

АЛГОРИТМ ОБУЧЕНИЯ СЕТИ ФУНКЦИОНАЛЬНЫХ СИСТЕМ В СТОХАСТИЧЕСКОЙ СРЕДЕ

Сорокин Артём Юрьевич

Аспирант

Кафедра Математических и Информационных Технологий СПб АУ РАН,

Санкт-Петербург, Россия

E-mail: griver29@gmail.com

Многие современные алгоритмы машинного обучения были вдохновлены тем, как обучаются живые организмы. Однако, несмотря на успехи последнего времени, базовые механизмы обучения животных еще недостаточно поняты. В частности, человек и многие животные способны адаптироваться к изменяющимся условиям среды и в процессе целенаправленного поведения не только выучивать новые стратегии поведения, но и при необходимости возвращаться к старым.

В данной работе в рамках поведенческой теории функциональных систем П.К. Анохина [1] рассматривается задача создания адаптивного агента способного успешно обучаться в стохастической среде и удовлетворяющего следующим критериям:

1. обеспечение достижения нескольких альтернативных результатов в среде;
2. возможность достижения одной цели разными способами;
3. самообучение новой последовательности действий, ведущей к получению результата, при невозможности достижения цели известным способом;
4. сохранение старых решений задачи при формировании и запоминании новых альтернатив.

Для решения этой задачи был разработан алгоритм обучения высокоуровневой модели сети функциональных систем [2]. Фактически обучение строится по аналогии с алгоритмами temporal-difference (td) learning (наиболее близким аналогом является Q-learning) [3], при этом часть весов к вторичным функциональным системам используется для хранения функции ценности. Несмотря на похожесть, предложенный алгоритм использует свой механизм распространения td-ошибки. Например, затухание td-ошибки происходит не равномерно со временем, а лишь в определенных состояниях, представляющих

Таблица 1: Эффективность тестируемых алгоритмов после обучения в средах первого и третьего типов стохастической изменчивости

N	Первый тип			Третий тип		
	FSN	RL	RND	FSN	RL	RND
25	6.4	18.3	87.1	6.4	4.5	68
100	10.3	37.2	358.8	9.2	5.7	276.4
400	23.3	194.3	1891	24.4	9.4	1240

из себя развилки между несколькими жестко выученными последовательностями действий.

Предложенный алгоритм, вместе с табличными td-алгоритмами (Q-learning, SARSA) и случайным поиском, был протестирован на дискретных средах разной топологии, размера и типа стохастической изменчивости среды. Первый тип изменчивости соответствовал среде, которая изменялась от одного эпизода обучения к другому, но в течении эпизода определенное действие агента в определенном состоянии среды всегда приводило к одним и тем же результатам. При втором типе среда медленно менялась в течении эпизода, и при возвращении агента в уже посещенное состояние реакция среды на действия агента могла измениться. Третий тип изменчивости совпадал с марковским процессом принятия решений (МППР) [3].

Частично результаты обучения представлены в таблице 1. В данном случае успешность стратегии, которой обучился агент на прямую зависит от того как быстро он достигает целевого состояния в среде, то есть сколько действий он на это затрачивает. В таблице приведено среднее число действий агента за последние 10 эпизодов обучения на всех средах с одинаковым числом состояний (параметр N). Аббревиатуры: FSN и RND — обозначают результаты агентов под управлением сети функциональных систем и алгоритма случайного поиска соответственно. В колонке RL (Reinforcement Learning) представлены результаты алгоритма Q-learning с параметрами $\alpha = 0.1$, $\gamma = 0.9$, $\epsilon = 0.1$, так как именно он показал наилучшие среди алгоритмов обучения с подкреплением результаты.

В средах первого и второго типа стохастической изменчивости алгоритм оказался способен выучить в 2–4 раза более эффективные стратегии достижения цели, чем протестированные td-алгоритмы,

причем его производительность практически не зависела от типа изменчивости среды.

Работа выполнена при поддержке гранта РФФ 15-11-30014.

Литература

1. Anokhin P. Biology and Neurophysiology of the Conditioned Reflex and Its Role in Adaptive Behavior. Pergamon, Oxford, 1974.
2. Komarov M. A., Osipov G. V., Burtsev M. S. Adaptive functional systems: Learning with chaos Chaos: An Interdisciplinary Journal of Nonlinear Science. 2010. Т. 20. No. 4. С. 045119.
3. Sutton R. S., and Barto A. G. Reinforcement Learning: An Introduction. The MIT Press, Cambridge, MA, USA. 1998.