aware dialogue systems: emotional chatting, empathetic responding, and emotional support conversation. Earlier works only addressed emotional chatting, which could determine the emotional signals of the seeker and could use those signals to provide a response. Later works were able to get more and more specific with the seeker’s emotions, to the point where they were able to get empathetic responses based on a better understanding of what the seeker was going through. Recent works utilized ConceptNet to get even more specific with emotions and relevant responses. The paper’s solution utilizes a generative commonsense model for their solution.

Since their solution strongly leans on a commonsense model, it’s relevant to also discuss prior work in using commonsense models. It’s stated that commonsense models have been recently used to solve a myriad of NLP problems but have been used on multiple occasions for dialogue systems. Most of these dialogue systems utilize ConceptNet so that physical knowledge can be used in responses. However, the paper instead decides to focus on ATOMIC, which is more relevant to their use-case since it deals more with emotional knowledge and event-centered causes. The paper’s unique solution utilizes COMET, a commonsense model trained on ATOMIC, for a better understanding of the seeker’s mental states. It also allows them to utilize strategy prediction. This can predict a strategy or strategies the chatbot should utilize in its response, like asking the user a question, divulging into personal experience, providing a suggestion, etc.

Finally, the paper discusses prior work in different conversation strategies. Strategies are often described in terms of dialog acts, which are utterances that serve some sort of purpose in the conversation (like questions or calls to action). Dialog acts have proven useful in dialogue chatbots in prior works. For empathetic conversations, the notion of different

strategies have been used to describe how a person responds when they want to be empathetic to the seeker. This paper directly adopts from Liu et al. (2021) the idea of 8 different strategies that are used by a person in conversation to help reduce someone's emotional distress. The idea of 8 different strategies validates the team's decision to use a combination of these strategies in a single response to construct a meaningful dialogue.

The paper's unique solution to the problem is to utilize COMET in order to create the model MISC, which first determines the seeker's fine grained emotions and then generates a mixed-strategy response based on the 8 strategies defined in Liu et al (2021). MISC first uses a mental state-enhanced encoder, which first gets the context by grabbing the entire history of the conversation, then uses a brief description of the seeker's general situation *s* as an event and feeds it with different relations from COMET (relations are the connections between the utterance and the seeker's or other's reactions, for instance a speaker's utterance can reveal that they are sad or what their wants are, as well as reveal what the other person should want or feel about the seeker's response). Next, the distribution probability for using each strategy is calculated. The paper decides to use VQ-VAE's codebook (from Oord et al., 2017) for strategy representation so that longer responses can be generated and so MISC can learn based on how each strategy is weighed. Finally, a multi-factor-aware decoder is used to utilize the knowledge from the previous two steps to generate a response using multiple factors.

The authors evaluated their model by running MISC alongside different models with the dataset ESConv. 5 different models were run alongside MISC, including a state of the art (SOTA) model BlenderBot-Joint which generates a strategy token before generating a response and was also trained on the ESConv dataset. To assess the performance of each model, both automatic metrics and human annotators were used. The results showed that the Transformer model performed the worst as it had no way to learn empathy and had trouble with longer contexts. The models MT Transformer, MoEL, and MIME also didn't perform well because they utilized static emotional labels for empathy and lacked the ability of consoling speakers. For the SOTA BlenderBot-Joint, its performance was also not as good as MISC as it utilizes only a single strategy for their response, rather than MISC's use of mixed strategies. The authors further evaluated their model by taking away one of the following added parts: the true strategy label *g*, the situation of the seeker *s*, and the seeker's last response *x*. They did this to determine if those added parts helped the model. By comparing the performances, they noted that each of those added parts helped the model to perform better. Next, they evaluated the responses of each model. Upon evaluation, their MISC model generated a response that was most consistent with the context and was the most empathetic compared to the others. Next, they evaluated MISC by creating a variant MISE which uses a course-grained label. MISC is shown to perform better than MISE. They also analyze the mixed-response strategy by comparing it with a variant of their model called Single (which is just MISC but replacing mixed strategy with single strategy), as well as comparing MISC with the SOTA single strategy model BlenderBot-Joint. They concluded that their mixed-strategy model MISC performed better.

The following are the number of citations each author has received (not including Yanran Li, who did not have a google scholar page): Quan Tu – 20 citations, Cui Jianwei – 189 citations, Bin Wang – 7535 citations, Ji-Rong Wen – 21017 citations (most citations of the group), Rui Yan – 9545 citations. Note that this particular paper has received 9 citations thus far.

I think that this work was important because there are a lot of people who struggle with trying events in their lives and may have nobody that they can talk to about it. Having a model that can not only empathize but also alleviate their stresses can help them process their feelings and aid them in moving forward. This model is also crucial since it could generate responses that sound more realistic than previous models. However, I would like to see how this model performs with actual people, as that seemed to be one area of experimentation that was lacking a bit.