

Markov Decision Process Toolbox

Shan peng,Li ran,Yang qin ,Ning shenghua

Abstract

With the development of science and technology, it is hoped that the future of the systems can be predicted, and the development of systems can be controlled or influenced.

Markov decision process (MDP) is established to study and control the future development of stochastic systems, and it is a very vitality of the subject. MDP is not only a branch of Stochastic Operations Research but a branch of Applied Probability. At the same time, as the theory of Markov optimal control. MDP also belongs to the optimal control of stochastic systems areas.

Markov decision process is a theory to study the issue of a series of random sequential decision-making. The so-called random sequential decision-making refers to making a decision on a series of successive or continuous time point. On each time point of decision-making, decision-makers based on the observed state, choose one from some available decisions; and implemented the decision-making, the system will gain a reward related to the current state and the decision chosen. The reward will affect the system state on the next decision point. The system state on the next time point is random, and on the new time point, the decision-makers should observe the new state of the system, and take the new decision, such a step by step continue. The decision taken on each time point will affect the operation of the system on the next point, and thereby influence the future. The purpose of the decision-making is to make the operation of the system optimal to some extent.

On each observation time point, decision-makers based on the observed state of the system choose one from the available actions, which will lead to two things happen: (1) the system will gain a certain degree of economic efficiency, (2) we can determine the probability of subsequent development of the system. In fact, MDP is the only method of dynamic control on the stochastic discrete event dynamic system.

This model has discrete-time and continuous-time Markov decision process. On the basis, considering the model approached the fact, it has some observable, multi-goal, adaptive Markov process. In terms of criterium function, there are the discounted criterium, the average criterium and the exception total reward criterium, and so on.

There is a wide range of application areas for the MDP, such as the production of storage systems, system replacement maintenance, scheduling of manufacturing systems control, computer /communications network system control, dynamic asset pricing, advertising optimization, the pricing of goods and services, quality control, sequential search, the highway management, etc. For its widespread application, MDP

is provided in the SCILAB of scientific computing software in the form of toolbox , which is very convenient to people .

The source code of Markov Decision Process Toolbox (MDPtoolbox) is based on the basic grammar and instructions of SCILAB, and use the converting the MATLAB to SCILAB which is one of the SCILAB toolboxes. MDPtoolbox containing 18 functions takes a modular design and tightly interface and, is convenient to solve the expectation total reward criterion, discounted criterium and average criterion problems for discrete-time Markov decision-making process.

The finite-horizon is based on the theory of expectation of reward dynamic programming that is backward recursion value. MDPtoolbox provides two modes: the verbose model and the silent mode. The verbose model can display the current phase and the corresponding optimal policy.

The discounted criterium aims to the situation that there are loss in the system to achieve the optimal model. The five main algorithms has taken: linear programming algorithm, iteration policy ,policy iteration algorithm, modified policy iteration algorithm, value iteration algorithm, Gauss-Seidel's value iteration algorithm.

The average criterion aims to the situation that discount factor is very close to 1, or it is difficult to measure the criterion by means of the economic situation. Policy-makers used to replace the discounted criterium with average criterion.

Key words: Markov decision process, The finite-horizon, The discounted criterium, The average criterion

References

- [1] Ci-Hua Liu. Random process [M]. Wuhan: Huazhong University of Science and Technology Press, 2001.
- [2] Qi-Ying Hu, Yong -Jian Liu. Primer on Markov decision process [M]. Xi'an: Xidian University Press, 2007.
- [3] Liu Ke. Practical Markov decision-making process [M]. Beijing: Tsinghua University Press, 2004.
- [4] Puterman M L. Markov decision processes: discrete stochastic dynamic programming [M]. New York: Wiley, 1994.
- [5] Minkoff A S. A Markov decision model and decomposition heuristic for dynamic vehicle dispatching [J]. Operation Research, 1993, 41: 77~90.