Brownian Snake Measure-Valued Markov Decision Process

Zhenzhen Wang¹ and Hancheng Xing²

¹ School of Information Technology,
Jinling Institute of Technology, Nanjing, 211169, China
² School of Computer Science & Engineering,
Southeast University, Nanjing 210096, China
{wangzhenzhen,xhc}@seu.edu.cn

Abstract. This paper presents a model called Brownian snake measure-valued Markov decision process (BSMMDP) that can simulate an important characteristic of human thought, that is, when people think problems, sometimes they can suddenly connect events that are remote in space-time so as to solve problems. We also discuss how to find an (approximate) optimal policy within this framework. If Artificial Intelligence can simulate human thought, then maybe it is beneficial for its progress. BSMMDP is just following this idea, and trying to describe the talent of human mind.

Keywords: Brownian snake measure-valued Markov decision process, Markov decision process, human mind, Brownian snake.

1 Introduction

Markov decision process (MDP) model, that integrates Markov stochastic process theory with reinforcement learning method, has been paid more and more attention in the theoretic and applied areas of Artificial Intelligence, and with one understanding MDPs more deeply one wonders to use it to simulate the human mind, so that it’s possible for AI to be really a kind of “human-like” intelligence.

Because the environment is generally unobservable or partially observable, the Markov decision process carried out in the totally observable environment is naturally replaced by partially observed Markov decision process (POMDP). Let us carefully investigate the exact solutions of POMDP (the first exact solution was proposed by E.J. Sondik in 1971.): its essence is to substitute belief state space for true state space, and substitute the transition probability over the belief state space for the transition probability over the true state space, and so under invariable action space, if the reward function (for instance) on the belief state space is defined as the expectation of true states, then solving a POMDP on a physical state space can be reduced to solving a MDP on the corresponding belief state space. It indeed is a good thing because an agent does not know which real state he is in and all he knows is just a belief state (See [1], pp.481-483).
The agent’s belief state in POMDP model is a probability distribution that reflects a whole grasp for the agent of the real space. The agent just depends on this grasp to help itself to think problem. Moreover it relies on prior knowledge (such as transition probability and reward function in previous MDP) and partial observation to transform its belief state, and thereby finds an optimal policy to problem. In this way it will become a real “intelligent person”. Concerning the human thought, under fully (or partially) unknown circumstance people think problems and find the way out relying on a vague grasp of problem. Accompanied by deeply understanding problem, this grasp, more exactly speaking, --a “measure” to different parts of the problem will become more and more clear, and finally a complete solution to the problem will be found. Now the belief state of agent in POMDPs just simulates such phenomenon of human thought. So POMDP model gives us an important illumination that simulating human thought not only should be an important task in AI, but also usually be a key point for acquiring success in AI.

Another important phenomenon of human thought is that often people are able to suddenly link things that are remote in space-time, that is, the associative function plays an important role in human mind. People usually think problem in a logical manner (that is due to the long-term learning and experiences of human beings), but when having no idea about problem, the activities of human thought always move randomly, i.e., the thought “particle” moves as something as Brownian movement and shows quanta’s characteristic, i.e., its locations and its proceeding directions, which are hardly controlled by rationality, go out of their way to seek unexpected way out in mind space; and that is the phenomenon we have described above: things distant from another in time or space will be linked together in order to search the route for solving problems. This paper presents a model, called Brownian snake measure-valued Markov decision process (BSMMDP) that tries to simulate this phenomenon, and a (approximate) meta-algorithm for finding optimal policy within this framework is also provided.

2 Brownian Snake Measure-Valued Markov Decision Processes

2.1 Reflective Brownian Movement on Semi-straight Line

Suppose \( \bar{w} \) is a trajectory of reflective Brownian movement on the semi-straight line, and, \( \bar{w}(0) = 0 \). We write \( n(d\bar{w}) \) as the Itô measure of the Brownian movement on straight line, and it is standardized as \( n(\sup_{s \geq 0} \bar{w}(s) > \varepsilon) = \frac{1}{2\varepsilon}, \quad \varepsilon > 0 \). Next we introduce the conception of local time. A process \( \{L_a^{\varepsilon}(\bar{w}), t \geq 0, a \geq 0\} \) is called local time process of a reflective Brownian movement \( \bar{w}(t) \) if it satisfies the following qualities: (i) \( \forall a \geq 0, t \rightarrow L_a^{\varepsilon}(\bar{w}) \) is non-descending, and it has an increment only at \( \bar{w}(t) = a \); (ii) for an arbitrary measurable function \( \Psi : R_+ \rightarrow R_+ \), \( s \geq 0 : \int_0^s \Psi(\bar{w}(u))du = \int_0^s \Psi(a) \bullet L_a^{\varepsilon}(\bar{w})da \). If we do not seek technical details (though this is required by mathematics for strictness), then intuitively, \( L_a^{\varepsilon}(\bar{w}) \) could be considered