Instructions for ACL-2016 Proceedings

Tom Byars, Cale Clark, Roman Fenlon, Charlie Lyttle, Katie McAskill, Jack Miller, Zein Said, Abdullah Sayed, Laura Schauer, Jason Sweeney, Aron Szeles, Xander Wickham
School of Mathematics and Computer Science, Heriot-Watt University, Edinburgh
tjb10, cc164, rf104, cl157, klm12, jjm7, zs2008,
as512, lms9, js418, as472, aw127@hw.ac.uk

Abstract

This should be a 6-8 page conference paper with appendices, if relevant. Good reports from last year: 1 and 7

1 Introduction

from coursework spec: main research or technical question addressed
Socially assistive robots (SARs) are a crucial part of the future of many sectors, for example, in education or healthcare (Gunson et al., 2022). Especially the latter depends on technology advancements as it is facing numerous obstacles in the future, such as increasing spendings and a growing percentage of older people. A serious lack of healthcare workers is already occurring, with 10 million more healthworkers needed worldwide by 2030 (Cooper et al., 2020; WHO, 2023). SARs can pose a solution to the problem, as they are able to support healthcare in various ways, such as encouraging older people to keep living independently for longer or reducing caregiver burden (Cooper et al., 2020).

These scenarios require SARs to be able to handle multi-party interactions as it is likely that more than one person will interact with the system. Compared to handling dyadic interactions, handling multi-party conversations includes more complex challenges, such as Speaker Recognition, Addressee Recognition, Response Selection (summarised in "who says what to whom") and turn-taking (Addlesee et al., 2023; Johansson and Skantze, 2015).

Include here what exactly we examined about turn-taking

In this work, we propose a model trained on multi-party human-human conversation data. We collected the data from recordings of special "Who wants to be millionaire?" episodes where two candidates collaborated to answer the host's questions.

Include results here.

2 Background

from coursework spec: literature review / related work, including a critical analysis of the field, and commentary on applicability of the technologies and methods used in emerging technologies and application areas

2.1 Socially Assistive Robots

For healthcare, as well as for any other sector, the difficulty of successfully designing SARs lies in creating robots that can effectively converse with humans and adhere to social norms (Moujahid et al., 2022). The more expressive a robot is, the more it will be perceived as intelligent, conscious and polite (Moujahid et al., 2022). To achieve such a positive perception, multiple parts need to be combined into one conversational system, such as the ability to carry out visually grounded as well as task-based dialogues, to perceive and discuss its environment and to chit-chat (Gunson et al., 2022).

The SPRING project conducts research on a SAR robot that is deployed in an eldercare hospital reception area (Addlesee et al., 2020). The conversational system is deployed on humanoid ARI robot produced by Pal Robotics (Robotics, 2023). ARIs capabilities can be extended with custom AI algorithms, in the case of SPRING-ARI a visual perception system, a dialogue system, and a social interaction planner (Addlesee et al., 2020). While the SPRING-ARI system successfully demonstrates that task-based, social and visually grounded dialogue can be combined with physical actions, it still lacks the ability to handle

conversations with more than one person simultaneously (Addlesee et al., 2020).

2.2 Multi-party Human Robot Interaction

As stated above, the endeavour to create conversational systems becomes considerably more difficult when dealing with multi-party interactions (Addlesee et al., 2023). Especially turn-taking poses a central problem. It is defined as follows:

> The rules of turn-taking organize the conversation into turns, during which one of the participants has the right to speak while the others agree to listen (Żarkowski, 2019)

In dyadic conversations, there are only two roles a participant can take: speaker or listener, hence it is clear when and to whom the turn is yielded. In multi-party conversations, participants can take multiple roles, therefore turn-taking needs to be coordinated (Johansson and Skantze, 2015). Humans signal their intents mostly through gaze, but also through pauses, prosody, and body positioning (Żarkowski, 2019). To copy this behaviour, earlier models for conversational systems relied on silence time-outs to coordinate turn-taking, however, this approach is found to be too simplistic (Skantze, 2021). Instead, mimicking human turn-taking behaviour better by using a combination of verbal and non-verbal cues leads to robots that are better perceived (Moujahid et al., 2022).

State exactly the gap that we will fill - whatever that will be

3 Data Collection

- (Laura) Talk about multi-party data collection
- (Aron and Katie) How we collected our data, describe the intents we used to label the data
- (Aron and Katie) Cohen's Kappa for our data collection method: probably need to annotate a couple of transcripts twice for reporting on this

4 Design and Implementation

- *from coursework spec:* design and implementation of the system: components and architecture
- Jack's System Graph

4.1 Natural Language Understanding

- Very short escription of RASA
- Refer to the intents described in 3 Data Collection
- Different versions of the model: show difference in F-Score, Confusion Matrix, ... (Laura: look up in literature what is used, what should be used)

4.2 Dialogue Management

- Clearly explain 2 parts (State-Machine and NN)
- State-Machine: high-level control, handles things we have no data for, eg. pauses
- NN: Report on differences between RNN and LSTM and why the choice for the LSTM has been made
- (optional): Comparison to RASA rule-policy

4.3 Natural Language Generation

5 Evaluation

from coursework spec: evaluation of the system and presentation of the results

5.1 Methodology

5.2 Experiment Layout

5.3 Results

6 Conclusion

6.1 Ethical Reflection

7 Future Work

from coursework spec: suggestions for future work

Acknowledgments

The acknowledgments should go immediately before the references. Do not number the acknowledgments section. Do not include this section when submitting your paper for review.

References

[Addlesee et al.2020] Angus Addlesee, Yanchao Yu, and Arash Eshghi. 2020. A comprehensive evaluation of incremental speech recognition and diarization for conversational AI. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3492–3503. International Committee on Computational Linguistics.

[Addlesee et al.2023] Angus Addlesee, Weronika Sieinska, Nancie Gunson, Daniel Hernández García, Christian Dondrup, and Oliver Lemon. 2023.

[Cooper et al.2020] Sara Cooper, Alessandro Di Fava, Carlos Vivas, Luca Marchionni, and Francesco Ferro. 2020. ARI: the social assistive robot and companion. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 745–751. ISSN: 1944-9437.

[Gunson et al.2022] Nancie Gunson, Daniel Hernandez Garcia, Weronika Sieińska, Angus Addlesee, Christian Dondrup, Oliver Lemon, Jose L. Part, and Yanchao Yu. 2022. A visually-aware conversational robot receptionist. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 645–648. Association for Computational Linguistics.

[Johansson and Skantze2015] Martin Johansson and Gabriel Skantze. 2015. Opportunities and obligations to take turns in collaborative multi-party human-robot interaction. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, page 305–314, Prague, Czech Republic, Sep. Association for Computational Linguistics.

[Moujahid et al.2022] Meriam Moujahid, Helen Hastie, and Oliver Lemon. 2022. Multi-party interaction with a robot receptionist. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 927–931.

[Robotics2023] Pal Robotics. 2023. Ari - the social and collaborative robot. Accessed on: 2023-02-08.

[Skantze2021] Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech and Language*, 67:101178.

[WHO2023] WHO. 2023. Global health workforce statistics. Accessed on: 2023-02-07.

[Żarkowski2019] Mateusz Żarkowski. 2019. Multi-party turn-taking in repeated human–robot interactions: An interdisciplinary evaluation. *International Journal of Social Robotics*, 11(5):693–707, Dec.

A Supplemental Material, Appendix