Armand Stricker

Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique
91400, Orsay, France
armand.stricker@lisn.upsaclay.fr

1 Research interests

My research interests lie at the **intersection between chitchat and task-oriented dialogues** (TODs), with a specific focus on **integrating capabilities typically associated with chitchat systems into task-oriented agents**. In the SDS literature, these two modes of communication are commonly depicted with a clear contrast. On the one hand, chitchat dialogues are characterized as open-domain and usually involve a wide range of topics; chitchat agents are expected to embody all the qualities of an ideal conversationalist and be empathetic, engaging, knowledgeable, and well-behaved. On the other hand, TODs are closed-domain and rely on specific databases and ontologies; task agents are designed to be efficient and effective tools.

Additionally, the goals of these respective systems are often presented as opposites: a lengthy conversation with a chitchat agent is generally perceived as successful, indicating user engagement and interest, whereas a prolonged conversation with a task-oriented agent is typically considered unsuccessful or suboptimal, suggesting that the user's needs were not met or were met with difficulty.

However, these distinctions are not as clear-cut when it comes to human communication. Most language is not purely transactional[1] or interactional[2] but a mix of both. In fact, exchanges are generally better described as *primarily* transactional or interactional (Brown and Yule, 1983). In the context of task-oriented dialogues, a system that lacks the ability to exhibit remorse when making errors, display empathy when a user's favorite restaurant is unavailable, or address additional context such as having dinner with one's boss vs. with a friend hampers human-system collaboration.

1.1 Towards a More Comprehensive User Understanding

Several studies aim to add chitchat into TODs, such as Accentor (Sun et al., 2021) and FusedChat (Young et al., 2022). However, these methods only add *general* chitchat and do not focus on any specific skill. Not all chitchat may be useful in TODs and a more focused approach should be considered.

Emotional State Understanding a user's task-oriented needs is undeniably crucial. Nevertheless, going beyond that and taking into account their emotional state can result in more suitable responses. It can also compensate for system errors, and even create the impression of a more capable system (Lutfi et al., 2013), ultimately leading to enhanced user satisfaction. Chitchat systems have greatly benefited from emotion detection, enabling them to generate more empathetic responses. I believe this skill can also enhance task-oriented systems, allowing them to better grasp the nuances of user utterances, resulting in more relevant and personalized responses.

To facilitate this, the EmoWOZ dataset (Feng et al., 2022) annotates user turns from the MultiWOZ corpus (Budzianowski et al., 2018) with emotion labels. In my research, I have explored an initial approach called JEmToD (Joint End-to-End Modeling of Emotion Detection and Task-Oriented Dialogue) which generates emotion labels, belief state, dialogue acts and system responses based on a given dialogue history. I have found that although this additional task does not hinder task-oriented performance, it does not improve empathy in system responses and does not provide enough grounding. To address this, I intend to investigate ways of more explicitly conditioning system responses on user emotions. For example, one approach could involve passing JEMTOD's output to a Large Language Model (LLM), instructing it to reformulate JEMTOD's response based on the predicted emotion label.

Beyond the Database A conversation is situated, meaning that contextual information may naturally be introduced. This act can be initiated by the system, to incorporate more diversity and make the dialogue more engaging. The KETOD dataset (Chen et al., 2022) focuses on this effort. System responses rely on relevant information retrieved from Wikipedia about proposed entities and annotators rewrite the original responses to integrate this new information. I plan to experiment with models

[1] the goal is external to the encounter and leads to performing an action, for example.

[2] the goal is internal to the encounter and pertains to the relationship between participants.

Proceedings of the 19th Annual Meeting of the Young Reseachers' Roundtable on Spoken Dialogue Systems, pages 30–32
September 11–12, 2023.

trained on this dataset.

This act may also be initiated by users, as they naturally provide contextual details. Indeed, it is important to acknowledge that users often have multiple, possibly underlying, goals such as needing to blow off steam after a long day, impressing one's significant other, or simply avoiding boredom. This background information may surface during a task-oriented conversation, as elements of backstory or justification of the request are introduced. However, these details tend to either be treated as noise by most task-oriented systems or cause confusion and break down the dialogue. This is quite unlike chitchat systems, trained on dialogues grounded in personas, situations, and general knowledge. We aim to explore avenues for enhancing TOD datasets in a similar manner, leveraging LLMs to do so automatically, as far as possible.

2 Spoken dialogue system (SDS) research

Predicting the state of SDS in the next 5 or even 10 years is challenging due to the rapid evolution of the field. However, it is clear that SDS and text-based dialogue systems are here to stay, offering a convenient means of interacting with machines for non-specialist users, in turn generating increased interest from actors in industry. This enthusiasm drives the direction of SDS towards more reliable systems (particularly in executing tasks like booking tickets or restaurant reservations) capable of providing accurate, relevant and non-hallucinated information. Another promising trend is the growing emphasis on personalization and adaptation of these agents to individual users' preferences, needs, and communication styles.

Leveraging the capabilities of LLMs for SDS presents an exciting opportunity for young researchers in the coming years. Understanding where LLMs fit in relation to TODs is an important topic. It raises questions such as: Can these models be employed in an end-to-end manner? Should they be utilized more prominently in specific components such as natural language understanding or natural language generation? Can they generate training data for early-stage prototyping while awaiting the collection of real-world data? Given their ability to follow instructions and generate coherent text, can they serve as user simulators that account for character traits/personas as well as task-oriented goals ?

While LLMs showcase powerful language capabilities, ensuring they provide precise, factual, and reliable information within narrow domains, as found in TODs, poses a non-trivial challenge. In this regard, research focusing on implementing safeguard mechanisms and constraints for these models needs to be carried out, especially if they are to be employed at scale in sensitive environments.

3 Suggested topics for discussion

- Where do LLMs fit in the task-oriented dialogue pipeline? Is it advisable to use them compared to smaller fine-tuned models?
- How can we enhance the contextual awareness of SDS, in the broad sense of taking into account user persona, situational factors such as a user's mood, open-domain knowledge and commonsense (Bosselut et al., 2019)?
- Human evaluation is often challenging and costly to conduct, however its significance is vital in assessing a system's performance. How can we simplify this evaluation? Is it viable for example to establish a platform that facilitates the pooling of efforts, allowing researchers to conveniently upload their systems for evaluation by fellow experts? Could a standard framework be created ?

References

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, pages 4762–4779. https://doi.org/10.18653/v1/P19-1470.

Gillian Brown and George Yule. 1983. *Teaching the spoken language*, volume 2. Cambridge university press.

Paweł 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. https://doi.org/10.18653/v1/D18-1547.

Zhiyu Chen, Bing Liu, Seungwhan Moon, Chinnadhurai Sankar, Paul Crook, and William Yang Wang. 2022. KETOD: Knowledge-enriched task-oriented dialogue. In *Findings of the Association for Computational Linguistics: NAACL 2022*. Association for Computational Linguistics, Seattle, United States, pages 2581–2593. https://doi.org/10.18653/v1/2022.findings-naacl.197.

Shutong Feng, Nurul Lubis, Christian Geishauser, Hsien-chin Lin, Michael Heck, Carel van Niekerk, and Milica Gasic. 2022. EmoWOZ: A large-scale corpus and labelling scheme for emotion recognition in task-oriented dialogue systems. In *Proceedings*

of the Thirteenth Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, pages 4096–4113. https://aclanthology.org/2022.lrec-1.436.

Syaheerah Lutfi, Fernando Fernández-Martínez, Jaime Lorenzo-Trueba, Roberto Barra-Chicote, and Juan Montero. 2013. I feel you: The design and evaluation of a domotic affect-sensitive spoken conversational agent. *Sensors* 13(8):10519–10538. https://doi.org/10.3390/s130810519.

Kai Sun, Seungwhan Moon, Paul Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, and Claire Cardie. 2021. Adding chit-chat to enhance task-oriented dialogues. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Online, pages 1570–1583. https://doi.org/10.18653/v1/2021.naacl-main.124.

Tom Young, Frank Xing, Vlad Pandelea, Jinjie Ni, and Erik Cambria. 2022. Fusing task-oriented and open-domain dialogues in conversational agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*. volume 36, pages 11622–11629.

Biographical sketch

Armand Stricker is a PhD student at Université Paris-Saclay in the LISN lab, working under the supervision of Patrick Paroubek. His research interests lie in task-oriented dialogues, open-domain dialogues, and in their interaction. Prior to his PhD, Armand completed a master's degree in English as well as a master's in Natural Language Processing at the Université de Paris. As part of his master's thesis, he worked on temporal question-answering. He enjoys swimming, biking, running and discovering new places.