The 'Icebreaker' Challenge for Social Robotics

Weronika Sieińska, Nancie Gunson, Christian Dondrup, and Oliver Lemon*
(w.sieinska;n.gunson;c.dondrup;o.lemon)@hw.ac.uk
School of Mathematical and Computer Sciences, Heriot-Watt University
Edinburgh, UK

ABSTRACT

Loneliness in the elderly is increasingly becoming an issue in modern society. Moreover, the number of older adults is forecast to increase over the next decades while the number of working-age adults will decrease. In order to support the healthcare sector, Socially Assistive Robots (SARs) will become a necessity. We propose the multi-party conversational robot 'icebreaker' challenge for NLG and HRI that is not only aimed at increasing rapport and acceptance of robots amongst older adults but also aims to tackle the issue of loneliness by encouraging humans to talk with each other.

1 INTRODUCTION

In recent years robots have been introduced into a variety of public spaces most notably museums [3, 15, 22], commercial malls [11], hospitals, retirement homes [13], etc. The latter two are becoming ever more important since according to Eurostat [9], by 2035 there will be an estimated 16.2 million people aged 80 and over in France, Germany, and the United Kingdom alone. At the same time the working population is expected to decline, which will put an additional burden on hospitals and care homes. We believe that Socially Assistive Robots (SARs) are a way of mitigating the strain put on the health services. While this HRI problem has been addressed in several projects on national and international level, e.g. [13], especially multi-party NLG has vastly been neglected apart from confirming given instructions. To increase acceptance and usability of SARs, however, they must be able to communicate via speech which is arguably the most natural way of communicating amongst humans. This is especially true when it comes to less technically savvy groups such as older adults. One way to overcome this might be the addition of a dialogue system to SARs such as "Alana" [6] including specialised NLU [24] and task execution [8] components, which is able to hold a conversation with the person not only about the specific scenario they are in but also about general topics (music, movies, news, fashion, celebrities, etc.) to build rapport. Most of these systems, however, are not built for multi-party interaction. Hence, we propose the HRI–NLG challenge of creating a multi-party conversational 'icebreaker' robot for use in care homes and hospitals.

The Icebreaker challenge. According to several sources, e.g. [1, 16, 23], one of the biggest current issues in the older population is loneliness, and here SARs provide a unique opportunity to improve social life. As part of the SPRING project [21], we will develop a robotic system with dialogue capabilities to act as an 'icebreaker' in specific social interactions. The project focuses on the well-being of older adults and more specifically on supporting their visits at a hospital in France. Patients attend the hospital for a full day at a time where they are sitting in a common waiting room, waiting to be called for their appointments. While waiting, the vast majority of patients sit quietly until their next appointment. We aim to deploy a robot to this waiting room that can not only give patients information about their next appointment (thus reducing anxiety) and guide them to it, but which also involves them in more general dialogue using open-domain social chit-chat. The challenge we propose is to use non-intrusive general social conversation [6] to involve other people, e.g. another patient sitting next to the patient we are already talking to, in the conversation with the robot. Hence we aim to 'break the ice' and get both patients to talk together in this multi-party interaction. At some point the robot will then remove itself from the conversation, hopefully having started a dialogue between the two patients themselves.

2 OPEN CHALLENGES

2.1 NLG and Response Generation challenges:

NLG for the elderly – Older people might be hard of hearing or suffer from memory impairments. This will require special care when generating speech such as repetition of certain information from previous turns, a simpler sentence structure, and shorter, slower utterances.

NLG for Conversation initiation – Should the robot initiate conversation one-on-one with one person before including the other, or approach both simultaneously? The former may be simpler in terms of initial conversation and addressee management but raises questions of when and how to include the other person. The latter raises the question of how to start a conversation with multiple strangers.

Topic selection – A first choice is in the selection of appropriate (neutral) conversational openers - typically in the UK a comment on the weather, or travel related ("Have you come far today?") but which may be culturally dependent. Moving on, it will be important to select topics of interest to the cohort [19] and which encourage self-disclosure.

Floor and addressee management – Understanding and managing turns in a open-world, multi-party setting of this nature is highly complex [2, 10, 26]. Addressee selection is a critical issue - who to address and when, and how to make the addressee clear. In this context, also, when to 'lead' the conversation e.g. asking users about their interests, when to 'withdraw' into the background e.g. as the conversation flourishes between the other parties (number of direct turns increases); and if and when to intervene when a problematic dialogue is detected e.g. there is evidence of misunderstanding between the other participants [20].

*All authors contributed equally.

2.2 HRI challenges:

Scheduling – A major challenge in HRI is to provide a robot that will be accepted by the target audience – older adults and hospital staff in our case. While social chat can help to build rapport with the patients, we also have to make sure that the robot remains a useful tool to staff. Hence, a big challange is to decide when to engage in rapport building and 'ice breaking'.

Willingness to interact – The robot has to identify one or multiple people to engage in conversation. The willingness to interact has to be evaluated to reduce annoyance amongst patients that do not want to be approached.

Approaching – The robot needs to approach either one or multiple people in a manner that communicates the willingness to engage in conversation. A lot of prior work has investigated this issue, e.g. [4, 7, 14], but no holistic approach has been found.

Active sensing – During multi-party interaction, to enhance visual perception and allow the robot to augment the speech signal with multimodal information such as visually perceived emotional states, it is necessary to turn the head towards the speaker to get them into the field of view.

Multimodal output – This includes generating and coordinating the speech signal, appropriate gaze behaviour, and gestures, for regulating the conversation and/or addressee management [5]. Head pose, in particular, will be vital in supporting addressee management. This, however, might conflict with the active sensing requirements.

Safety and reliability – Hospitals are ever-changing and chaotic environments. The robot has to be able to reliably navigate and execute tasks to not become a burden to staff. It cannot be a roadblock in case of fire or other emergencies. Hence, the robot has to comply with all safety standards and regulations and has to be aware of its surroundings.

2.3 Related challenges:

Addressee identification – Determining who the "you" is in "what do you think?" is a key question [12]. Is the speaker referring to a specific individual, the group, or using the word in a generic sense? E.g. "Well, usually what **you** do is...". Linguistic-based features can potentially be supported by multimodal output as mentioned earlier.

Ellipsis resolution – In order to resolve statements such as e.g. "I agree" in multi-party conversations the system will need access to keep a shared dialogue context / conversation history of all participants.

Split utterances – The occurrence of single utterances split between two or more dialogue turns or speakers [18], is likely to increase in a multi-party situation. Moreover, it may be a particular issue in this user group, some of whom may have cognitive issues and companions who tend to speak on their behalf and perform utterance completion.

Monitoring conversation status – To build rapport between humans and robots, and between humans and humans, detecting agreement and sentiment analysis are important challenges to tackle for multi-party conversations.

Speech recognition – Speech signals from potentially multiple sources are intermingled with significant background noise which requires specialised hardware and software. On top of this *speaker diarization* [25] has to be performed in order to identify who is speaking and when, focusing only on those speakers that have actively engaged with the robot. This can be aided by vision components but that might interfere with the multimodal output mentioned earlier or suffer from a narrow field of view.

3 EVALUATION METHODOLOGY AND METRICS

Given the desire to promote social interaction, the *duration* of the conversation achieved between humans is an obvious evaluation metric. There are however subtleties within this e.g. the point at which participants are deemed to be talking to each other, rather than the robot. Moreover, a long(er) conversation may not always be a happy one. Of interest is the degree of rapport between participants. Previous work has suggested that interactional features such as participants' word count per sentence and number of interruptions are useful indicators of a relationship [17]. The energy of the speech signals themselves may be relevant — is a raised voice a sign of disagreement, of enthusiasm, or in this context simply related to difficulty hearing? With the addition of the robot's capabilities visual factors such as users' body language (closed/open, shifts in orientation towards/away from the other person) will also provide valuable information on the conversation status. Exploring and optimising the hallmarks of a 'successful' icebreaker conversation is another interesting aspect of this challenge.

ACKNOWLEDGMENTS

This research has been funded by the EU H2020 program under grant agreement no. 871245 (SPRING).

REFERENCES

[1] AgeUK. 2018. All the Lonely People: Loneliness in Later Life. https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/loneliness/loneliness-report_final_2409.pdf.

[2] Dan Bohus and Eric Horvitz. 2011. Multiparty Turn Taking in Situated Dialog: Study, Lessons, and Directions. In *Proceedings of the SIGDIAL 2011 Conference: 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 98–109.

[3] Wolfram Burgard, Armin B Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. 1999. Experiences with an interactive museum tour-guide robot. *Artificial intelligence* 114, 1 (1999), 3–55.

[4] J. T. Butler and A. Agah. 2001. Psychological Effects of Behavior Patterns of a Mobile Personal Robot. *Autonomous Robots* 10, 2 (March 2001), 185–202. https://doi.org/10.1023/A:1008986004181

[5] Justine Cassell. 1998. Embodied Conversation: Integrating Face and Gesture into Automatic Spoken Dialogue Systems.

[6] Amanda Cercas Curry, Ioannis Papaioannou, Alessandro Suglia, Shubham Agarwal, Igor Shalyminov, Xinnuo Xu, Ondrej Dušek, Arash Eshghi, Ioannis Konstas, Verena Rieser, et al. 2018. Alana v2: Entertaining and informative open-domain social dialogue using ontologies and entity linking. *Alexa Prize Proceedings* (2018).

[7] Kerstin Dautenhahn, Michael Walters, Sarah Woods, Kheng Lee Koay, Chrystopher L Nehaniv, A Sisbot, Rachid Alami, and Thierry Siméon. 2006. How may I serve you?: a robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 172–179.

[8] Christian Dondrup, Ioannis Papaioannou, and Oliver Lemon. 2019. Petri Net Machines for Human-Agent Interaction. *arXiv preprint arXiv:1909.06174* (2019).

[9] Eurostat. 2017. People in the EU – population projections. https://ec.europa.eu/eurostat/statistics-explained/index.php/People_in_the_EU_-_population_projections.

[10] Raquel Fernández, Matthew Frampton, Patrick Ehlen, Matthew Purver, and Stanley Peters. 2008. Modelling and detecting decisions in multi-party dialogue. In

ACL-08: HLT - Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue. 156–163. https://doi.org/10.3115/1622064.1622095

[11] Mary Ellen Foster, Rachid Alami, Olli Gestranius, Oliver Lemon, Marketta Niemelä, Jean-Marc Odobez, and Amit Kumar Pandey. 2016. The MuMMER project: Engaging human-robot interaction in real-world public spaces. In *International Conference on Social Robotics*. Springer, 753–763.

[12] Surabhi Gupta, John Niekrasz, Matthew Purver, and Dan Jurafsky. 2007. Resolving "You" in multi-party dialog. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*. 227–230.

[13] Nick Hawes, Christopher Burbridge, Ferdian Jovan, Lars Kunze, Bruno Lacerda, Lenka Mudrova, Jay Young, Jeremy Wyatt, Denise Hebesberger, Tobias Kortner, et al. 2017. The strands project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine* 24, 3 (2017), 146–156.

[14] Jens Kessler, Christof Schroeter, and Horst-Michael Gross. 2011. Approaching a person in a socially acceptable manner using a fast marching planner. In *Intelligent Robotics and Applications*. Springer, 368–377.

[15] Irene Macaluso, Edoardo Ardizzone, Antonio Chella, Massimo Cossentino, Antonio Gentile, R Gradino, Ignazio Infantino, Marilia Liotta, Riccardo Rizzo, and Giuseppe Scardino. 2005. Experiences with CiceRobot, a museum guide cognitive robot. In *AI* IA 2005: Advances in Artificial Intelligence*. Springer, 474–482.

[16] NHS. 2018. Loneliness in older people. https://www.nhs.uk/conditions/stress-anxiety-depression/loneliness-in-older-people/.

[17] Florian Pecune, Jingya Chen, and Justine Cassell. 2018. Field Trial Analysis of Socially Aware Robot Assistant. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems, AAMAS*. 1241–1249.

[18] Matthew Purver, Christine Howes, Patrick G.T. Healey, and Eleni Gregoromichelaki. 2009. Split utterances in dialogue: A corpus study. *Proceedings of the SIGDIAL 2009 Conference: 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue* September (2009), 262–271.

[19] S. Zahra Razavi, Lenhart K. Schubert, Kimberly A. Van Orden, and Mohammad Rafayet Ali. 2019. Discourse Behavior of Older Adults Interacting With a Dialogue Agent Competent in Multiple Topics. arXiv:1907.06279 http://arxiv.org/abs/1907.06279

[20] Alexander Schmitt, Carolin Hank, and Jackson Liscombe. 2008. Detecting Problematic Dialogs with Automated Agents. In *Perception in Multimodal Dialogue Systems, 4th IEEE Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems, PIT 2008, Kloster Irsee, Germany, June 16-18, 2008, Proceedings*. 72–80. https://doi.org/10.1007/978-3-540-69369-7_9

[21] SPRING. 2020. Socially Pertinent Robots in Gerontological Healthcare. http://spring-h2020.eu/.

[22] Sebastian Thrun, Michael Beetz, Maren Bennewitz, Wolfram Burgard, Armin B Cremers, Frank Dellaert, Dieter Fox, Dirk Haehnel, Chuck Rosenberg, Nicholas Roy, et al. 2000. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *The International Journal of Robotics Research* 19, 11 (2000), 972–999.

[23] UNECE. 2017. REPORT ON THE FRENCH AGEING POLICY. https://www.unece.org/fileadmin/DAM/pau/age/country_rpts/2017/FRA_Report_ENG.pdf.

[24] Andrea Vanzo, Danilo Croce, Emanuele Bastianelli, Roberto Basili, and Daniele Nardi. 2020. Grounded language interpretation of robotic commands through structured learning. *Artificial Intelligence* 278 (2020), 103181.

[25] Quan Wang, Carlton Downey, Li Wan, Philip Andrew Mansfield, and Ignacio Lopz Moreno. 2018. Speaker diarization with LSTM. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* 2018-April (2018), 5239–5243. https://doi.org/10.1109/ICASSP.2018.8462628 arXiv:1710.10468

[26] Rui Zhang, Honglak Lee, Lazaros Polymenakos, and Dragomir Radev. 2018. Addressee and response selection in multi-party conversations with speaker interaction RNNs. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018* 2016 (2018), 5690–5697. arXiv:1709.04005