Towards Language Learning for Safety during Human-Robot Social Interaction

Chandrakant Bothe, Sven Magg, Cornelius Weber and Stefan Wermter
Knowledge Technology, Department of Informatics
University of Hamburg
Hamburg, Germany
{bothe, magg, weber, wermter}@informatik.uni-hamburg.de

***Abstract*—This document is a progress report provided as a requirement of the SECURE EU Project. In this work, we tackle the problem of developing a dialogue system for robots with multivariate behavioral adaptation as a preliminary step towards potentially learning safety concepts for safer human-robot interaction. With the concern of safety during the human-robot verbal interaction, the aim is to study and research different linguistic aspects. vAs language is very complex, different linguistic features could be used to assess the behavior. We start with sentiment guided learning of the safety concepts. We also found that for the language interaction, we need a dialogue system which drives dialogue flow. In natural language understanding, dialogue act, which represents a functional type of utterance, plays a very important role in a dialogue system. We developed neural inference models to recognize and classify the dialogue acts. We also follow up with discourse analysis, which is one of the important processes in the development of dialogue systems. Results of research in this direction allow us to revisit the dialogue systems, develop and deploy on a robot to demonstrate a proof of concept.**

***Index Terms*—natural language processing, human-robot interaction, dialogue systems, dialogue acts, discourse analysis**

I. Introduction

In a conversation, humans use changes in a dialogue to predict safety-critical situations and use it to react accordingly. We propose to use these kinds of cues for safer human-robot interaction through early detection of dangers. In the section below, you will find the list as a research progress from learning the linguistic feature to developing a dialogue system for the robot which can adapt their behavior based on linguistic features. The features learned using learning approaches such as artificial neural networks and deep learning.

II. Approaches

A. Sentiment Guided Language Learning

Sentiment can drive conversation based on their polarity. For example, being sentimentally positive in the language can bring positive utterances and vice versa. We attempt to model such a model to learn to estimate the sentiment of the next upcoming utterance based on a few preceding utterances [1]. Due to a low availability of sentiment annotated dialogue corpora, we use a sentiment classification for utterances, to learn sentiment changes within dialogues and ultimately predict the sentiment of upcoming utterances [2], [3].

We show that training a recurrent neural network on context sequences of words, defined by two preceding utterances of each speaker with the sentiment class of the next utterance, leads to useful predictions of the sentiment class of the upcoming utterance. See the example in Figure 1 to relate the safety learning process using sentiment as a guiding cue. We also explore the emotion intensity detection by using character- and word-level recurrent neural network models [4].

B. Dialogue Act Recognition

Dialogue act represents a functional importance of an utterance. It is an aspect of natural language understanding where its recognition plays an important role in building the dialogue systems (DS). We develop several neural models to learn to recognize and classify the dialogue acts. For the recognition of dialogue act, the context within the dialogue is very important, hence, modelling the neural models the same way is crucial [5]. We develop a recurrent neural model which uses a character level language model feature for each utterance. This model surpasses some of the state-of-the-art results on the Switchboard Dialogue Act corpus [6], [7].

However, we also attempt to answer the research question that how much context information is needed while recognizing the dialogue act of utterance [8]. Hence, we develop a similar neural model with attention mechanism on the top, which computes the weights of the contribution of preceding utterances while recognizing the dialogue act of current utterance [9]. The architecture uses a bi-directional recurrent neural network with attention mechanism [10].

R: Hello, how can I help you?	*Neutral*
P: Can you bring me tea?	*Neutral*
R: Yes, I can make some tea.	*Positive (context)*
P: Be careful, that cup seems broken.	*Neutral*
R: Shall I continue the action.	*Neutral*
P: No, don't use the broken cup.	*Negative (context)*
R: Okay, I will find another one.	*Neutral*

Fig. 1. Example for preparing the contexts

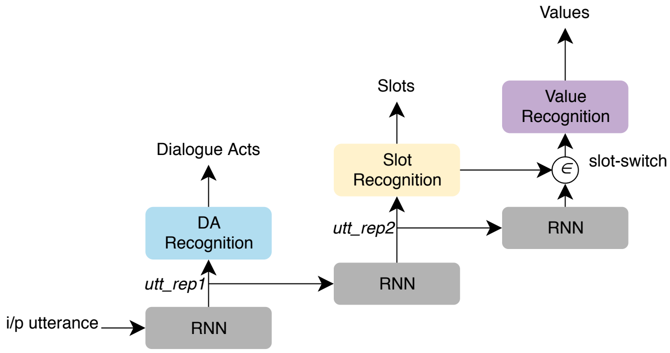


Fig. 2. Hierarchical recurrent neural networks for dialogue acts and slot-value pair recognition

C. Discourse Analysis

Discourse analysis can be performed by analyzing the dialogue act of sequence of utterances in a conversation. Our live web-demo called Discourse-Wizard is available[1] for discourse analysis. The backend used for this live web-demo is similar to our previous work at the dialogue act recognition, and more details can be found in [11].

D. Dialogue Systems (DS)

As a result, we aim to develop a dialogue system for the social robots which could take several linguistic features into account and infer accordingly. We developed a simple dialogue system which uses deep learning as a backend for spoken language understanding. As a first step, we develop a natural language interface for the simulated agent in AI2Thor environment [12]. The language understanding module is able to decode the input utterance into the symbolic representation using hierarchical recurrent neural networks as shown in Figure 2. For example, the utterance *"please move to the right"* can be decoded as $\{$*da : moveRobot, direction : right*$\}$ where *da* represents the dialogue act or intention, *direction* is a slot and *right* its value [13].

E. DS with Politeness as a Social Cue

Service robots need to show appropriate social behaviour in order to be deployed in social environments such as healthcare, education, retail, etc. [14]. Some of the main capabilities that robots should have are navigation and conversational skills [15], [16]. If the person is impatient, the person might want a robot to navigate faster and vice versa. Linguistic features that indicate politeness can provide social cues about a person's patient and impatient behaviour [17]. We developed the next part as a result of the dialogue system where we add another module like politeness detection, as shown in Figure 3. Dialogue act module is as same as spoken language understanding described previously. The response manager picks an appropriate response from the data file based on intention and the degree of politeness.

[1]https://secure-robots.eu/fellows/bothe/discourse-wizard-demo/ and full demo website at https://crbothe.github.io/discourse-wizard/

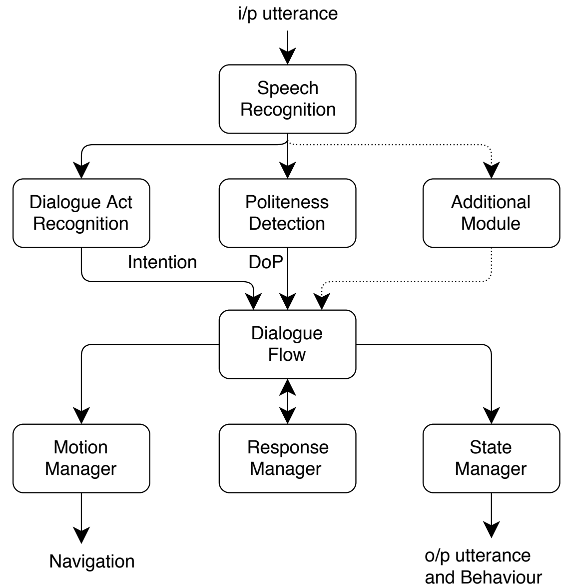


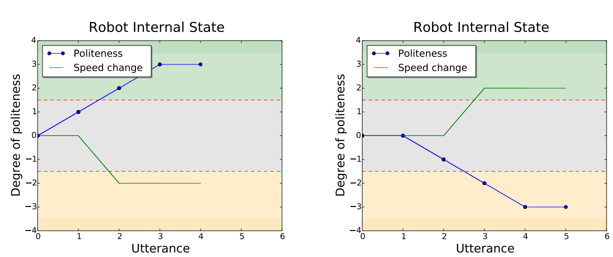
Fig. 3. Dialogue system with different modules

F. DS for Robot Adapting Behavior based on Politeness

As a proof of concept, we have developed and deployed our DS on the robot which adapts its behavior based on a degree of politeness [18]. It is demonstrated with practical experiment as a part of the project during secondment[2] [19]. DS communicated with the robot through state and motion managers for appropriate actions such as behavioral changes and navigation. We tested our system on the Pepper robot with different users, expressing different levels of politeness.

The behavioural changes and adaptation to speed change based on a change in DoP are shown in Figure 4. The robot behavioural adapts to the human being polite; the robot slows down and spends more time with the user. When the user is impolite, the robot speeds up and executes motion faster. The proposed behaviours of the robot for different situations, shown in Figure 5, are mainly to demonstrate the developed system and the efficacy of the proposed framework. The results indicate that the system is able to consider the

[2]We have accomplished this experiment as a secondment in collaboration with the industrial partner SoftBank Robotics in Paris, France with their semi-humanoid robot Pepper.



(a) (b)

Fig. 4. Robot internal state for polite (a) and impolite (b) interactions.

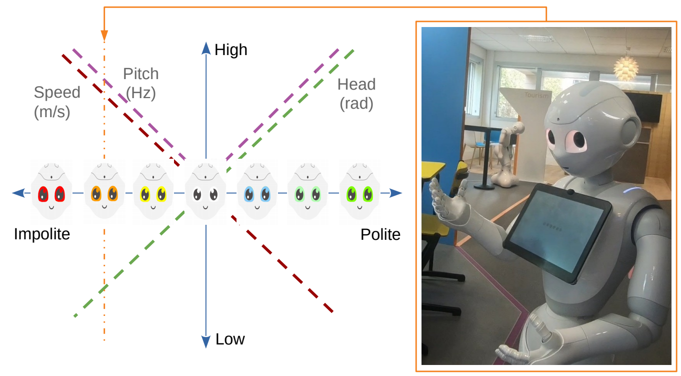


Fig. 5. The behavioural model used to create the verbal and non-verbal responses based on the cumulative sum of the DoP. The Pepper robot shown in the right is in the position of the vertical orange line in the plot during the interaction.

linguistic features to modulate the navigation behaviour of the robot in a coherent theoretical and functional framework. As aforementioned, to the best of our knowledge such a framework and implementation in a practical situation is one of the first attempts of its kind. However, it is important to mention that the validation of the hypotheses about the most appropriate behaviours of the robot is not within the scope of this paper and it will require further investigation and user studies.

III. Conclusion

We discussed most of the stages of progress in our research in the direction of the language learning for safety during human-robot interaction. We gave the pointers to deal with the language processing for dialogue system development and integration of those with the robot that shall accordingly adapt its behavior. The experimental frameworks open up a new challenge for the study of the effect of politeness or any socio-linguistic features in human-robot social interaction.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 framework programme for research and innovation under the Marie Sklodowska-Curie Grant Agreement No. 642667 (SECURE).

References

[1] C. Bothe, S. Magg, C. Weber, and S. Wermter, "Dialogue-based neural learning to estimate the sentiment of a next upcoming utterance," in *Artificial Neural Networks and Machine Learning – ICANN 2017*, A. Lintas, S. Rovetta, P. F. Verschure, and A. E. Villa, Eds. Cham: Springer International Publishing, 2017, pp. 477–485.

[2] C. Danescu-Niculescu-Mizil and L. Lee, "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs." in *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL*, 2011.

[3] B. MacWhinney, *The CHILDES project: Tools for analyzing talk.* Lawrence Erlbaum Associates, Inc, 1991. [Online]. Available: http://childes.psy.cmu.edu/

[4] E. Lakomkin, C. Bothe, and S. Wermter, "GradAscent at EmoInt-2017: Character and Word Level Recurrent Neural Network Models for Tweet Emotion Intensity Detection," in *Proc. of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis at the Conference EMNLP*. ACL, 2017, pp. 169–174.

[5] C. Bothe, C. Weber, S. Magg, and S. Wermter, "A Context-based Approach for Dialogue Act Recognition using Simple Recurrent Neural Networks," in *Proc. of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ERLA), 2018, pp. 1952–1957.

[6] J. J. Godfrey, E. C. Holliman, and J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, 1992, pp. 517–520.

[7] D. Jurafsky, "Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, draft 13," *Technical Report 97-01, University of Colorado Institute of Cognitive Science*, pp. 225–233, 1997.

[8] M. Sbisà, "Speech acts in context," *Language & Communication*, vol. 22, no. 4, pp. 421–436, 2002.

[9] C. Bothe, S. Magg, C. Weber, and S. Wermter, "Conversational Analysis using Utterance-level Attention-based Bidirectional Recurrent Neural Networks," in *Proc. of the International Conference INTERSPEECH 2018*, 2018.

[10] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based Bidirectional Long Short-term Memory Networks for Relation Classification," in *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 207–212.

[11] C. Bothe, S. Magg, C. Weber, and S. Wermter, "Discourse-Wizard: Discovering Deep Discourse Structure in your Conversation with RNNs," *preprint arXiv:1806.11420*, 2018.

[12] ——, "Natural Language Interface to Control an AI2Thor Simulator Agent using a Deep Learning Backend," in *[submitted to] Proc. of the International Conference IROS 2018: Workshop – Towards Intelligent Social Robots*, 2018.

[13] S. Ultes, L. M. Rojas Barahona, P.-H. Su, D. Vandyke, D. Kim, I. Casanueva, P. Budzianowski, N. Mrkšić, T.-H. Wen, M. Gasic, and S. Young, "PyDial: A Multi-domain Statistical Dialogue System Toolkit," in *Proceedings of ACL 2017, System Demonstrations*. Vancouver, Canada: Association for Computational Linguistics, July 2017, pp. 73–78.

[14] C. Adam, W. Johal, D. Pellier, H. Fiorino, and S. Pesty, "Social Human-Robot Interaction: A New Cognitive and Affective Interaction-Oriented Architecture," in *International Conference on Social Robotics*. Springer, 2016, pp. 253–263.

[15] M. Salem, M. Ziadee, and M. Sakr, "Marhaba, how may I help you?: Effects of Politeness and Culture on Robot Acceptance and Anthropomorphization," in *International Conference on Human-robot Interaction*, 2014, pp. 74–81.

[16] P. S. Rogers and S. M. Lee-Wong, "Reconceptualizing Politeness to Accommodate Dynamic Tensions in Subordinate-to-Superior Reporting," *Journal of Business and Technical Communication*, vol. 17, no. 4, pp. 379–412, 2003.

[17] C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts, "A computational approach to politeness with application to social factors," in *Proceedings of ACL 2013 (Volume 1: Long Papers)*, 2013, pp. 250–259.

[18] A. Beck, L. Cañamero, and K. A. Bard, "Towards an Affect Space for robots to display emotional body language ," in *19th International Symposium in Robot and Human Interactive Communication*, 2010, pp. 464–469.

[19] C. Bothe, F. Garcia, A. Cruz-Maya, A. K. Pandey, and S. Wermter, "Towards dialogue-based navigation with multivariate adaptation driven by intention and politeness for social robots," in *International Conference on Social Robotics*, 2018.