TOWARDS HUMAN-LIKE DOMAIN-LIKE MULTIMODAL DIALOGUE AGENTS

Diogo Silva (dmgc.silva@campus.fct.unl.pt)
PhD Student
NOVA LINCs

**Motivation**

Task-oriented agents are increasingly common in several aspects of our daily lives. However, they often lack the naturalness and deep domain knowledge of human experts. These problems hinder their adoption, as it can be frustrating to users who might have questions that the agent is not able to answer accurately. This work focuses on bridging the human-machine gap by producing conversational agents capable of dialoguing with the user in a natural, informative, and factual manner, about recipes and DIY tasks.

**Methods and Techniques**

To produce a Large Language Model (LLM) capable of dialoguing about recipes we needed to create a conversational dataset and then train a system on it.

**Data Creation**

To produce high-quality data, we leveraged user-system conversations about recipes. With these conversations we mapped the user behaviors when transitioning between different conversational intents. To determine the intents, we used an intent detector trained on conversational cooking data.

**Agent Training**

We trained a Vicuna-7B, a transformer-based LLM, using a Fully Sharded Data Parallel approach. During the training, the model was given the recipe information, current recipe step, previous 4 conversation turns, current user request, and target tone of voice.

**Results**

To test the proposed models, we performed automatic evaluation, using GPT4, on several type of user requests:

- Plan Navigation
- Grounded Question Answering
- Open Requests
- Conversational Norms

Additionally, we performed a user study, with Vicuna-DPO, on both recipes and DIY tasks. We found that the trained model showed significant dexterity in an unseen domain (DIY tasks) even in Question Answering requests.

**Future Work**

The ongoing and future work is focusing on bringing multimodal capabilities:

- Visual QA
- Video Moment Retrieval
- Video QA

**Scientific Output:**


**Awards:**

2nd Place in the 1st Edition of the Alexa Prize Taskbot Challenge, 2022
1st Place in the 2nd Edition of the Alexa Prize Taskbot Challenge, 2023

**Acknowledgements:**

This work has been partially funded by the #4ESt project, Ref. 4ESt20, co-financed by ERDF, COMPETE 2020, NORD2020 and FCT under UIDB/04516/2020, and by the FCT project NOVA LINCs Ref. (FCT/07017/2022).