Stockholm, 2024-01-15

Review of PhD thesis "Low resource methods for dialogue systems applications" by Vojtěch Hudeček

The doctoral thesis by Vojtěch Hudeček concerns the area of dialogue systems, which are systems that users can interact with through natural language. Being able to communicate with machines through conversation, in the same way we naturally communicate with each other, has been a long-standing vision in both science fiction and research labs, and it has been considered a hallmark of human intelligence. This research is particularly timely given the surge in public interest and technological advancement in large language models (LLMs) illustrated by OpenAI's ChatGPT. It is important to stress though that this is quite a large area of research, where different researchers focus on different application domains, or different modes of interaction (text, speech, multi-modal, etc.). The field is also inherently interdisciplinary, as it involves knowledge from computer science, machine learning, linguistics, etc.

The thesis under review here targets written interaction with task-oriented dialogue systems. Task-oriented dialogue refers to dialogue such as restaurant booking or information seeking. The focus is on how these systems can be developed without extensive need for annotated material to train on. The thesis has four main contributions. The first contribution concerns the processing and automatic annotation of raw dialogue data to make it suitable as material for supervised machine learning. This is done through a process of identifying suitable "slots" (corresponding to semantic concepts) in dialogue utterances. These identified slots are then used to train a domain-specific slot tagger. The evaluation is done over multiple different domains (using already existing corpora).

The second contribution is a hierarchical variational model for task-oriented dialogue. Through analysis of the latent space, it is shown that the hierarchical modelling is beneficial. The third contribution is an exploration of how multiple domains can be merged into one ontology for training sequence-to-sequence task-oriented dialogue systems over multiple domains.

The fourth main contribution is an exploration of how pre-trained LLMs (like ChatGPT) can be used in a zero-shot or few-shot fashion to model task-oriented dialogue. The model is used in several different ways to determine the topic, identify

KTH
Lindstedtsvägen 24 | 100 44 Stockholm
08 790 78 74 | skantze@kth.se | www.kth.se/profile/skantze

slots and generate a response. The evaluation shows that the use of pre-trained LLMs for task-orient dialogue systems is very promising.

While these contributions are substantial, I do miss more evaluations using real interactions with human subjects. The author does acknowledge in the background section that “Human evaluation plays a crucial role in developing and refining dialogue systems. Automated tools or metrics often fail to capture human communication’s nuanced responses and language variations.” However, I can only find one very limited human evaluation in the thesis, where the participants are actually interacting with the system. And in this case, there are only 6 people who are also experts.

Another concern is perhaps more related to the field in general, where the tasks and applications that are being studied here are perhaps not the most appropriate for dialogue systems in the first place. I would for example argue that it is questionable whether restaurant booking is best done through a natural language interface (rather than, say, just using Google Maps). While I cannot criticize this PhD student for not expanding the horizon beyond those standardized tasks, it would have been an even more exciting read if the thesis had contained such attempts.

In summary, the contributions of the thesis are valuable to the field of dialogue systems. While there are rapid advancements and large commercial efforts in the developments of LLMs and services like ChatGPT, there is still a need to make those systems better grounded to be able to have task-oriented interactions with them and avoid hallucinations. Thus, there is still a big need for this line of research. Seeing that Vojtěch Hudeček has been able to adapt his research agenda towards those new tools (in his final contribution) is a good sign, even if I miss more thorough human evaluations and considerations of the practical use of these systems. In conclusion, I can confirm that the thesis shows that Vojtěch Hudeček is able to formulate relevant research questions and carry out and report the research in a scientific manner, and he therefore deserves a PhD title.

Gabriel Skantze
Professor in Speech Technology
Division of Speech Music and Hearing
Department of Intelligent Systems
School of Electrical Engineering and Computer Science
KTH