

In most instances whenever a person is making use of a chatbot they do so knowing exactly what they wish to do. Whether it be making a call using Siri or using the chatbot on a bank's website to learn about their interest rates, most of the interactions between the user and the chatbot follow the pattern of "User gives order, chatbot complies." This however, is not effective in reality. Oftentimes when people decide to do things, they do so without knowing precisely what and how exactly they plan to do it (As a comparison, compare "I am going to build a computer." with "I am going to build a computer with an Intel Core I9, 32GB of RAM, and an NVIDIA RTX 3090 that I have purchased from Newegg."). The creation of a corpus (called "DuClarifyDial") that can handle these situations effectively is the topic of the paper *Where to Go for the Holidays: Towards Mixed-Type Dialogs for Clarification of User Goals* authored by Zeming Liu from the Harbin Institute of Technology in Harbin, China and Jun Xu, Zeyang Lei, Haifeng Wang, Zheng-Yu Niu, and Hua Wu from the Chinese technology company Baidu, Inc.

However, it is important to recognize the prior work that needed to be done in order to make DuClarifyDial a reality. These include multi-domain task-oriented dialog datasets, knowledge grounded dialog datasets, and multi-tasking dialogs. These are all fairly self explanatory. Multi-domain task-oriented dialog datasets, have been a long standing research topic in NLP and consists of the creation of very large scale datasets with the goal being that they will be able to be used to perform tasks in various different domains. However, these are hindered because they assume that the user knows the exact details of what they wish to do. Another piece of prior work is knowledge grounded dialog datasets, which provide digital assistants with enough information about their topics in order to answer specific questions about their domain and even potentially provide recommendations. Finally, multi-domain task-oriented

dialog datasets aim to bring the capability for the chatbot to work on multiple tasks simultaneously. However, DuClarifyDial is different in that alongside providing information to and performing tasks for the user it also is able to help the user in clarifying what exactly their goals are if the user is unsure.

In order to create the dataset and have it have the information it needs to operate as described, special data collection must take place. This data collection process can be broken down into four steps. The first is collecting information from the web to create a knowledge base for the corpus' five domains (hotels, attractions, restaurants, food, and movies) to allow it to provide detailed information to the user. Second, a series of templates of dialogues are created for the crowdsource workers to work off of. These templates are designed as such so that a single conversation can include discussion over multiple domains and can be broken up into different "sub scenarios" with different dialogue types (these types being general chit chat, decision making assisting, task oriented, and knowledge grounded). Using these templates, crowdsource workers are able to generate over 5000 dialogue sessions that move from type to type and sometimes include separate domains as well. Finally, a unified dialogue schema that blends these different dialogue types together is created. The dialogue sessions are then annotated in accordance with the schema.

Work evaluation took several forms, first the dataset's dialogue sessions themselves were manually judged. Three workers judged 200 random dialogue sessions in the dataset, rating them a 1 if they were fluent and followed the schema, and a 0 if they did not. The dataset was then used with multiple language models, including a pre trained mixed type dialog model known as PLATO-MT. Each model was judged based on their type accuracy, domain accuracy, slot accuracy, and joint accuracy. Also humans evaluated 50 sessions with the models using the

dataset, scoring them on their appropriateness, informativeness, the accuracy of the information presented, and whether or not it was helpful in assisting the user in figuring out clear goals. The results of these evaluations, as well as automatic evaluations such as METEOR, BLEU 1-2, and CIDER, showed that PLATO-MT outperformed the other models by most metrics.

Between all the authors that have Google Scholar accounts (accounts for Zeming Liu, Jun Xu, and Zeyang Lei unfortunately could not be found), these authors have received 23,566 citations. The leader is Haifeng Wang with 11,871 citations. Closely behind him is Hua Wu with 10,079 citations. I believe the work done by the authors is incredibly important. Not only does it represent a new way to allow for chatbots to provide recommendations and help the user make decisions, it also showcases a new way for chatbots to deal with conversations that naturally flow to and from different types and domains.