論文の内容の要旨

Policy Search and Model Learning in Reinforcement Learning via Evolutionary Computation
（進化計算を用いた強化学習における政策探索とモデル学習）

氏名　森口　博貴

Autonomous motion planning and control is one of the fundamental problems in robotics. There have been huge demands on robot's application to challenging problems in human society such as manufacture, disaster-relief, environmental observation, nursing, etc. Autonomous motion planning and control is necessary for realizing such applications. Industrial robots in factories would need to move their arms to assembly parts on belt conveyors. Successful repair robots at damaged Fukushima Daiichi nuclear plant will run over rubble, open a door, fix pipe leakage, for instance.

This thesis attempts to push the state-of-the-art of reinforcement learning forward to improve robot automation. Four major challenges in reinforcement learning in robotic applications are discussed: a) continuous high-dimensional state-action space, b) lack of the perfect dynamics models, c) sample efficiency, and d) undesirable convergence into locally optimal policy. There have been two major approaches, policy search and model learning, to solve these challenges,

This thesis proposes two policy search and one model learning algorithms. All these algorithms are proposed based on the idea of utilizing evolutionary computation, that is a general optimization framework using a population of candidate solutions.

First policy search algorithm tries to improve robustness against convergence into local optimal policy. The main idea is to diversify the search population in terms of behavior in the environment. This approach successfully improve the robustness, resulting in a significantly good performance in robot soccer domain.

The other policy search algorithm is proposed to enhance sample efficiency. This algorithm has shown success in popular cart-pole balancing task. The task could be solved efficiently with expert domain knowledge. Otherwise, it required a lot of training runs so far. Experimental results show that this algorithm can solve this task efficiently without any domain knowledge.

Both two policy search algorithms naturally handle continuous high-dimensional state-action space with policies represented with neural networks. Since they optimize the policies through experience, they do not require the dynamics models at hand. While they have tradeoffs between sample efficiency and robustness against local optima, they show significant performances in their extreme respectively.

Model learning method I propose attempts to further improve sample efficiency. By learning dynamics models with symbolic regression and effectively integrate learned model in motion planning, the algorithm could achieve intelligent autonomous behavior with extremely small amount of experience. Although this algorithm cannot scale to problems with exceptionally high-dimensional state-action space, its sample efficiency is among the best of existing algorithms to date.

The contribution of this thesis is twofold. One contribution is that each proposed algorithm updated the state-of-the-art of each directions, e.g. sample efficiency and robustness against local optima. The other is that it provides a toolbox for robot automation, from which users can choose according to their requirements and the characteristics of the target tasks.