Adaptive function approximation using critical state for reinforcement learning in large-dimensional continuous space

Makoto Futamoto (510089)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 8, 2007

Keywords: Reinforcement learning, NGnet,Two legged-robot,Adaptive function approximation,Critical state.

Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the greatest reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics – trial-and-error search and delayed reward – are the two most important distinguishing features of reinforcement learning.

Reinforcement learning is different from supervised learning, the kind of learning studied in most current research in machine learning, statistical pattern recognition, and artificial neural networks. Supervised learning is learning from examples provided by a knowledgeable external supervisor. This is an important kind of learning, but it is not alone sufficient for learning from interaction. In interactive problems, it is often impossible to obtain examples of desired behavior that are both correct and representative of all the situations in which the agent has to act. In uncharted territory – where one would expect learning to be most beneficial – an agent must be able to learn from its own experience.

Compared to other machine learning methods, the reinforcement learning has the big advantage that a learner can learn how to act without supervision through interaction with the environment. Instead of preparing appropriate examples by a knowledgeable external supervisor, the designer needs to define a reward function that maps 'states' to a numerical number indicating a shortsighted desirability of the state. The purpose of an agent is to obtain an optimal policy that maximizes the future rewards obtained for each state. Although the agent cannot know the consequences of a current action, the agent must take an action that maximizes the cumulative reward in the long run. So, the agent estimates state values defined in each state, which are an estimation value of the cumulative future reward under the current policy. The agent learn a policy that maximizes the state values.

Many interesting real-world control tasks, such as driving a car, require smooth continuous actions taken in response to high-dimensional, real valued sensory input. In applications of reinforcement learning to continuous problems, we must define the value function as a continuous function. The value function is usually represented by a certain function approximation. In the learning process, the value function must be estimated correctly to obtain a good policy. However, in general, it is difficult to approximate this function accurately when the state space dimensionality is high known as the 'curse of dimensionality'.

In this study, we propose a reinforcement learning method that can work efficiently for high-dimensional continuous problems. For this, we employ the normalized Gaussian network (NGnet) to approximate the value function. The NGnet consists of a number of basis functions that are located in a reticular pattern. As the number of basis functions become large, the accuracy of the value function approximation becomes accurate. However, when considering en 8 dimensional problem with 5 units for every axis, $5^8(390625)$ units are needed to cover the state space. However, this allocation of the basis function is too coarse to approximate the value function appropriately in real-world control tasks. Indeed, most of the state space is infrequently-used in the learning process. Thus, Allocating basis functions adaptively is a natural idea, i.e., allocating them densely around important states and coarsely otherwise.

In this thesis, we propose such kind of method and apply it to the problem that is en two legged-robot walking contorol where the state space is defined by en 8 dimensional continuous space. The task of learning walking movements in this problem using traditional RL methods is very hard. we used this problem to confirm the effectiveness of the proposed method.