Towards a Conversational Corpus for Human-Robot Conversations

Dagoberto Cruz-Sandoval[2], Friederike Eyssel[1],Jesus Favela[2],Eduardo B. Sandoval[1]
[1] CITEC,University of Bielefeld, Germany
[2] Computer Science Department,CICESE, Mexico
dagoberto@cicese.edu.mx,feyssel@cit-ec.uni-bielefeld.de,
favela@cicese.mx,ed@hoiho.co

ABSTRACT

Conversational corpora based on human-human dialogues have often been used for training of data-driven dialogue systems. However, human-human conversations might not be the optimal inputs for machine learning training aims used in HRI. This paper suggests the creation of a conversational corpus based on Human-Robot conversations as input for the training of a dialogue system used in future conversational robots. We propose that the significant differences between Human-Human Conversation (HHC) and Human-Robot Conversation (HRC) in terms of used language and other aspects (e.g., humanlikeness, embodiment, etc.) might affect the quality of the responses from a conversational robot. Hence, the use of HRCs as an input could improve the responses of the robots when the conversational machine learning system is trained using a more realistic model of HRI conversations rather than a HHI model. Future applications of conversational robots in education and health care could be enhanced by using an appropriate HRC corpus.

1. INTRODUCTION

Humans do not talk to robots like they talk to other humans. Apparently, humans tend to adapt their language to the social interface depending on the features of this interface [1]. Hence, to use Human-Human Conversations (HHC) as an input for the creation of corpus for use in HRI contexts might create a significant bias when human-robot conversational systems are designed. Similarly, a very natural, human-like response from a robot could negatively affect the interaction. For instance, previous research suggested that the use of indirect, typically human language by robots results in less effective Human-Robot Interaction (HRI) and less reciprocity interactions [4]. Embodiment, degree of anthropomorphism, and non-verbal language could be also factors that affect the language used in conversational robot system compared with other types of interactive systems. Hence, to use HHC could be not the optimal input for the training of a conversational robot system.

In this paper, we propose to build a conversational corpus using Human-Robot Conversations (HRC) as input. The future aim of this research is to use this corpus for training dialogue systems based on deep learning methods to endow robots with the ability to engage in natural conversation with humans (see **Figure 1**). A Wizard-of-Oz framework will be used to generate HRC. Each conversation must be recorded, transcripted, and processed before it will added to the corpus. We consider that the quality of the outputs in the robotic conversational system would be significantly better using a HRI input. To our knowledge, a dialogue corpus based on HRC has not been proposed yet.

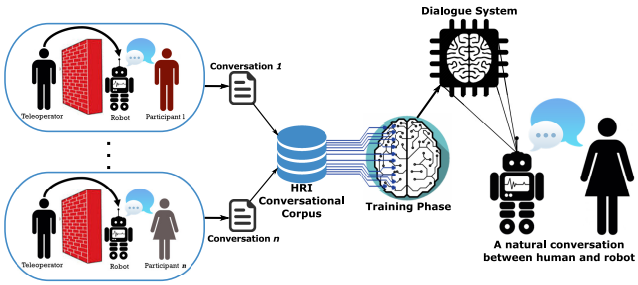

Figure 1: Human-robot conversations used for training of the corpus for a robotic dialogue system.

2. PREVIOUS WORK

Creating natural, smooth human-robot dialogues has been a technical challenge for years. However, the development of deep neural networks architectures, has made it possible to build dialogue systems to establish a natural conversation between humans and machines. These dialogue systems require a corpus of conversations for the training of neural networks.A conversational corpus is defined a collection of transcriptions of many different conversations [2]. One of the most important distinctions between conversational corpora is concerning the nature of the conversational agents. There are corpora based on human-human, human-computer, and even computer-computer conversations [7]. However, not so much work about of HRC corpora has been done. An early contribution about this topic can be found in [3]. These authors suggest an specific system design for scenarios of Multiparty Dialogue management (MDM) using HRC as an input for a Bayesian system aiming to measure the *focus*

HRI '17 Companion March 06-09, 2017, Vienna, Austria

ACM ISBN 978-1-4503-4885-0/17/03.
DOI: http://dx.doi.org/10.1145/3029798.3038344

of attention of the user. Hence, the way that individuals communicate with a robot matters in terms of language adaptations made during the interaction. If we assume that most of the robotic conversational systems aim to use spoken natural language, dialogue systems based on Generative Response Models (GRM) might be useful. These GRMs attempt to generate responses by keeping possible utterances. GRMs can generate entirely novel responses. These models learn to assign a probability to every possible conversation. Furthermore, since GRMs generate responses word by word, they must learn to simulate the behaviour of the agents used to build up the corpus . Thus, training is a key process for these dialogue systems because the features of the corpus impact in the output's quality of the dialogue system [5]. Finally, we consider that a training corpus based on HRC is not directly transferable to HRI conversational model. According to Williams, human-human turn-taking is much richer than human-machine dialogue [7]. In contrast, HRC do not contain the same distribution of understanding errors, pauses, second intentions, hidden meaning and other human features present in HHC. Hence, it is important to create high quality inputs to be used in future HRC.

3. DESIGN OF A CORPUS BASED ON HRC

The design of a training corpus based on HRC would be done in a Wizard-of-Oz scenario using a humanoid robot with an embedded automatic speech recogniser and text-to-speech synthesiser controlled by a hidden operators with a high level of expertise in interaction using their native language. In order to capture the different conversational styles, we propose a cross-cultural study that involves the record of significant samples (30-40 participants) to obtain conversations in three languages: English, Spanish and German. We consider these languages due to their international influence, and level of penetration as a second language in several regions of the world. The operator would be a native speaker of each language and the conversational topics would be related to the main future uses of social robots such as: education, health care, entertainment and housekeeping. In our proposed setup, the robot exhibits the ability to enact a natural, open domain conversation with a human. Moreover, the robot can exhibits gestures to enrich the interaction, for example, nodding and gesticulation.To generate personalised responses to the user, it is useful to have access to either long-term dialogues or to dialogues with returning users. Our proposal is based on a hybrid of these two approaches. Thus, we consider a middle-size group of participants who will converse with the robot at least three times over a period of one week. Before of the study, we will conduct a contextual enquiry based on interviews with each subject. Contextual information (e.g., gender, age, hobbies, activities) will be used to enact conversations via the robot. During a new interaction between a participant and the robot, the tele-operator will take information from past conversations to deep in the specific topics.

Once the HRCs are recorded we aim to perform a data processing stage to standardise the corpus with minimal loss of information for the GRM system. During the transcription process, we must follow a defined guideline, the most relevant aspects to be modelled are: **a) Turn-taking:** Add explicit labels to each speaker (human or robot) in the turn-taking during a conversation. **b) Use of spoken language:** Normalise all the entries it is also necessary. Thus, a transcription must do not include acronyms, abbreviations, fillers (e.g um, uh, oh) and phonemicizations.**c) Anaphora:** We need to manage the use of anaphora - the use of a word such as a pronoun that has the same reference as a word previously used in the same discourse - in conversations. **d) Spelling errors:** If we avoid spelling errors, we expect to reduce data sparsity.

The training methods proposed for this conversational system using HRC would be based on prediction of the next sentence given the previous sentence or sentences in a conversation. The system will train to maximise the cross entropy of the correct sequence given its context. This approach use deep neural networks to map sequences to sequences [6].

4. RESULTS AND FUTURE WORK

We started to build a multilingual robust open-source training corpus for a generative dialogue robotic system using human-robot conversations. We consider that our contribution in HRI field lies in the quality of the responses of this system. We expect our HRI corpus offer a better performance compared with existing systems due to the real HRC inputs and keep it open source. Hence, humans interacting with our system could have more effective, productive and natural interactions with it. We also aim to an ample range of applications. For instance, health care focused in mental diseases as dementia, education and entertainment. As future work, we would like to extend the HRC to other languages such as Chinese and Arabic to cover different conversational styles.

5. REFERENCES

[1] Has voice control finally started speaking our language? https://www.theguardian.com/technology/2016/dec/04/voice-control-amazon-echo-digital, 2016.

[2] G. Kennedy. *An Introduction to Corpus Linguistics*. Addison-Wesley-Longman, London, 1st edition, 1998.

[3] C. Kennington, K. Funakoshi, Y. Takahashi, and M. Nakano. Probabilistic multiparty dialogue management for a game master robot. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 200–201, New York, NY, USA, 2014. ACM.

[4] E. B. Sandoval, J. Brandstetter, and C. Bartneck. Can a robot bribe a human? The measurement of the negative side of reciprocity in human robot interaction. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 117–124. IEEE, Mar 2016.

[5] A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, and W. B. Dolan. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In *Naacl-2015*, pages 196–205, Jun 2015.

[6] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems (NIPS)*, pages 3104–3112, 2014.

[7] J. D. Williams and S. Young. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422, 2007.