Experimental Verification of Learning Strategy Fusion for Varying Environments

Akihiko Yamaguchi Masahiro Oshita Jun Takamatsu Tsukasa Ogasawara

Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara 630-0192, JAPAN
{akihiko-y,ogasawar}@is.naist.jp

ABSTRACT

We investigate a real robot applicability of our method, general-purpose behavior-learning for high degree-of-freedom robots in *varying environments*. Our method is based on the learning strategy fusion proposed in [3], and extended theoretically in [4]. This report discusses its applicability to real robot systems, and demonstrates some positive experimental results.

Keywords

Robot learning, learning strategy, crawling

1. INTRODUCTION

In near future, robots will be used in human daily life where they are engaged in household chores, supporting and taking care of humans. In contrast to manufacturing robots, such robots are required to be more flexible; they need to adapt to their owners' requests. One difficult problem is how the robot finds a policy to achieve a given task. We are taking reinforcement learning approach which is more general than planning since it is applicable even if a model of the system is unknown. Generally to say, it is hard to define a system model when a robot is working with people. Examples of reinforcement learning researches are [1, 2]. However, successful results for complicated systems, especially high degree-of-freedom (DoF) systems like humanoid robots, are obtained only when combining with imitation learning.

In contrast, we are tackling to solve the policy learning problem in a *learning-from-scratch* case, where no prior information, e.g. a demonstration trajectory, is given to the robot. The robot needs to find a policy that maximizes the sum of *rewards* encoding task objective.

Especially, we are developing a method for the case where a robot works in varying environments. So far, we proposed a method with which (1) a robot can learn a policy quickly in an environment even if the robot has high DoF, (2) when the robot does the same task in a different environment, it can quickly adapt to the new environment, (3) the robot preserves the policies for each environment, and recalls a suitable policy for the current environment. The proposed method for (1) is named *Learning Strategy (LS) fusion* [3], and it is extended to achieve (2) and (3) in [4]. An advantage of our method is that it does not use visual or geometrical information to recognize an environment, but estimates the environment by executing learned policies and matching the observed rewards. Thus, our method is more generally applicable. However, we have applied it only for a simulated robot [4].

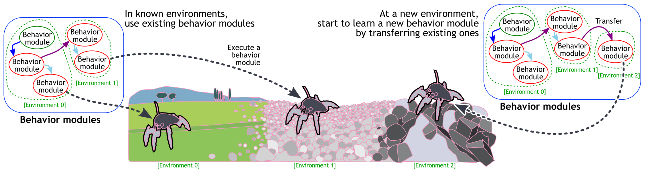

Figure 1: Illustration of how the proposed method works. Each behavior module corresponds a policy.

In this report, in order to show its wider capability, we investigate the method with the real robots' task. The experimental results show that our method works as well as in the simulation. We believe that these results are beneficial for many researchers to consider applying our method to many robots' tasks, including human robot interaction (HRI) scenarios. The uncertainty of the dynamics increases when humans are involved in a system, where a change of user (imagine to share a robot with family) is considered as a change of environment. Our method has a capability to treat such a situation.

2. LEARNING STRATEGY FUSION

LS fusion is a general architecture to combining *learning strategies*; each learning strategy is a way to generate a policy. There are different variations of such strategies; the most simple one is generating a policy for random exploration when there is no policy for a task. *Transferring* strategies are important for efficient learning; sometimes humans start to learn a policy with slow and restricted (reduced DoF) movements, then incrementally increase the speed and DoF of movements. LS fusion is a meta-framework to fuse such learning strategies. It has a wide applicability, but the most successful results are obtained when combining three learning strategies: learning from scratch, accel-

HRI'15 Extended Abstracts, March 2–5, 2015, Portland, OR, USA.
ACM 978-1-4503-3318-4/15/03.
http://dx.doi.org/10.1145/2701973.2702056.

erating a movement, and freeing (increasing) the DoF of a movement.

The reason why we chose the LS fusion as a basis of the learning method for varying environments is that since the best policy for an environment is too specialized to that environment, a policy in middle-stage of learning sometimes has a better generalization ability. LS fusion preserves the past policies in order to get back to them when a transferring strategy does not work. Thus, LS fusion is suitable for the extension to varying environments.

The major problem is how to model and estimate an environment. Rather than using visual or geometrical information of an environment, our approach is using learned policies to *test* the current environment, then check the correspondence between the learned environment and the current environment from rewards observation. The reason of this approach is that the performance (measured as a sum of rewards) is only the index to categorize the environments. If the same policy works well in visually or geometrically different environments, these environments should be categorized into the same class. Thus in our method, a probabilistic model of the sum of rewards conditioned by a policy and a environment class is also learned during learning a policy. This model tells how the current environment is close to the learned one. Thus, executing *test* with learned policies several times, the robot can estimate the class of the current environment; of course, it can be a new class. Fig. 1 illustrates how the proposed method works.

3. EXPERIMENTS

We investigate the performance of the LS fusion for multiple environments with a real robot task. We employ a spider robot with 6 legs and engage it in a crawling task. We setup three different environments as shown in Fig. 2. The reward of the crawling task is proportional to the forward speed of the robot; thus, the robot will obtain a crawling motion by maximizing the sum of rewards. A basic unit of learning is an *episode* where the robot starts moving from an initial pose, and finishes after 50 sec or at some troubles. During each episode, a policy is updated by Q(λ)-learning at each action. Policy generation and environment estimation of LS fusion are executed at the beginning of each episode.

Each trial (run) consists of three learning stages and a test stage. In 1st stage, the robot starts to learn from scratch in the *plain* environment; i.e. the robot does not have any policies. In this stage, the proposed method works as same as the LS fusion for a single environment. After learning ends, the robot is put on the *rough* environment (2nd stage), and starts learning with the result of 1st stage. Similarly, in the 3rd stage, the robot learns on the *slip* environment. In the 2nd and the 3rd stages, the robot is expected to use the learned policies as the starting points. In the 2nd and the 3rd stages, we set the current environment class as *unknown* to enforce the environment estimation. In the test stage, the last result of the 3rd stage is executed in each terrain. Here, the robot is expected to choose a suitable policy for the environment.

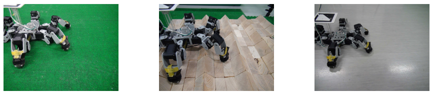
Carpet floor(plain) Rough terrain(rough) Slippery floor(slip)

Figure 2: Terrains.

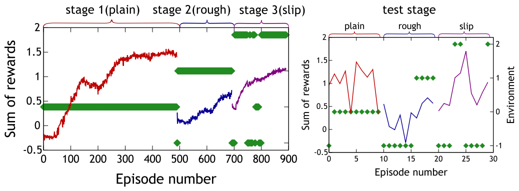

Figure 3: Result–1st run: learning curve (sum of rewards), environment estimations (diamond points).

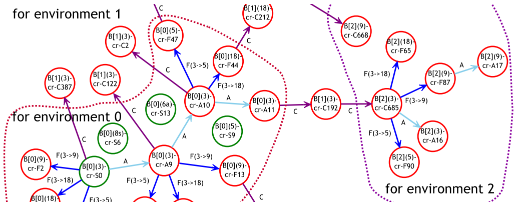

Figure 4: Part of *fusion tree* obtained in the 1st run.

Fig. 3 shows the results of learning curves and estimated environment classes. Compare the beginning of the 1st stage and the 2nd/3rd stage; the sum of rewards of 2nd/3rd stage is greater. This is because the robot uses policies learned in the previous stage. In 2nd and 3rd stages, the robot could estimate the environment classes almost correctly. Fig. 4 shows the automatically generated behavior modules. As we illustrated in Fig. 1, the policies were obtained as we expected. Video is here: `http://youtu.be/h3mzPsEnYYc`

4. CONCLUSION

In this report, we investigated the real robot applicability of learning strategy fusion extended to varying environments [3, 4]. Since our method does not use visual or geometrical information to recognize an environment, we could apply it to a real robot system with minimum number of sensors. This advantage will be also beneficial when applying to HRI scenarios, because this method can treat a change of human as a change of environment. We need further investigation of its applicability to HRI scenarios.

Acknowledgments

Part of this work was supported by Artificial Intelligence Research Promotion Foundation of Japan.

5. REFERENCES

[1] J. Kober, et al. Reinforcement learning to adjust parametrized motor primitives to new situations. *Autonomous Robots*, 33:361–379, 2012.

[2] P. Kormushev, et al. Robot motor skill coordination with EM-based reinforcement learning. In *IROS'10*, pages 3232–3237, 2010.

[3] A. Yamaguchi, et al. Learning strategy fusion to acquire dynamic motion. In *Humanoids'11*, pages 247–254, 2011.

[4] A. Yamaguchi, et al. Learning strategy fusion for acquiring crawling behavior in multiple environments. In *ROBIO'13*, pages 605–612, 2013.