Z-learning of linearly-solvable Markov Decision Processes*

Aleksandr Beznosikov*[1], *Yury Maximov, Michael Chertkov, Vadim Strijov
beznosikov.an@phystech.edu
[1]Moscow Institute of Physics and Technology

Considered methods for solving the problem of discrete Markov Decision Process. For certain class of MPDs which greatly simplify Reinforcement Learning. In this paper we adapt a modification (Z - learning) to the case of Markov Decision Process discussed in the context of energy systems and solve the optimal control problem by incomplete data. Comparing with standard Q-learning, show that modification of algorithm gives faster and more reliable solution.

Keywords: *Markov Decision Process, Z - learning, Q-learning.*

1 Introduction

In the area of power systems there is a huge demand on fast reinforcement learning algorithms, but there is still a lack of that. In this paper we solve the problem of optimal energy system consisting of a set of devices from [1].

The behavior of the system of devices in time is considered as a discrete Markov process. In general, this problem is not solved simply. But in [4] there are several ways to solve it optimally. Most methods require knowledge of what happens if a system "left alone"long enough. But in practice it often happens that this information is hidden from us.

In this paper it is proposed to use the Z-learning method (stochastic modification of Q-learning from [2, 3]). Together with the solution of the main task of the MDP in parallel to restore unknown data on the behavior of the system. With this algorithm, a working model of system management will be built based only on limited samples of representative behavior. We compare the speed and quality of the two algorithms: Z-learning and Q-learning, when solving the MDP, describing via transition probability matrix. Given initial state vector (probability of being in a state at time zero), we generate data for the time evolution of the state vector.

Литература

[1] Michael Chertkov and Vladimir Y. Chernyak. Ensemble control of cycling energy loads: Markov decision approach. *CoRR*, abs/1701.04941, 2017.

[2] Chi Jin, Zeyuan Allen-Zhu, Sébastien Bubeck, and Michael I. Jordan. Is q-learning provably efficient? *CoRR*, abs/1807.03765, 2018.

[3] Csaba Szepesvári. *Algorithms for Reinforcement Learning.* Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2010.

[4] Emanuel Todorov. Linearly-solvable markov decision problems. In Bernhard Schölkopf, John C. Platt, and Thomas Hofmann, editors, *NIPS*, pages 1369–1376. MIT Press, 2006.

*Supervisor: Vadim Strijov Task author: Michael Chertkov Consultant: Yury Maximov