In: *Proceeding of the International PhD Conference on Safe and Social Robotics (SSR-2018)*, pp. 31-33, 29-30 September 2018, Madrid, Spain.

Progress Report: Language-modulated Actions using Deep Reinforcement Learning for Safer Human-Robot Interaction

Mohammad Ali Zamani[1], Sven Magg[1], Cornelius Weber[1] and Stefan Wermter[1]

***Abstract*— Spoken language can be an efficient and intuitive way to warn robots about threats. Guidance and warnings from a human can be used to inform and modulate a robot's actions. An open research question is how the instructions and warnings can be integrated in the planning of the robot to improve safety. Our goal is to address this problem by defining a Deep Reinforcement Learning (DRL) agent to determine the intention of a given spoken instruction, especially in a domestic task, and generate a high-level sequence of actions to fulfill the given instruction. The DRL agent will combine vision and language to create a multi-modal state representation of the environment. We will also focus on how warnings can be used to shape the DRL's reward, concentrating on the recognition of the emotional state of the human in an interaction with the robot. Finally, we will use language instructions to determine a safe operational space for the robot.**

I. INTRODUCTION

In the future, robots are expected to work as companions with humans in various areas including domestic scenarios such as care-giving. Human-robot interaction safety has not been well studied [1]. Even with well-engineered robots, it would be unrealistic to move robots directly from factories to home environments to perform complex tasks [2] [3] due to safety [4]. Moreover, robots also have to continuously adapt to new environments to avoid hazardous actions since using experts to program a robot for every environment is impossible. Hence, we need adaptive learning algorithms.

Spoken language can be considered one of the most effective communication channels to warn robots about threats. For example, robots may not notice an external threat or mis-planning that may harm a human or the robot itself. However, a human can warn or guide the robot by a verbal utterance toward a safer interaction. How robots react to safety warnings is not addressed exhaustively in the literature. The closest related research area is assigning tasks to robots by verbal instructions [5], [6], [7]. They follow rule-based methods to utilize spoken language instructions which can cover only a limited number of scenarios.

Our goal is to train a robot to safely perform complex tasks with the ability of processing environmental feedback, including guidance and warnings by a human, to shape a proper signal for updating its own policy. Therefore, our research is focused on three capabilities of the robot: generating high-level actions from verbal instructions, extracting reward from prosodic/sentiment features of the human speaker, and learning a safe workspace for the robot.

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

[1]Knowledge Technology, Department of Informatics, University of Hamburg, Germany. zamani, magg, weber, wermter @informatik.uni-hamburg.de

II. FOCUS AREAS

A. Mapping Spoken Instruction to a Sequence of Actions

We introduced a framework to obtain the intention of a given spoken instruction (e.g. "boil water") and generate the sequence of actions ("moveto kettle", "grasp kettle", ...) to fulfill the task [8], [9]. The intention detection was implemented with a 2 layer perceptron with 20% dropout and trained by the TellMeDave corpus to predict one of 10 predefined classes. Our model could achieve 89.57% accuracy in a 5-fold cross-validation.

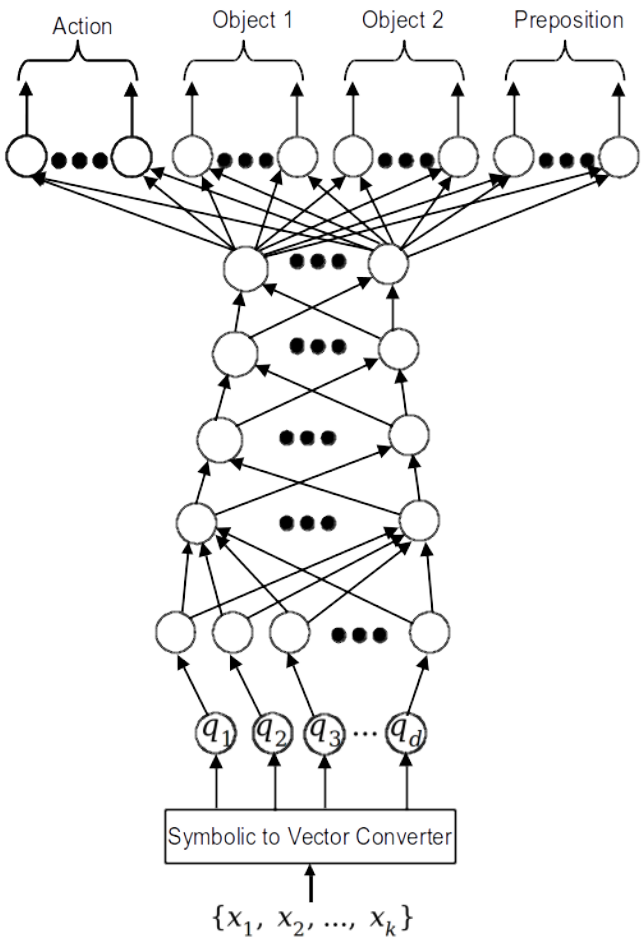

Fig. 1. The deep reinforcement learning architecture generates the sequence of actions. An MLP neural network is trained to approximate the action-value functions. The compositional linguistic state, $\chi(t)$, is presented to the network as a compositional vector which is a binary vector.

We developed a symbolics environment from the "Tell Me Dave" Corpus [10] to train the RL agent. The main contribution was to use a distributed symbolic state representation (e.g. {On Kettle Sink}, {Near Robot Sink}, ...) which reduced the learning time on given tasks. Our Reinforcement Learning was built based on the Deep Q-Network [11] architecture with modifications to support multiple Q functions and different types of value estimation. As shown in figure 1, there are four output groups in the network architecture. However, the actions in the corpus have different number of arguments (i.e. object1, object2, and preposition) from zero to three. Therefore, we masked the gradient based on the performed action.

In our case, the environment state was directly accessible through the simulation while this needs to be extracted in a real life scenario. Therefore, we will extend by encoding vision and instruction in a fused state similar to Shu et al. [12] in a more realistic simulator like AI2Thor [13] (see figure 2).

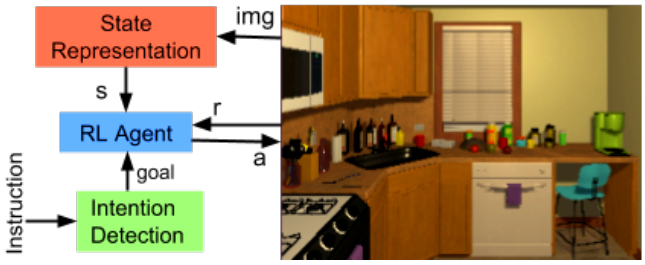

Fig. 2. The modular approach using intention detection and reinforcement learning trained for each objective to generate the sequence of actions [9].

B. Extracting Reward from the Human Speech

The robot needs to continuously process human speech to detect implicit interruptions or any change in the instruction. The robot is expected to be able to stop (both soft and emergency) with a minimum latency in an unsafe situation (see figure 4). We developed a reinforcement learning approach to optimize the accuracy and latency concurrently [14].

The model (see figure 3) was consist of recurrent neural network with Gated Recurrent Units [15] which learns a temporal representation from the extracted features of speech. The Emotion classification module (θ_c) used the GRU's output to determine emotion as angry or neutral. The action selection (θ_a) which is Monte Carlo Policy Gradient (or REINFORCE) [16] decides to either wait for the next speech frame or terminate the processing and read the emotion classification module. We also used the baseline estimation (θ_b) to estimate a baseline reward. Similar to [17], [18], this helps to lower the variance of the gradient signal.

As a result, our model achieved about 50% latency reduction with the same level of accuracy evaluated on the iCub recorded data in our lab. We also improved the robustness of emotion recognition by proposing data augmentation techniques like overlaying background noise [19].

As future work, emotion recognition will be used to filter warnings and to record this experience in the RL's memory

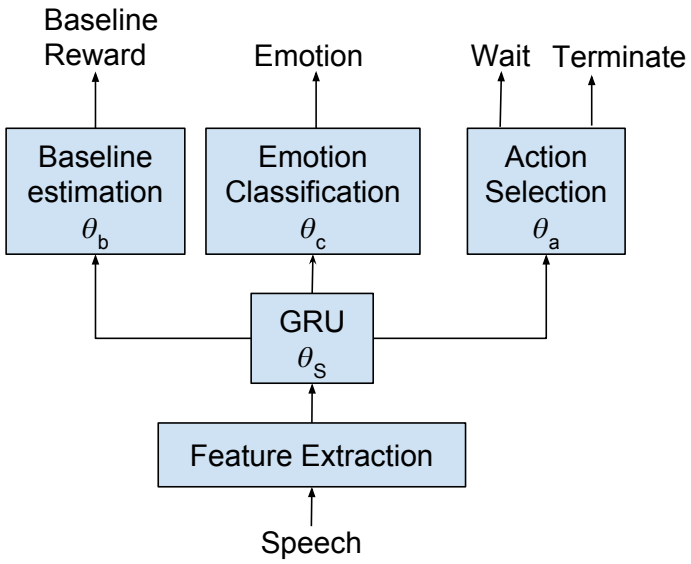

Fig. 3. The EmoRL model consists of 4 components: *Gated Recurrent Unit (GRU)*, *Emotion Classification (EC)*, *Action Selection (AS)* and *Baseline Reward Estimator (BRE)*. The GRU encodes the acoustic information of a speech signal which is used as a state representation. *EC* uses the state representation to evaluate the probability of the human speaker being in an angry state. *AS* and *BRE* determine the probability distribution over possible actions and the estimation of the baseline reward [14].

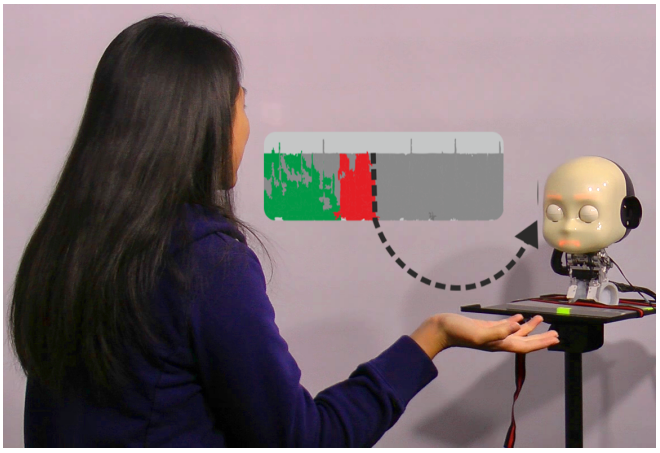

Fig. 4. Extracting reward from the human speaker. The robot analyzes continuously arriving acoustic input and only when it has enough information to evaluate the affective state of the speaker it will output the person's specific emotion. The robot is trained using reinforcement learning to make the dynamic decision: wait for more data or trigger a response [14].

for updating the agent's policy. We will use a pre-trained model in simulation to focus on learning new safety cases in the real scenario.

C. Safe Human-Robot Collaboration in Manual Tasks

Safety becomes more important when humans work with robots collaboratively. For shaping such a collaborative scenario incrementally, as an initial step, we improved the learning of the Deep Deterministic Policy Gradient (DDPG) [20] in a reach-for-grasp task by introducing an adaptive (larger-than-life) augmented target [21]. Later, we used it to train a 2-DOF arm in an interactive scenario to reach multiple target points which improved the learning time by solving the

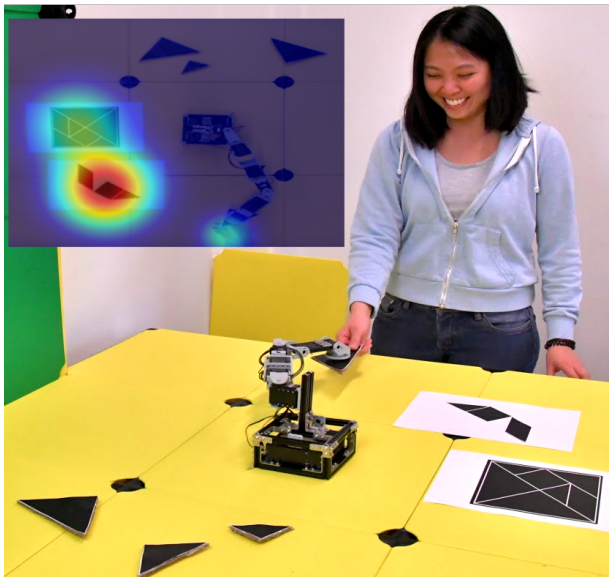

Fig. 5. A person is solving a tangram puzzle in collaboration with a robot arm. The robot arm is instructed to avoid the person's workspace while fetching puzzle pieces from the far end of the table. The top right image shows the top view overlaid with a spatial representation which can be learned by interaction with the user. The robot plans its motion incorporating the adaptive spatial constraints [23].

problem in simulation and deploying it on the robot when it gained enough confidence [22]. In a preliminary experiment (see figure 5), we demonstrated how spoken instructions can be mapped to a spatial representation of the robot's workspace which can be used as constraints for the path planner [23]. As a next step, we are also interested to learn grasping with verbally described spatial constraints in an end-to-end approach.

III. CONCLUSIONS

In this PhD project, spoken instructions are used in different areas and we focused on the high level action sequences for performing tasks in a domestic scenario. As a next step, we will concentrate on obtaining state representations in real life scenarios. In parallel, we proposed a model to detect angry emotions rapidly, which can be used as an implicit interruption to planning to lead to a safer human-robot interaction. As future work, we will focus on how the robot can learn from experience to immediately avoid the same behavior. We also investigate how teaching the operational space to the robot can be performed intuitively. We plan to extend this to a small kitchen scenario which can bring together all these ways of using spoken instructions/warnings to guide the robot towards safer interaction.

REFERENCES

[1] M. Vasic and A. Billard, "Safety issues in human-robot interactions," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 197–204.

[2] S. Schaal, "The new robotics–towards human-centered machines," *HFSP journal*, vol. 1, no. 2, pp. 115–126, 2007.

[3] S. Schaal and C. G. Atkeson, "Learning control in robotics," *IEEE Robotics & Automation Magazine*, vol. 17, no. 2, pp. 20–29, 2010.

[4] J. Peters and S. Schaal, "Learning to control in operational space," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 197–212, 2008.

[5] S. Lauria, G. Bugmann, T. Kyriacou, J. Bos, and E. Klein, "Converting natural language route instructions into robot executable procedures," in *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*. IEEE, 2002, pp. 223–228.

[6] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein, "Mobile robot programming using natural language," *Robotics and Autonomous Systems*, vol. 38, no. 3, pp. 171–181, 2002.

[7] T. Nishizawa, K. Kishita, Y. Takano, Y. Fujita *et al.*, "Proposed system of unlocking potentially hazardous function of robot based on verbal communication," in *System Integration (SII), 2011 IEEE/SICE International Symposium on*. IEEE, 2011, pp. 1208–1213.

[8] M. A. Zamani, S. Magg, C. Weber, and S. Wermter, "Deep reinforcement learning using symbolic representation for performing spoken language instructions." *In 2nd Workshop on Behavior Adaptation, Interaction and Learning for Assistive Robotics (BAILAR) on Robot and Human Interactive Communication (RO-MAN), 26th IEEE International Symposium on.*, 2017.

[9] ——, "Deep reinforcement learning using compositional representations for performing instructions." *Submitted to the Paladyn Journal of Behavioral Robotics*, 2018.

[10] D. K. Misra, J. Sung, K. Lee, and A. Saxena, "Tell Me Dave: Context-sensitive grounding of natural language to manipulation instructions," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 281–300, 2016.

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[12] T. Shu, C. Xiong, and R. Socher, "Hierarchical and interpretable skill acquisition in multi-task reinforcement learning," *arXiv preprint arXiv:1712.07294*, 2017.

[13] E. Kolve, R. Mottaghi, D. Gordon, Y. Zhu, A. Gupta, and A. Farhadi, "Ai2-thor: An interactive 3d environment for visual ai," *arXiv preprint arXiv:1712.05474*, 2017.

[14] E. Lakomkin, M. A. Zamani, C. Weber, S. Magg, and S. Wermter, "Emorl: Continuous acoustic emotion classification using deep reinforcement learning," *accepted at the International Conference on Robotics and Automation (ICRA)*, 2018.

[15] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *ICLR*, 2015.

[16] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine learning*, vol. 8, no. 3-4, pp. 229–256, 1992.

[17] V. Mnih, N. Heess, A. Graves *et al.*, "Recurrent models of visual attention," in *Advances in neural information processing systems*, 2014, pp. 2204–2212.

[18] J. Gu, G. Neubig, K. Cho, and V. O. K. Li, "Learning to translate in real-time with neural machine translation," in *15th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics (ACL), 2017.

[19] E. Lakomkin, M. A. Zamani, C. Weber, S. Magg, and S. Wermter, "On the robustness of speech emotion recognition for human-robot interaction with deep neural networks," *accepted to the International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[20] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, pp. 1–14, 2015.

[21] M. Kerzel, H. Beik Mohammadi, M. A. Zamani, and S. Wermter, "Accelerating deep continuous reinforcement learning through task simplification," *accepted at the International Joint Conference on Neural Networks (IJCNN)*, 2018.

[22] H. Beik Mohammadi, M. A. Zamani, M. Kerzel, and S. Wermter, "Online continuous deep reinforcement learning for a reach-to-grasp task in a mixed-reality environment," *to be submitted*, 2018.

[23] M. A. Zamani, H. Beik Mohammadi, M. Kerzel, S. Magg, and S. Wermter, "Learning Spatial Representation for Safe Human-Robot Collaboration in Joint Manual Tasks," *accepted in WORKMATE 2018: the WORKplace is better with intelligent, collaborative, robot MATEs on International Conference on Robotics and Automation (ICRA).*, 2018.