Poster at the IEEE International Conference on Developmental Learning (ICDL), Prague, 2025.

Qualitatively Guided Training of Skills

Jan Lemeire[1], Ruben Spolmink[1], Domen Šoberl[2], Jure Žabkar[3] and Marco Van Cleemput[1]

***Abstract*— We propose using qualitative causal models to steer an autonomous agent's learning process of skills. Embodied learning of motor skills is traditionally approached using reinforcement learning. While reinforcement learning has achieved several notable results since the emergence of deep learning, it still requires a substantial amount of data and repetitions compared to human learning.**

Qualitative models describing the causal relations between the variables can guide embodied learning [1]. We apply our approach to the learning of skills: throwing a ball into a basket, parking a car backwards, and reaching towards an object. For each learning task, first, a qualitative causal model is learned from numerical data; in the second stage, qualitative inference guides the tuning of the quantitative action parameters. Trials quickly converge: after 5 to 15 trials, skills are successfully. We compare our approach to a handful of traditional methods and show that qualitative constraints substantially reduce learning complexity. The proposed approach provides a hypothesis of how humans learn skills: qualitative reasoning happens consciously, while quantitative tuning happens subconsciously.

I. Overview

Three skills were studied: throwing a ball into a basket, parallel parking a car, and reaching with a coordinated move of multiple joints. For each of these skills, a set of action parameters must be tuned to achieve the goal. We propose a training procedure that is steered by a qualitative model that represents the effect of changes made to the values of action parameters. Cause-effect relations are modeled qualitatively by the *direction of change*. This allows qualitative reasoning: 'throwing harder will bring the ball higher', or 'turning with a larger angle brings the car deeper into the parking spot'. The qualitative relations can be learned efficiently and iteratively learned from data [2]. Context is added when relations appear to be incorrect in certain parts of the state space. For example, throwing above or below an angle of 45 degrees results in relationships with different signs.

Figure 1 shows the approach that steers the training procedure with a qualitative model. A set of action parameters determines the action's execution and the end state. A skill is a policy that maps the goal state onto effective values for the action parameters. This *quantitative model* is learned by an iterative trial-and-error training process that is guided by a *causal qualitative model*. An unsuccessful trial results in a state at a certain distance from the goal state, represented by the error vector. To improve the next trial, the direction in which the action parameters should be changed is inferred from the qualitative model by backpropagating the signs of the errors.

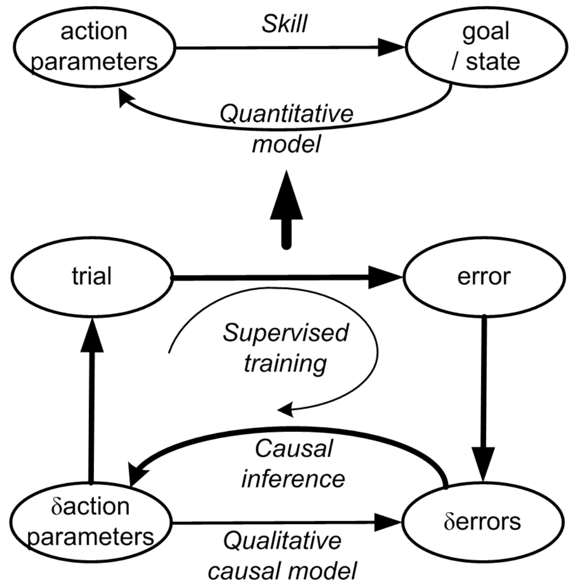

Fig. 1. Qualitative supervised training.

[1]Jan Lemeire, Ruben Spolmink and Marco Van Cleemput are with the Dept. of Industrial Sciences (INDI) and the Dept. of Electronics and Informatics (ETRO) of the Vrije Universiteit Brussel (VUB) Pleinlaan 2, B-1050 Brussels. `jan.lemeire@vub.be`

[2]Domen Šoberl is with the the Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Primorska, Slovenia. `domen.soberl@famnit.upr.si`

[3]Jure Žabkar is with the Faculty of Computer and Information Scienceof the University of Ljubljana, Ljubljana, Slovenia. `jure.zabkar@fri.uni-lj.si`

II. Results

During experimentation, it took the causal model 5 to 15 trials to reach a stable throwing action in comparison to traditional reinforcement learning, where a soft actor-critic algorithm requires around 100,000 samples to reach a similar stable throwing action. However, as the causal model is only capable of adjusting the throw, it is more akin to stochastic gradient descent. During testing, we found that COBYLA[3] gave the best performance with a mean of 229.1 trials with a standard deviation of 237.3.

References

[1] T. Wiley, C. Sammut, B. Hengst, and I. Bratko, "A Planning and Learning Hierarchy using Qualitative Reasoning for the Online Acquisition of Robotic Behaviors," *Advances in Cognitive Systems*, 2016.

[2] J. Žabkar, M. Možina, I. Bratko, and J. Demšar, "Learning qualitative models from numerical data," *Artificial Intelligence*, 2011.

[3] M. J. Powell, *A direct search optimization method that models the objective and constraint functions by linear interpolation*. Springer, 1994.