Sopan Khosla

AWS AI Labs
Santa Clara, CA, US

sopankh@amazon.com
https://sopankhosla.github.io/

1 Research interests

My research interests broadly lie in the area of **Information Extraction** from Spoken Dialogue, with a spacial focus on **state modeling**, **anaphora resolution**, **program synthesis & planning**, and **intent classification** in **goal-oriented conversations**. My aim is to create embedded dialogue systems that can interact with humans in a collaborative setup to solve tasks in a digital/non-digital environment.

Most of the goal-oriented conversations usually involve experts and a laypersons. The aim for the expert is to consider all the information provided by the layperson, identify the underlying set of issues or intents, and prescribe solutions. While human experts are very good at extracting such information, AI agents (that build up most of the automatic dialog systems today) not so much. Most of the existing assistants (or chatbots) only consider individual utterances and do not ground them in the context of the dialogue. My work in this direction has focused on making these systems more effective at extracting the most relevant information from the dialogue to help the human user reach their end-goal.

1.1 Information Extraction from Doctor-Patient Dialogue

Following each patient visit, physicians draft long semi-structured clinical summaries called SOAP notes. While invaluable to clinicians and researchers, creating digital SOAP notes is burdensome, contributing to physician burnout. Physicians spend more than 2 hours creating and updating these SOAP notes for every hour of direct patient care.

To automate this arduous task of SOAP note generation, I worked on a pipeline that converts the dialogue into transcripts, performs speaker diarization, extracts the most important utterances from the physician-patient conversation, and then summarizes them in the required format and structure. We built state-of-the-art transformer-based extractive and abstractive summarization architectures to extract the most relevant information from the conversation transcripts (Krishna et al., 2021).

First, the extractive summarization module clusters and classifies the transcript utterances into the SOAP section they contain information for e.g., Past Medical History, Assessment, etc. One of the main novelty points of this module was that it learns contextual representations for each utterance in the conversation by grounding them onto the UMLS (a medical ontology) concepts and conditioning them on the information flow and asymmetric roles/ expertise of the speakers (patient vs physician) (Khosla et al., 2020). Finally, the abstractive module creates a summary for each cluster conditioned on the predicted SOAP note section. This conditioning tailors the output summary to the format expected by each section. Overall, our system was one of the first complete pipelines to automatically generate SOAP notes from conversation transcripts between patients and physicians.

1.2 Anaphora Resolution in Dialogue

Most of the earlier work in the Anaphora Resolution community has focused on expository text. Some example datasets include (most domains within) ONTONOTES (Pradhan et al., 2012), GAP (Webster et al., 2018), etc. The systems built on these datasets often focused only on identity anaphora resolution. More recently, research has been carried out for interpretations beyond identity anaphora in datasets like ARRAU (Poesio et al., 2018).

During my Masters, I worked on creating new benchmarks and systems for three types of anaphoric relations (identity, bridging, discourse deixis) in a dialogue setting. I spearheaded the creation of multiple dialogue datasets labeled with these different types of anaphoric relations. These datasets were then used to host the CODI-CRAC 2021 (Khosla et al., 2021) and 2022 (Yu et al., 2022a) Shared-tasks where we invited other researchers in the community to build new systems that solve this problem. I also worked on the metrics that were used to score the different systems that were submitted to the shared task (Yu et al., 2022b). We created a first of its kind state-of-the-art benchmark dataset, and a baseline system to perform automatic resolution of these three different types of anaphoric relationships in dialogue. Our system was built on top of a transformer-based encoding layer, trained, and evaluated to perform generalizable anaphora resolution in different types of dialogue settings.

1.3 Intent Classification in Dialogue

My ongoing work is in performing contextual intent classification in spoken & written dialogue between humans and an agent. Most of the existing production-ready as-

Proceedings of the 19th Annual Meeting of the Young Reseachers' Roundtable on Spoken Dialogue Systems, pages 51–53
September 11–12, 2023.

sistants are not good at grounding the interactions in the context of the dialogue.

I am actively researching on creating dialogue systems that can perform context-dependent intent classification on the incoming user utterance, and interact with external tools/ APIs to perform further processing conditioned on that intent. I worked on a transformer-based state-of-the-art intent classification system that not only classifies incoming utterances into different intents that the assistant can handle, but also detect utterances that are out-of-scope for the assistant's current capabilities to gracefully convey to the customer (Khosla and Gangadharaiah, 2022b). In our recent work, we also created a new intent-classification dataset that evaluates the prowess of state-of-the-art models on samples that can prove to be adversarial in the production scenario Khosla and Gangadharaiah (2022a). The dataset was a significant contribution as it was a first of its kind that evaluated intent classification systems on non-iid distributions.

2 Spoken dialogue system (SDS) research

Owing to the fast-paced innovations in language and speech, SDSs are likely to transition into becoming useful assistants for the human users. They might go one step further, and be able to interact with their environment to perform tasks and help the user achieve their goal. To get to this stage, however, SDS research has to focus on dialog management modules that are capable of accurately modeling user's intents and goals, translating those intents into actionable steps (or programs), executing those steps in their (digital) environment, and deploying remedial measures when needed. All of which will need to happen in a transparent, verifiable, and controlled setting.

3 Suggested topics for discussion

- End-to-end vs Modular methodologies for Spoken Dialogue Assistants.
- Program Synthesis and Planning in Dialogue Assistants to perform complex tasks.
- Methods for Efficient Interaction with Digital/Non Digital APIs and Tools.
- Modeling multi-modal context in Dialogue.

References

Sopan Khosla and Rashmi Gangadharaiah. 2022a. Benchmarking the covariate shift robustness of open-world intent classification approaches. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Association for Computational Linguistics, Online only, pages 14–23. https://aclanthology.org/2022.aacl-short.3.

Sopan Khosla and Rashmi Gangadharaiah. 2022b. Evaluating the practical utility of confidence-score based techniques for unsupervised open-world classification. In *Proceedings of the Third Workshop on Insights from Negative Results in NLP*. Association for Computational Linguistics, Dublin, Ireland, pages 18–23. https://doi.org/10.18653/v1/2022.insights-1.3.

Sopan Khosla, Shikhar Vashishth, Jill Fain Lehman, and Carolyn Rose. 2020. MedFilter: Improving Extraction of Task-relevant Utterances through Integration of Discourse Structure and Ontological Knowledge. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, pages 7781–7797. https://doi.org/10.18653/v1/2020.emnlp-main.626.

Sopan Khosla, Juntao Yu, Ramesh Manuvinakurike, Vincent Ng, Massimo Poesio, Michael Strube, and Carolyn Rosé. 2021. The CODI-CRAC 2021 shared task on anaphora, bridging, and discourse deixis in dialogue. In *Proceedings of the CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue*. Association for Computational Linguistics, Punta Cana, Dominican Republic, pages 1–15. https://doi.org/10.18653/v1/2021.codi-sharedtask.1.

Kundan Krishna, Sopan Khosla, Jeffrey Bigham, and Zachary C. Lipton. 2021. Generating SOAP notes from doctor-patient conversations using modular summarization techniques. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, pages 4958–4972. https://doi.org/10.18653/v1/2021.acl-long.384.

Massimo Poesio, Yulia Grishina, Varada Kolhatkar, Nafise Moosavi, Ina Roesiger, Adam Roussel, Fabian Simonjetz, Alexandra Uma, Olga Uryupina, Juntao Yu, and Heike Zinsmeister. 2018. Anaphora resolution with the ARRAU corpus. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*. Association for Computational Linguistics, New Orleans, Louisiana, pages 11–22. https://doi.org/10.18653/v1/W18-0702.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In *Joint Conference on*

EMNLP and CoNLL - Shared Task. Association for Computational Linguistics, Jeju Island, Korea, pages 1–40. https://aclanthology.org/W12-4501.

Kellie Webster, Marta Recasens, Vera Axelrod, and Jason Baldridge. 2018. Mind the GAP: A balanced corpus of gendered ambiguous pronouns. *Transactions of the Association for Computational Linguistics* 6:605–617. https://doi.org/10.1162/tacl$_{a0}$0240.

Juntao Yu, Sopan Khosla, Ramesh Manuvinakurike, Lori Levin, Vincent Ng, Massimo Poesio, Michael Strube, and Carolyn Rosé. 2022a. The CODI-CRAC 2022 shared task on anaphora, bridging, and discourse deixis in dialogue. In *Proceedings of the CODI-CRAC 2022 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue*. Association for Computational Linguistics, Gyeongju, Republic of Korea, pages 1–14. https://aclanthology.org/2022.codi-crac.1.

Juntao Yu, Sopan Khosla, Nafise Sadat Moosavi, Silviu Paun, Sameer Pradhan, and Massimo Poesio. 2022b. The universal anaphora scorer. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pages 4873–4883. https://aclanthology.org/2022.lrec-1.521.

Biographical sketch

Sopan Khosla is an Applied Scientist at AWS AI Labs. His research focuses on Information Extraction from Spoken Dialogue, with a spacial focus on state modeling, anaphora resolution, program synthesis & planning, and intent classification in goal-oriented conversations. He holds a Masters degree in Language Technologies from Carnegie Mellon University, where he was advised by Prof. Carolyn Rose. During his masters, he worked on problems relating to anaphora resolution, discourse modeling, and knowledge grounding in dialogues. In his free time, he enjoys playing Badminton and Tennis.