In a paper written for the 2021 ACL conference titled *I like fish, especially dolphins: Addressing Contradictions in Dialogue Modeling*, authors Yixin Nie, Mary Williamson, Mohit Bansal, Douwe Kiela, and Jason Weston discuss the topic of conversational inconsistencies within the scope of natural language processing as it pertains to open-domain chatbots. Their paper outlines the issues associated with dialog inconsistencies, previous attempts to address those issues, and how their new method, the DialoquE COntradiction DEtection task (DECODE), handles the problem more effectively.

Access to great quantities of dialog data has caused a “resurgent interest” in creating intelligent, open-domain chatbots[1]. One of the most notorious and frustrating errors that these chatbots produce is the error of contextual inconsistency [1]. Chatbots are often held to the same standards as human to human interaction [1]. Inconsistencies in human to human dialog can make it feel like a person isn’t paying attention, and it disrupts the flow of the conversation. It is the same with chatbots. Contradictions in conversation with a chatbot can be “jarring”, ruining the illusion [1]. They garner mistrust in a user and can disrupt long term confidence in the conversation [1]. Issues arise not only with the user of the chatbot, but they also arise on the development side. Many techniques have been employed to address consistency issues, but measurement of the efficacy of these techniques has proven difficult because things like the accuracy of conversational logic are hard to track [1]. Despite the multitude of attempts, the issue of contradiction hasn’t been resolved, nor has a regularized method been created to gauge any potential consistency improvements [1]. The continued inconsistencies in chatbot dialog have raised concerns about the ability of generative models to ever fully understand human speech [1].

Prior to the writing of *I like fish,* there were many works focused on improving the chatbot dialogue consistency problem. Some utilized implicit models where consistency related information was stored in the form of distributed embeddings, neural long-term memories, hierarchical neural architecture, latent variables, topical attention, or feature vectors [1]. Others have used text generative models based on user personas [1]. A previously used method that is most similar to the authors’ DECODE method is a model called natural language inference (NLI) [1]. This type of model uses unlikelihood training

to try and reduce inconsistencies in generative approaches [1]. While NLI is similar to DECODE, it fails miserably in some domains and has no way to measure contradiction metrics [1].

The task defined in *I like fish,* DECODE, handles the two main problems associated with contradiction reduction in chatbots. The task uses a supervised model where the input is a list of “utterances” that represent a portion of chat dialog [1]. The model is similar to NLI sequence-to-label modeling except in place of a three part output, DECODE uses cross-entropy loss to create a boolean output indicating whether or not the last utterance contradicts any utterance contained within the dialog [1]. DECODE explored two approaches to contradiction detection using the author’s modifications. The first was a simple unstructured approach where all utterances in a dialog were strung together to form a single context that was fed into the model to determine the probability of contradiction [1]. The second approach was a structured approach that separated the dialog into sets of utterances by each speaker that was then separately fed into the model [1]. Testing was evaluated using the two part labeling offered by DECODE, and the results indicated that both approaches of DECODE performed better than other NLI tasks at supervising contradiction in dialog with approach one providing the best results on “in-domain” test sets[1].

What makes DECODE special is the use of humans to not only create data but also to identify inconsistencies [1]. The input of contradictory utterances were created by asking annotators to add additional utterances to existing dialog that were inconsistent with the given conversation. People directly contributed to the creation and labeling of the dataset, and there is no better way to determine contextual inconsistencies, and the results of using DECODE are evident of that.

The paper *I like fish* was written by people affiliated with Facebook AI Research and UNC Chapel Hill, and according to Google Scholar, has been cited 33 times in the past two years. Of the authors, Jason Weston has been cited the most at 99,276 times and is a part of the Facebook AI Research team. The work done using DECODE is important because it seems to be the first of its kind to use humans to solve human problems. Everything can seemingly be automated, but some problems are uniquely human and can only be solved by people. One of these problems is determining

whether the context of a conversation is inconsistent or not. DECODE introduced human created datasets into generative models, and the results proved that sometimes people can solve problems better than computers.

Resources

[1] Y. Nie, M. Williamson, M. Bansal, D. Kiela, and J. Weston, “I like fish, especially dolphins: Addressing contradictions in dialogue modeling,” *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)*, 2021.