Léo Jacqmin

Orange & Aix-Marseille University
2 Av. Pierre Marzin
22300 Lannion
France

leo.jacqmin@orange.com
https://jacqle.github.io/

1 Research interests

My research interests lie in the area of task-oriented dialogue (TOD) understanding. I am interested in developing systems that can effectively discern meaningful information from a complex dialogue context. More specifically, my thesis focuses on dialogue state tracking (DST) and robustness to distribution shifts. I plan to explore several research avenues based on different possible variations in the input in the context of this task.

1.1 Spoken dialogues

My first research track is concerned with robustness to spoken inputs. Spoken dialogues pose additional challenges due to automatic speech recognition (ASR) errors and speech phenomena such as disfluencies and repetitions. Though previously widely studied, spoken dialogues have been neglected in favor of textual dialogues in past years in the context of DST (Faruqui and Hakkani-Tür, 2022). As a result, current DST models struggle when faced with ASR hypotheses (Kim et al., 2021). However, dealing with spoken dialogue is arguably one of the main goals of dialogue research.

As part of this effort, I am participating to one of DSTC11 tracks intended to stimulate work on this aspect by providing a spoken version of MultiWOZ (Budzianowski et al., 2018). I plan to explore and compare traditional cascade approaches which first transcribe speech to text before proceeding with DST, along with end-to-end approaches which directly extract semantic representations from speech. This latter approach can benefit from the recent progress on self-supervised learning of speech representations (Baevski et al., 2020).

1.2 Long dialogues

DST requires selecting relevant information from the dialogue context and this can become more difficult as the dialogue length increases. Originally intended as the main focus of my thesis, this aspect will be explored at a later stage. It has been shown that a large portion of turns in the main DST datasets can be considered as non-conversational, i.e. they can be parsed successfully without considering previous turns (Jakobovits et al., 2022). Moreover, joint goal accuracy (JGA) is a difficult metric to interpret in this context. Several studies have pointed out that JGA decreases as the dialogue length increases. However, this observation can be misleading as JGA requires a strict match and a single error made at an early turn can lead to invalid dialogue states for the rest of the dialogue. As a result, the decrease in JGA in long dialogues is mainly due to error accumulation rather than the difficulty of selecting information from a complex dialogue context. Hence, this aspect is left for further consideration in the hope that TOD datasets that are conversational are released later.

1.3 Real dialogues from industry

I am interested in applied research and ultimately, I would like to apply the methods explored in the two previous avenues to real dialogues from industry. While dialogues from standard TOD datasets have become increasingly complex, due to the type of data collection they follow rigid scenarios and do not entirely reflect real dialogues. Real dialogues tend to exhibit much more complex scenarios and rarely follow a happy path. As an industrial PhD candidate, I am fortunate to have access to real anonymized conversations between clients and customer service agents from a telecom company. These dialogues are usually lengthy and quite noisy, e.g. due to misspellings. As such they pose additional challenges, which I plan to address in the final part of my PhD.

2 Spoken dialogue system (SDS) research

Smartphones and smart home appliances have become ubiquitous and come with a strong interest to further research in the field of SDSs. Recent deep learning approaches to SDSs have enabled considerable advances and it is difficult to foretell where the field will be in 5 to 10 years. In this digital age, SDSs offer a convenient way of interacting with machines and they have the potential to become widely used across various applications.

As SDSs become increasingly efficient, they will be more readily adopted by the public. They hold great promise for companies, which could automate various tasks through their use. As a result, high expectations have been set on SDSs. At the same time, users tend to have a low tolerance to mistakes and misunderstandings. Disappointment and criticism could ensue if SDSs do not meet these expectations, akin to previous AI winters.

With the current paradigm of pre-trained language models coupled with supervised learning, SDSs could potentially reach a point where they can have seamless interactions with users, but I believe they will still be far from actually *understanding* anything. This generation of researchers could work towards actual *understanding* through e.g. grounding, interactive self-learning, and multimodality. Another interesting avenue is integrating TOD, chit-chat and conversational QA systems for more natural interactions. We tend to anthropomorphize many objects, SDSs that are engaging and natural could provide smoother interactions.

Many people are dissatisfied with chatbot-based customer services, which are typically rule-based. SDSs based on large pre-trained language models could potentially improve user experience, though there is still progress to be made before we can implement and deploy reliable and generalizable systems. Such approaches require large amounts of annotated data, which cannot be acquired in a practical setting. Hence, another interesting research question for this generation of young researchers is efficient and generalizable fine-tuning.

3 Suggested topics for discussion

Reconsidering TOD systems benchmarking As is the case with many NLP problems, benchmarking of TOD systems is inadequate. Biased systems score highly on popular benchmarks and there is little room for improvement. At the same time, these systems are flawed and struggle with their task. Some questions to consider: How do we ensure that good performance on a TOD benchmark leads to robust performance on the task, and ultimately to a robust SDS? How can we obtain larger and more difficult benchmark datasets with accurate and unambiguous annotations?

Disparities between research and industrial applications of SDSs There is a wide gap between SDSs developed in research and those used in industrial applications. This aspect is related to the previous topic in the sense that popular SDS benchmarks do not reflect real SDS applications. Moreover, deep learning-based SDSs lack the controllability needed in an industrial setting. How to ensure that industrial applications of SDSs benefit from the progress in research? How to promote research on more realistic scenarios?

Collaborations between speech and NLP communities As pointed out in Section 1, much TOD research in recent years has focused on textual datasets. Though initially close in the context of SDSs, speech and NLP communities have diverged, with each community publishing in their respective venue thus reducing collaborations (Faruqui and Hakkani-Tür, 2022). How to promote such collaborations and further research in dialogue understanding that is aware of spoken inputs?

References

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. Wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., volume 33, pages 12449–12460.

Pawe\l Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, pages 5016–5026.

Manaal Faruqui and Dilek Hakkani-Tür. 2022. Revisiting the Boundary between ASR and NLU in the Age of Conversational Dialog Systems. *Computational Linguistics* 48(1):221–232.

Alice Shoshana Jakobovits, Francesco Piccinno, and Yasemin Altun. 2022. What Did You Say? Task-Oriented Dialog Datasets Are Not Conversational!? *arXiv:2203.03431 [cs]* .

Seokhwan Kim, Yang Liu, Di Jin, A. Papangelis, Karthik Gopalakrishnan, Behnam Hedayatnia, and Dilek Z. Hakkani-Tür. 2021. “How Robust R U?”: Evaluating Task-Oriented Dialogue Systems on Spoken Conversations. *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* .

Biographical sketch

Léo Jacqmin is an industrial PhD candidate at Orange and Aix-Marseille University working under the supervision of Lina M. Rojas Barahona and Benoit Favre. His research interests lie in the area of task-oriented dialogue systems and dialogue state tracking.

Prior to his PhD, Léo completed a master’s degree in natural language processing at the University of Lorraine. As part of his master’s thesis, he worked on multilingual opinion mining at Orange. Léo’s extracurricular interests include meditation, surfing, and climbing.