

## **Review of doctoral dissertation**

### **Low resource methods for dialogue systems applications**

**Submitted by Vojtěch Hudeček to Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics**

The thesis focuses on the development and improvement of task-oriented dialog systems. The author proposes techniques that can reduce development and deployment costs, speed up development and deployment, and improve the quality of the dialogs in terms of the nature and targeted business goals. The research focuses on data by introducing a weakly supervised automatic data annotation pipeline, on models investigating latent variable models, pre-trained models, end-to-end training, and popular large language model-based approaches. The range of addressed topics is very wide and covers almost all aspects of task-oriented dialogs and the development of the research field over the past few years.

The thesis is a unified document with 159 pages divided into eight chapters.

This review first deals with the technical content of the thesis, then summarizes its technical quality, comments on the formal points, and finally presents overall conclusions and recommendations to the University.

### **Technical content of the thesis and remarks to chapters**

Chapter 1: Introduction – this chapter gives the introduction and provides some motivation for the research work. The selected topic is very relevant, not solved yet, and the motivation is well-grounded.

Chapter 2: Background – this chapter provides an overview of all techniques used in this thesis. It introduces neural network architectures, pre-trained models, variational autoencoders, memory networks, dialog system architectures, datasets, and evaluation metrics. The chapter is heavy in the number of techniques and content covered. I appreciate the author's deep understanding of the presented techniques, good summaries, and well-estimated balance between the detail of information shared and text length. It is necessary to go to the referred materials occasionally, but in most cases, the provided text is enough for good understanding. From the content, the only part I would appreciate and did not find is a short discussion comparing Memory Networks and Transformers. Both were used in this thesis, and transformer-based architectures are very popular today.

Chapter 3: Related Work – this relatively short chapter describes the modular dialog system architecture and end-to-end architecture. It also mentions instruction tuning for large language models. Then, it presents techniques for working with less or without supervision. These include automatic labeling for NLU and Variational Autoencoders.

Chapter 4: Discovering dialogue slots – this is one of the key chapters. This chapter evaluates several methods of getting slots from raw unlabeled data. It includes using a frame semantic parser, named entity recognition, slot filtering, training of standalone tagger on selected candidates, and use of large language models for candidate identifications. All the methods were implemented and evaluated on some dialog data sets. People try to extract slots using large language models often these days. But fight with hallucination. Better slot detection can bring better understanding to data in the dialog structure discovery task, more fluent dialog, and help introduce methods reducing hallucination in large language models.

Chapter 5: Dialog modeling with less supervision – this is the second key chapter, reducing the amount of supervision and making the dialog states more explainable through variational models. These are hot topics from the research point of view and also because of European values and legislation. Variational Recurrent Neural Networks or Hierarchical Variational Models are explored here. The work is well described, although, for some

technical details, it is necessary to go to the author's published articles or to other cited articles. I appreciate the visualization of latent variables through t-SNE for better imagination.

Chapter 6: Sequence-to-Sequence Task-Oriented Dialog Modelling – this chapter gives an overview of how the dialog systems can be seen as some sequence-to-sequence models and how these models can be decomposed into multiple tasks to facilitate the use of external APIs. It uses GPT-2 for both belief state prediction and response generation. Here, I understand that the author was a member of a wider team building this system, and the author focused mainly on data preparation. The Diaser corpus was created. It is publicly available on a GitHub repository, and the creation is well described. The main question here is how to preserve the usability of this corpus for future research. The public availability through web download is an advantage on one side. However, the web data are collected for other data sets. For example, it was verified during the JSALT 2023 workshop that MultiWOZ data is in the training data of the Llama 2 large language model. It is impossible to discover the facts from the model description. So, the results of future research may be biased due to data contamination.

Chapter 7: Large Language Models for Task-Oriented Dialog – this chapter connects the previous research with the recent hype around large language model (LLM) based chatbots. It nicely demonstrated how to design such a system through a series of LLM calls for domain detection, state update, and response generation. Prompts for zero-shot and few-shot LLM queries were used. The results nicely present the advantages and disadvantages of this architecture of dialog systems compared to the previous architectures. It shows that the LLMs do not overperform classical methods everywhere and have difficulties with some tasks. And they suffer from hallucinations. Here, for example, techniques from Chapter 4 can help mitigate the problems. This chapter uses LLMs as they are without any finetuning. Investigation of some finetuning methods might be an extension of the research work.

Chapter 8: Conclusion – concludes the work and indicates possible future research directions.

### **Summary of the technical content of the thesis**

The thesis clearly demonstrates the qualities of the candidate – the capability to study non-trivial literature from several fields, suggest own novel solutions, implement them, and carefully test and discuss the results of experiments. I highly appreciate the clarity of the presentation, quantity and quality of experiments done on several datasets, and thorough discussions. I also appreciate that the experiment's source code and data were made public for easy reproduction. The research answered important questions for developing and deploying practical systems and indicated several future research directions.

### **Comments on the formal aspects**

The thesis is well-structured and well-written. I appreciate the well-estimated level of presentation detail that is good enough for understanding and does not prolong the text. I would appreciate better referencing between text and figures in several places. Also, it would be nice to introduce the frequent TOD abbreviation better the first time and then use it consistently later. There are also several errors, mainly in the description of figures and referencing. But nothing critical. These will be shared with the author for possible correction before publication.

### **Summary and recommendation**

I have carefully examined the doctoral thesis of Vojtěch Hudeček. In my opinion, it is solid work that contributes to the progress in the field of task-oriented dialog systems. The addressed research topics are important and actual. I see the main parts and uniqueness in Chapter 4 "Discovering dialogue slots" and Chapter 5 "Dialog modeling with less supervision", and partly also in Chapter 6 "Sequence-to-Sequence Task-Oriented Dialog Modelling", in the corpora preparation work. This work brings significant new contributions. The research methods and approaches used are appropriate and trustworthy. There has been substantial progress in this research field in the recent few years. The thesis nicely maps this world's research progress and shows that the

author was always at the state-of-the-art edge. The work clearly showed prerequisites for independent work. There are several possible directions for the follow-up research work that I see. Some of them are indicated by the author in the conclusion. The author's strong devotion to research is also proved by his new involvement in a newly started European research project, Eloquence, which focuses on this research area.

**To conclude, I find the dissertation ready for publication and printing and recommend accepting the dissertation as a partial requirement for granting Vojtěch Hudeček the Doctoral degree at Charles University.**

In Brno, January 24<sup>th</sup>, 2024

Petr Schwarz, Assistant Professor  
Department of Computer Graphics and Multimedia  
Faculty of Information Technology, Brno University of Technology  
Bozetechova 2, 612 66 Brno, Czech Republic  
Tel: +420 733 532 891,