Generalizing Multimodal Policies Learned in a Human-Robot Interaction

Lorenzo Cazzoli[1], Jacqueline Hemminghaus[2], Stefan Kopp[2], and Mauro Gaspari[1]

[1]Department of Computer Science, University of Bologna, Bologna, Italy
[2]CITEC, Bielefeld University, 33619 Bielefeld, Germany

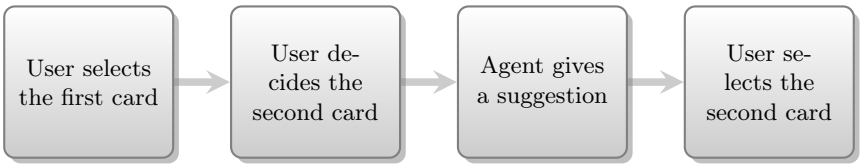
Human-Robot-Interaction is an increasingly studied field with many disciplines involved. The different categories of targets influence the type of challenges that need to be accomplished. For instance, robot platforms designed for children have to adapt to their maturing behavior and communication abilities. On the other hand, platforms for elderly people have to accommodate to the digressive visual and hearing capacities. The robots need to adapt their behavior to the capabilities and deficits of their interaction partners. It is expected that if the user has some visual lack the robot will use more voice suggestion and in the opposite case, with lacking hearing abilities, more facials expressions or gestures are required.

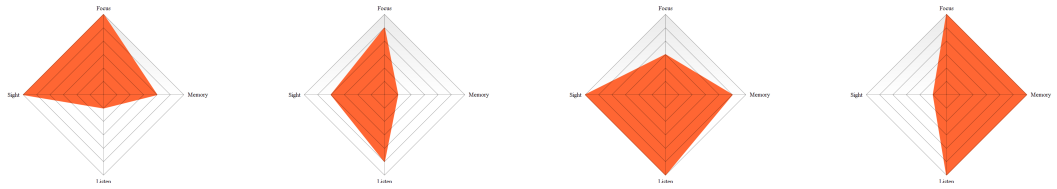
Using Reinforcement Learning (Sutton and Barto [2011]) robots can easily learn the correct behavior for each group of users independently from their sensorial differences. The problem is that information learned interacting with one user can't be used in a learning process of another user because they presumably have different capabilities, thus it is not guaranteed that two actions are perceived in the same way and cause the same results. The learned policy allows the robot to infer a correct behavior only for the current user. But when the robot interacts with a new user it has to update its policy to adapt to the new situation. If the two users are not too different, the robot is able generate a new reliable behavior policy. Otherwise the generated behavior could be misleading or frightening due to an inaccurate policy.

Combining policies we can overcome this problem guaranteeing that the robot behaves properly, with a sub-optimal solution, with all encountered users and without the need to learn a policy for the current user from scratch again. This fact is even more important when learning online while interacting with a human. In this case the combined policy is used to initialize the learning process, to speed it up and to limit initial wrong suggestions because of exploration.

In this abstract we present a way of combining previously learned robot behavior policies of different users. The main idea is to combine a set of policies, in tabular representation, into a final sub-optimal solution for the problem all users have contributed to. We assume that the features/differences of users are unknown and need to be extracted from the different policies generated from same user. This information is used to weight the importance of a set of actions to sum up two policies.

As a starting point policies were collected from simulated users that complete the matched pairs memory game, assisted by the robot. Each time the user looks for pairs on the game board, the robot, knowing the position of the matching card, gives suggestions helping the user to finish the game in minimum number of turns. To enable behavior adaptation of the robot, it is equipped with Q-learning and ϵ-greedy to handle exploration and exploitation (Watkins and Dayan [1992]). The robot communicates through multimodal actions, including gaze, speech, head movements and facial expressions.





(a) Deaf (b) Old (c) Teenager (d) Blind

Figure 1: The charts show the features of four different personas. The features are measured on a scale 1 to 6. A value of 1 means that the user completely lacks this sense or is severely impaired. On the opposite side a value of 6 represents full perception of the sense. The labels on the corners are in clockwise order focus, memory, listen, sight beginning at the top.

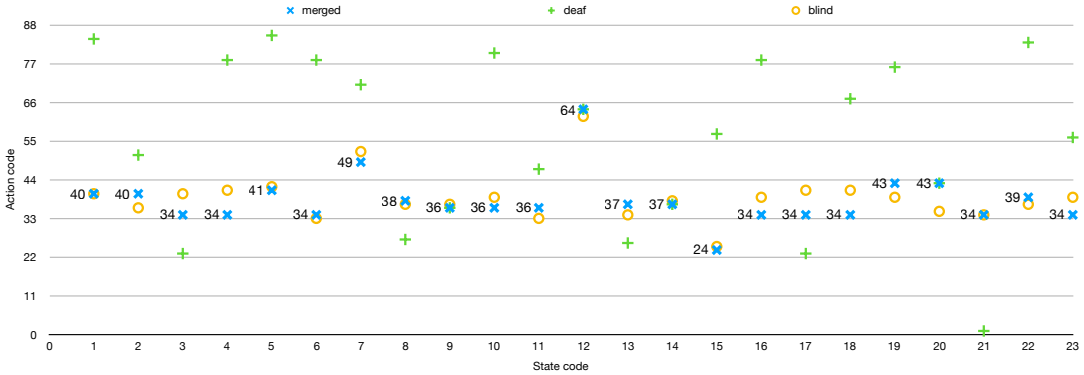


Figure 2: The graphic represents three policies. The green plus and the yellow circles represent respectively the deaf and blind user policy. The blue cross represents the policy gained from combining both. The axes represent the index number of the action/state in an array representation.

A set of personas (Schulz and Fuglerud [2012]) is generated to cover different groups of users and to raise awareness about users' needs. For each of these personas a set of features is extracted from their stories. For example an old man will have less memory than a teenager even if he could have an higher level of attention. These features are focus, memory, listen, sight associated to a value between 1 to 6 creating a chart (see Figure 1) that defines the simulated user. The actions taken by the user are based on those attitudes and have a high impact on the learning of the agent. For example a deaf user will ignore all speech suggestions while he percepts and follows all other multimodal parts of the actions.

Analyzing the individual policies for each simulated user groups, the impact of their different personality traits on the learning process were clearly visible. For example, users representing personas with a good memory are performing better in terms of number of actions per game while the robot is still giving wrong suggestions.

After combining different policies it emerges that the combined policy works as a general policy suitable for all users, as it always selects actions that are satisfying the users at the border of the defined sensorial possibilities. In Figure 2 a graphic representation of three policies is given. One of them, the blue cross, is the result of combining the actions of the other two policies from a deaf and a blind user. This combined policy converges to a set of actions where the most powerful and perceptible suggestions for each modality are given. The combined policy and the policy of the blind are similar because they use the same speech suggestions but with different head movements that work out for the deaf as well. In this way it is guaranteed that all users receive effective help.

Our first results show that it is possible to combine and generalize learned behavior policies for different user groups. Our next step will be to train and test the presented approach to generate generalized multimodal robot behavior in a real user study, whose results will be presented at the symposium.

References

Trenton Schulz and Kristin Skeide Fuglerud. Creating personas with disabilities. In *International Conference on Computers for Handicapped Persons*, pages 145–152. Springer, 2012.

Richard S Sutton and Andrew G Barto. Reinforcement learning: An introduction. 2011.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.