1 УДК 004.853

А.Н. Алфимцев [1]***, А.Р. Питикин*** [1]

[1]*Московский государственный технический университет им. Н. Э. Баумана (национальный исследовательский университет), г. Москва, Российская Федерация*

ЭМЕРДЖЕНТНЫЕ СВОЙСТВА МУЛЬТИАГЕНТНОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Аннотация. В данной работе представлены десять эмерджентных свойств мультиагентного обучения с подкреплением. Каждое свойство формализовано с использованием марковских процессов принятия решений и представлено в виде формулы. Выдвинуто предположение, что такая формализация позволит в дальнейшем прицельно обучать мультиагентую систему для получения необходимых эмерджентных свойств. Установлено, что эмерджентность в данной сфере является слабой. Проанализированы высокоцитирумые публикации по теме мультиагентного обучения с целью проверки наличия сформулированных свойств. По результатам проделанной работы представлена сводная таблица свойств с указанием алгоритма, в котором свойство было обнаружено, среда, для которой алгоритм создан, архитектура нейронной сети агента, а также используемая схема обучения с подкреплением.

Ключевые слова: искусственный интеллект, глубокое машинное обучение, обучение с подкреплением, мультиагентная система, прикладное программное обеспечение.

A.N. Alfimtsev [1]***, A.R. Pitikin*** [1]

[1]*Bauman Moscow State Technical University, Moscow, Russian Federation*

EMERGENCY PROPERTIES OF MULTI-AGENT REINFORCEMENT LEARNING

Abstract. This paper presents ten emergent properties of multi-agent reinforcement learning. Each property is formalized using Markov decision processes and presented as a formula. It has been suggested that such a formalization will allow further targeted training of a multi-agent system to obtain the necessary emergent properties. It has been established that emergence in multi-agent reinforcement learning is weak. Highly cited publications on the topic of multi-agent learning were analyzed in order to check the presence of the formulated properties. Based on the results of the work done, a summary table of properties is presented indicating the algorithm in which the property was discovered, the environment for which the algorithm was created, the architecture of the agent's neural network, and the reinforcement learning scheme used.

Keywords: artificial intelligence, deep machine learning, reinforcement learning, multi-agent system, application software.

Введение. С точки зрения системной инженерии главную теоретическую основу современных исследований в области искусственного интеллекта составляют исследования и разработки в области машинного обучения [1]. Однако до недавнего времени исследования алгоритмов машинного обучения имели серьезное ограничение, связанное с очень большим количеством возможных состояний реальной среды, что ограничивало применение алгоритмов областью простых задач в табличном мире-сетке [2]. Исследования и разработки компаний уровня OpenAI и DeepMind расширили применение этих алгоритмов для высоко размерных и сложных искусственных сред — компьютерных видеоигр [3]. Алгоритмы машинного обучения с подкреплением продемонстрировали достижения, сравнимые с достижениями человека в аркадных играх Atari, стратегических играх Dota 2 и StarCraft II. Один и тот же агент обучался в самых различных играх с почти одинаковым успехом, что говорит о его высокой способности обобщать данные и знания, полученные из среды.

В данном случае под агентом понималась компьютерная программа, расположенная в виртуальной среде и способная автономно действовать и обучаться в этой среде для достижения целей функционирования мультиагентной системы [4]. При этом мультиагентной системой называют систему, состоящую из двух и более взаимодействующих агентов [5]. В целом, совокупность алгоритмов и моделей, которые позволяют агенту без предварительного

программирования находить закономерности в данных и самостоятельно принимать решения на основе опыта взаимодействия, относят к области машинного обучения [6].

Академический интерес к мультиагентному обучению с подкреплением растет на протяжении последних 5 лет. Ниже, на рисунке 1, приведен график, показывающий динамику роста количества статей по ключевым словам «multi agent reinforcement learning» относительно общего числа статей научной конференции «Нейронные системы обработки информации» (англ. The Conference and Workshop on Neural Information Processing Systems, NIPS) [7].

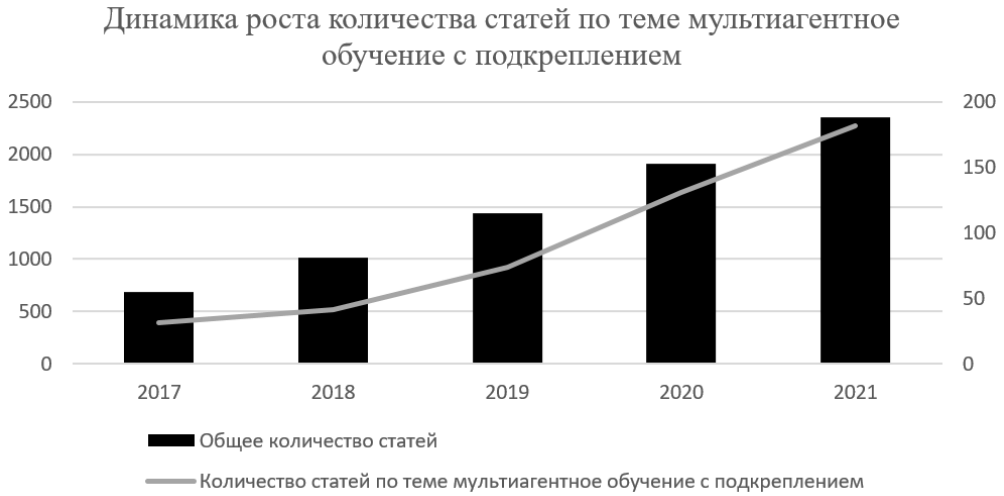

Рис. 1. Динамика роста количества статей по теме мультиагентное обучение с подкреплением

Ряд задач из реального мира могут быть удобно представлены в виде кооперативных мультиагентных систем, например, управление трафиком и координация автономных устройств в умном городе. В задачах, где задействовано множество одновременно обучающихся или взаимодействующих агентов сегодня объективно наблюдаются эффекты эмерджентности. Согласно декомпозируемому подходу к определению эмерджентности, это понятие определяется как наличие у системы свойств, которыми не обладают ее компоненты, а также несводимость свойств системы к сумме свойств ее частей [8]. В настоящий момент выделяют два основных вида эмерджентности – это сильная эмерджентность и слабая [9].

Слабая эмерджентность подтверждает реальность сущностей и особенностей, установленных в специальных науках, а также подтверждает физикализм, тезис о том, что все природные явления полностью конституированы и полностью метафизически детерминированы фундаментальными физическими явлениями, из чего следует, что любой физический эффект фундаментального уровня имеет чисто фундаментальную физическую причину [10]. Сторонники сильной эмерджентности утверждают, что, по крайней мере, некоторые явления более высокого уровня демонстрируют более слабую зависимость/более сильную автономию, чем допускает слабая эмерджентность. Это часто принимает форму отказа от физической реализации, утверждения фундаментальных причинных сил более высокого уровня или и того, и другого [11].

Таким образом, в данной статье под термином «эмерджентность» будет пониматься именно слабая эмерджентность. Возникновение высокоуровневых свойств системы при более сильной автономии, а также форма отказа от возможности причинно-следственной реализации эмерджентных свойств не рассматриваются в контексте данной статьи. Кроме того, в мультиагентных системах часто под эмерджентностью понимают самоорганизацию агентов в мультиагентной системе, что также не является предметом исследования данной работы [8].

Однако, свойства эмерджентности не классифицированы и не формализованы в рамках мультиагентного обучения с подкреплением. Методо-ориентированная классификация эмерджентных свойств позволит в будущем решать следующие задачи. Во-первых, появится воз-

можность автоматически проверять программную систему на наличие эмерджентных свойств и формально доказывать их наличие, верифицируя переход программной системы на новый, более качественный уровень развития. Во-вторых, появится способ прицельно обучать группы агентов с целью проявления эмерджентных свойств в программной системе. В-третьих, формализация в данной работе эмерджетных свойств на базе модели марковских процессов принятия решений позволит в дальнейшем сразу использовать технологии глубоко обучения с подкреплением.

Далее структура статьи следующая. В разделе 2 даны основные понятия марковских процессов принятия решений, как основы машинного обучения с подкреплением. В разделе 3 на основе сравнительного экспериментального анализа современных высокоцитируемых публикаций в области мультиагентного обучения с подкреплением выявлены, классифицированы и формализованы десять базовых свойств слабой эмерджентности. Раздел 4 является заключительным.

Марковский процесс принятия решений как базовая модель обучения с подкреплением. В области информационных технологий для формализации эмерджентных свойств программных систем использовались подходы, основанные на теории графов [12], марковских процессах [13], темпоральной логике [14]. Однако основные алгоритмы обучения с подкреплением были разработаны именно для математической модели марковского процесса принятия решений (МППР), представляющей собой кортеж, вида (S, A, T, R), где S – конечное множество состояний, A – конечное множество действий, T: $S \times A \times S \to [0; 1]$ – функция переходов, $R: S \to \mathbb{R}$ – функция наград [15].

Марковский процесс принятия решений обладает марковским свойством, будущие переходы зависят только от текущего состояния. Вероятность перехода в состояние s' после выполнения действия a в состоянии s обозначается $T(s, a, s')$. Для всех действий a и для всех состояний s и s' выполняются неравенства $0 \le T(s, a, s') \le 1$ и $\sum_{s' \in S} T(s, a, s') = 1$. Функция наград R возвращает награду $R(s, a, s')$ после выполнения действия a в состоянии s и перехода в состояние s'. Целью обучения в МППР является поиск оптимальной стратегии $\pi^*: S \to A$, которая задает действия агента, приводящие к получению максимальной награды. Качество же стратегии определяется функцией ценности, обозначаемой V^π. Оптимальной стратегией $\pi^*(s)$ называется стратегия, при которой $V^{\pi^*}(s) \ge V^\pi(s), s \in S$ для всех π.

Функции ценности состояния для стратегии π определяет суммарное количество награды, которую ожидает агент получить, начав из состояния s следовать стратегии π:

$$Q^\pi(s, a) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma_t R(s_t) | s_0 = s].$$

где γ – параметр дисконтирования, $\gamma \in [0; 1]$.

Функция ценности действия для стратегии π определяет общее количество награды, которую агент ожидает получить, если выполнит действие a в состоянии s, следуя стратегии π:

$$Q^\pi(s, a) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma_t R(s_t, a_t) | s_0 = s, a_0 = a], \gamma \in [0; 1].$$

Поиск оптимальной стратегии с помощью МППР и децентрализованных модификаций этой модели является классической задачей как для одногагентного, так и для мультиагентного обучения с подкреплением, и в настоящее время существует большое количество алгоритмов, успешно решающих эту задачу в различных средах [15]. В следующем разделе с использованием МППР будут формализованы эмерджентные свойства мультиагентного обучения с подкреплением.

Классификация свойств слабой эмерджентности. В данном разделе приведем основные свойства слабой эмерджентности в контексте мультиагентного обучения с подкреплением.

I. Возникновение нового поведения.

Большинство работ в области мультиагентного обучения с подкреплением направлены именно на то, чтобы обучить агента из простых доступных ему действий составлять более сложные стратегии поведения. Возникновение нового поведения на уровне агента или всей мультиагентной системы сигнализирует об эффективности накопления опыта за счет обуче-

ния с подкреплением и достижении нужной абстракции знаний о среде. Например, если базовыми действиями агента является перемещения в простом табличном мире-сетке на одну клетку вперед, назад, то стратегии «перейти в другую зону», «перегородить проход», «держаться подальше от оппонента» представляет собой качественно новую комбинацию из существующих действий:

$$(\forall A_i \in A)\exists Y = \{a_i, \dots a_{i+n}\}, Y \subset a,$$

где A – множество всех агентов, a – множество всех действий, Y – новое поведение агента, определяемое как некоторая комбинация из уже существующих действий или стратегий агента.

Данное эмерджентное свойство прослеживается во многих алгоритмах мультиагентного обучения с подкреплением, пытающихся повторить инновационный алгоритм AlphaStar [16].

II. Возникновение ролей.

Гомогенные кооперативные агенты имеют единое множество возможных действий [15]. Под появлением ролей понимается ситуация, в которой каждый агент формирует свое уникальное множество действий, которое является подмножеством основного множества и при этом не пересекается с множествами других агентов. То есть:

$$\big(\forall A_i, A_j \in A\big)\exists X_i = \{a_k, \dots a_{k+n}\}, X_j = \{a_l, \dots a_{l+m}\}, X_i \cap X_j = \emptyset, X_i \subset a, X_j \subset a,$$

где A – множество всех агентов, a – множество всех действий, X_i – набор действий i-го агента, определяющий его роль.

В данном случае имеется в виду именно эмерджентное появление ролей [17], а не использование агентом заранее определённых ролей [18].

III. Возникновение языка.

При коммуникации обучающихся агентов возможно появление между ними некоторого языка, представляющего собой связь между алфавитами двух различных агентов. Под алфавитом каждого агента в этом случае понимается набор некоторых символов, которые агент способен передать по каналу связи:

$$\big(\forall A_i, A_j \in A\big)\exists c_i = \{x_0 \dots x_n\}, c_j = \{y_0 \dots y_n\}, \exists\, \theta_{ij} = \{(x_0, y_0), \dots, (x_n, y_n)\}, \exists\, \omega_{ij} = \bigcup \theta_{ij},$$

где c_i, c_j – наборы символов, которые могут передавать агенты A_i, A_j, $x_0 \dots x_n$, $\{y_0 \dots y_n\}$ – множества символов, θ_{ij} – множество сообщений, которыми агенты могут обмениваться, θ_{ij} – язык агентов A_i, A_j.

Данное эмерджентное свойство возникает из-за необходимости агентам обмениваться информацией, чтобы получить максимальную награду в среде [19].

IV. Возникновение социальной организации.

Под данным эмерджентным свойством подразумевается появление некоторой социальной иерархической организации в системе взаимодействий агентов. В иерархическом мультиагентом обучении с подкреплением это свойство реализуется через способность агента к выделению локальных подзадач в своей глобальной задаче. Данная декомпозиция позволяет агенту достичь цели наиболее эффективным образом в сложной среде [20].

$$(\forall A_i \in A)\ \exists\, T_i = \{a_n, \dots, a_m\}, \exists\, t_{i_j} = \{a_x, \dots, a_y\}, \bigcup_{j=1}^{|T_i|} t_{i_j} = T_i, T_i \subset a,$$

где A – множество всех агентов, T_i – задача агента A_i, представленная в виде набора некоторых действий, t_{i_j} – задача, являющаяся подзадачей для задачи T_i.

V. Возникновение награды.

В некоторых средах может возникнуть ситуация, при которой внешняя награда по каким-либо причинам либо отсутствует, либо является очень редкой. В таком случае агенты используют различные подходы для формирования внутренней награды, называемой также внутренней мотивацией агента. В качестве внутренней мотивацией агента может использоваться степень новизны вновь полученного из среды состояния [21]:

$$(\forall A_i \in A)\exists Buf_i = \{s_k, \dots s_n\}, \exists S_{new} \notin \mathrm{M}_i \rightarrow \exists f_R\ (A_i, S_{new}) = R_{in}, \mathrm{M}_i \in S,$$

где A – множество всех агентов, S – общее множество состояний среды, M_i – множество всех состояний среды, в которых агент A_i был, S_{new} – новое состояние для агента A_i, $f_R\ (A_i, S_{new})$ – функция внутренней награды агента, R_{in} – внутренняя награда агента.

VI. Системная трансформация.

Кооперативные агенты могут объединяться в группы. При группировке возможны ситуации, в которых некоторый агент меняет свою идентичность или поведение. В качестве примера рассмотрим юнитов из стратегической игры StarCraft II. Агент-мародер расы терранов имеет возможность атаки как бронированных целей, так и пехоты противника. При формировании подразделения, то есть группы агентов, объединенных общей задачей, целью которого является атака пехоты противника, данный агент теряет возможность атаки бронированных целей.

$$\exists A_i \in A, \exists G = \{A_i, \dots A_k\}, X_{i_G} \neq\ X_{i_0}, G \subset A,$$

где G – группа агентов, X_{i_G} – роль агента в составе группы, то есть набор действий, выполняемых агентом в группе, X_{i_0} – роль агента до поглощения группой.

Проявление данного эмерджентного свойства присутствует в алгоритме RODE [18].

VII. Возникновение нового качества.

Это свойство описывает появление нового качества группы агентов не за счет появления новых функций, а за счет их пространственно-временных отношений. Примером рождения такого свойства может служить результат работы алгоритма ROMA [17]. В этой среде возникают ситуации, в которых агентам выгодно занимать позиции на плоскости в форме различных геометрических фигур, таких как треугольник или полукруг.

$$\forall A_i, A_j \in A\ \exists\ P_{i_t} = \left(x_{i_t}, y_{i_t}\right), P_{j_t} = \left(x_{j_t}, y_{j_t}\right), \exists\ W_{ij_t} =\ P_{i_t}P_{j_t}, \cup_{i=1,j=1}^{|A|} W_{ij_t} = \varphi, \varphi \in \Phi,$$

где P_{i_t}, P_{j_t} – позиции агентов A_i, A_j в момент времени t, $\left(x_{i_t}, y_{i_t}\right)$, $\left(x_{j_t}, y_{j_t}\right)$ – координаты агентов в момент времени t, W_{ij_t} – отрезок, соединяющий позиции агентов в момент времени t, φ – геометрическая фигура на плоскости, Ф – множество всех геометрических фигур на плоскости.

VIII. Трансформация индивида.

В среде итеративной мультиагентной игры если два различных гомогенных агента оказываются в одинаковом состоянии в один и тот же момент времени, то оба агента имеют одинаковый набор действий в этом состоянии. В этом случае трансформации индивида не происходит, оба агента остаются гомогенными. Однако, в случае если два агента, находясь в одном и том же состоянии, имеют разный набор действий, то можно сказать, что произошла трансформация одного из агентов и теперь они не гомогенны [22]:

$$\exists A_i, A_j \in A, s_{i_t} =\ s_{j_t}, a_{i_t} \neq\ a_{j_t}, \rightarrow A_i \in T_k, A_j \in T_n, T_k\ \neq T_n, T_k \subset A, T_n \subset A,$$

где A_i, A_j – два агента, s_{i_t}, s_{j_t} – состояние агентов в момент времени t, a_{i_t}, a_{j_t} – множества доступных им действий в момент времени t, T_k, T_n – типы агентов.

IX. Возникновение координации.

В некоторых задачах мультагентного обучения присутствует такое понятие, как множество совместных действий агентов. Под таким множеством понимается набор из согласованных действий агентов, приводящих к общей награде, данное поведение агентов возникает эмерджентно. Данное свойство можно формализовать в следующем виде:

$$\left(\forall A_i, A_j \in A\right)\exists a_{joined_{ij}} = \left\{\left(a_{i_k}, a_{j_n}\right), \dots, \left(a_{i_x}, a_{j_y}\right)\right\}, \cup_{u=1}^{|a_i|} a_{i_u} = a_i, \cup_{u=1}^{|a_j|} a_{j_u} = a_j, a_i\ \subset a, a_j\ \subset a,$$

где A – множество всех агентов, a – множество всех действий, $a_{joined_{ij}}$ – множество всех совместных действий для агентов A_i, A_j.

Примером проявления такого эмерджентного свойства является ситуация в среде компьютерной игры футбол, когда некоторый агент, находящийся в не самой удачной позиции для удара по воротам, решает отдать передачу агенту, находящемуся в более выгодной для удара позиции [23].

X. Возникновение неравенства.

В кооперативных средах обычно существует некоторое множество совместных действий агентов. При этом может возникнуть ситуация, когда два агента выполняли действие из множества совместных действий, но получили разную награду за эти действия. Это свойство ведет к нарушению кооперации и возникновению социальной дилеммы [24]:

$$(\forall A_i, A_j \in A)\exists a_{joined_{ij}} = \left\{\left(a_{i_k}, a_{j_n}\right), \dots, \left(a_{i_x}, a_{j_y}\right)\right\}, a_{i_t} \in a_{joined_{ij}}, a_{j_t} \in a_{joined_{ij}}, R_{i_t} \neq R_{j_t},$$

где A – множество всех агентов, a – множество всех действий, $a_{joined_{ij}}$ – множество всех совместных действий для агентов A_i, A_j, a_{i_t}, a_{j_t} – действия, совершенные агентами в момент времени t, R_{i_t}, R_{j_t} – награды агентов полученные после совершения действия в момент времени t.

Рассмотренные выше свойства объективно наблюдаются в различных средах, в которых обучаются разные архитектуры агентов разными алгоритмами мультиагентного обучения с подкреплением (таблица 1).

Таблица 1.

Алгоритмы, архитектуры, среды мультиагентного обучения с подкреплением и их эмерджентные свойства

Эмерджентное свойство	Название алгоритма или фамилия автора	Среда	Основная архитектура нейронной сети агента	Схема обучения с подкреплением
Возникновение нового поведения	AlphaStar [16]	StarCraft II	Сверточная, рекуррентная	TD(λ), V-trace, UPGO
Возникновение ролей	ROMA [17]	SMAC	Рекуррентная	QMIX
Возникновение языка	Havrylov S. [19]	ReferItGame	Генеративно-состязательная	Q-learning
Возникновение социальной организации	ACB&OPRE [25]	Melting Pot 2.0	Рекуррентная	A2C
Возникновение награды	LIIR [21]	SC2LE, SMAC	Рекуррентная	A2C
Системная трансформация	RODE [18]	SMAC	Рекуррентная	Q-learning
Возникновение нового качества	ROMA [17]	SMAC	Рекуррентная	QMIX
Трансформация индивида	Joel Z. Leibo [22]	Melting Pot	Рекуррентная	A3C
Возникновение координации	Siqi L. [23]	MUJOCO Soccer Environment	Рекуррентная	PBT
Возникновение неравенства	Johanson M. [24]	The Fruit Market Environment	Многослойная, рекуррентная	A2C

Заключение. В данной работе классифицированы и формализованы десять эмерджентных свойств мультиагентного обучения с подкреплением. Такая формализация позволит в дальнейшем прицельно обучать мультиагентую систему для получения необходимых эмерджентных свойств. Кроме того, представляется интересной возможность автоматической проверки наличия свойств у системы, а также их формальное доказательство. Свойства были сформулированы на основе модели марковских процессов принятия решений, что позволяет применять к ним алгоритмы обучения с подкреплением.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Jiang H. et al. Applications and development of artificial intelligence system from the perspective of system science: A bibliometric review // Systems Research and Behavioral Science. 2022. Vol. 39. №. 3. pp. 361-378.

2. Busoniu L., Babuska R., De Schutter B. A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). IEEE, 2008, vol. 38 (2), pp. 156–172. DOI: https://doi.org/10.1109/TSMCC.2007.913919

3. Jaderberg M., Czarnecki W.M., Dunning I., et al. Human-level performance in 3D multiplayer games with population-based reinforcement learning. Science, 2019, vol. 364, no. 6443, pp. 859–865. DOI: https://doi.org/10.1126/science.aau6249

4. Du W., Ding S. A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications. Artificial Intelligence Review, 2021, vol. 54, no. 5, pp. 3215–3238.

5. Goodfellow I., Bengio Y., Courville A. Deep learning. New York, MIT press, 2016, 800 p.

6. Hernandez-Leal P., Kartal B., Taylor M.E. A survey and critique of multiagent deep reinforcement learning. Autonomous Agents and Multi-Agent Systems, 2019, vol. 33, no. 6, pp. 750-797.

7. Neural Information Processing Systems, NeurlIPS, https://papers.nips.cc/

8. Kalantari S., Nazemi E., Masoumi B. Emergence phenomena in self-organizing systems: a systematic literature review of concepts, researches, and future prospects // Journal of Organizational Computing and Electronic Commerce. 2020. Vol. 30. №. 3. pp. 224-265.

9. O'Connor, Timothy, "Emergent Properties", The Stanford Encyclopedia of Philosophy (Winter 2021 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2021/entries/properties-emergent/.

10. Цветков В.Я. ЭМЕРДЖЕНТИЗМ // Международный журнал прикладных и фундаментальных исследований. – 2017. – № 2-1. – С. 137-138; https://applied-research.ru/ru/article/view?id=11234.

11. Wilson, Jessica M., 2015, "Metaphysical Emergence: Weak and Strong", in Tomasz Bigaj and Christian Wüthrich (eds.), Metaphysical Emergence in Contemporary Physics, Amsterdam: Rodopi, 251–306.

12. Moncion T., Amar P., Hutzler G. Automatic characterization of emergent phenomena in complex systems // Journal of Biological Physics and Chemistry. 2010. Vol. 10. pp. 16--23.

13. Zeigler B. P., Muzy A. Some modeling & simulation perspectives on emergence in system-of-systems // Spring Simulation Multi-conference (SpringSim'16). 2016. pp. 1-5.

14. Chen C. C., Nagl S. B., Clack C. D. Specifying, detecting and analysing emergent behaviours in multi-level agent-based simulations // Summer Computer Simulation Conference 2007, SCSC'07. Vol. 2. pp. 969-976.

15. А.Н. Алфимцев Мультиагентное обучение с подкреплением. М.: МГТУ им. Н. Э. Баумана, 2021. 224 с.

16. Vinyals O. et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning // Nature. 2019. Vol. 575. №. 7782. pp. 350-354.

17. Tonghan Wang, Heng Dong, Victor Lesser, Chongjie Zhang. ROMA: Multi-Agent Reinforcement Learning with Emergent Roles, 2020, ICML(2020). DOI: https://doi.org/10.48550/arXiv.2003.08039

18. Tonghan Wang, Tarun Gupta, Anuj Mahajan, Bei Peng, Shimon Whiteson, Chongjie Zhang. RODE: Learning Roles to Decompose Multi-Agent Tasks, 2020, DOI: https://doi.org/10.48550/arXiv.2010.01523

19. Havrylov S., Titov I. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols // Advances in neural information processing systems. 2017. Vol. 30. pp. 1-11.

20. Heechang Ryu, Hayong Shin, Jinkyoo Park. Multi-Agent Actor-Critic with Hierarchical Graph Attention Network, 2020, AAAI(2020). https://ojs.aaai.org/index.php/AAAI/article/view/6214

21. Yali Du, Lei Han, Meng Fang, Ji Liu, Tianhong Dai, Dacheng Tao. LIIR: Learning Individual Intrinsic Reward in Multi-Agent Reinforcement Learning, 2019, NeurlIPS (2019).

22. Joel Z. Leibo, Edgar Duéñez-Guzmán, Alexander Sasha Vezhnevets, John P. Agapiou, Peter Sunehag, Raphael Koster, Jayd Matyas, Charles Beattie, Igor Mordatch, Thore Graepel. Scalable Evaluation of Multi-Agent Reinforcement Learning with Melting Pot, 2021, International Con-

ference on Machine Learning 2021 (pp. 6187-6199).

23. Siqi Liu, Guy Lever, Josh Merel, Saran Tunyasuvunakool, Nicolas Heess, Thore Graepel. Emergent Coordination Through Competition, 2019, ICLR(2019). DOI: https://doi.org/10.48550/arXiv.1902.07151

24. Michael Bradley Johanson, Edward Hughes, Finbarr Timbers, Joel Z. Leibo. Emergent Bartering Behaviour in Multi-Agent Reinforcement Learning, 2022, DOI: https://doi.org/10.48550/arXiv.2205.06760.

25. John P. Agapiou, Alexander Sasha Vezhnevets, Edgar A. Duéñez-Guzmán, Jayd Matyas, Yiran Mao, Peter Sunehag, Raphael Köster, Udari Madhushani, Kavya Kopparapu, Ramona Comanescu, DJ Strouse, Michael B. Johanson, Sukhdeep Singh, Julia Haas, Igor Mordatch, Dean Mobbs, Joel Z. Leibo. Melting Pot 2.0, 2022, https://doi.org/10.48550/arXiv.2211.13746

REFERENCES

1. Jiang H. et al. Applications and development of artificial intelligence system from the perspective of system science: A bibliometric review // Systems Research and Behavioral Science. 2022. Vol. 39. №. 3. pp. 361-378.

2. Busoniu L., Babuska R., De Schutter B. A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). IEEE, 2008, vol. 38 (2), pp. 156–172. DOI: https://doi.org/10.1109/TSMCC.2007.913919

3. Jaderberg M., Czarnecki W.M., Dunning I., et al. Human-level performance in 3D multiplayer games with population-based reinforcement learning. Science, 2019, vol. 364, no. 6443, pp. 859–865. DOI: https://doi.org/10.1126/science.aau6249

4. Du W., Ding S. A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications. Artificial Intelligence Review, 2021, vol. 54, no. 5, pp. 3215–3238.

5. Goodfellow I., Bengio Y., Courville A. Deep learning. New York, MIT press, 2016, 800 p.

6. Hernandez-Leal P., Kartal B., Taylor M.E. A survey and critique of multiagent deep reinforcement learning. Autonomous Agents and Multi-Agent Systems, 2019, vol. 33, no. 6, pp. 750–797.

7. Neural Information Processing Systems, NeurIPS, https://papers.nips.cc/

8. Kalantari S., Nazemi E., Masoumi B. Emergence phenomena in self-organizing systems: a systematic literature review of concepts, researches, and future prospects // Journal of Organizational Computing and Electronic Commerce. 2020. Vol. 30. №. 3. pp. 224-265.

9. O'Connor, Timothy, "Emergent Properties", The Stanford Encyclopedia of Philosophy (Winter 2021 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2021/entries/properties-emergent/.

10. Tsvetkov V.Ya. Emergence // International Journal of Applied and Fundamental Research. – 2017. – № 2-1. – pp. 137-138; https://applied-research.ru/ru/article/view?id=11234.

11. Wilson, Jessica M., 2015, "Metaphysical Emergence: Weak and Strong", in Tomasz Bigaj and Christian Wüthrich (eds.), Metaphysical Emergence in Contemporary Physics, Amsterdam: Rodopi, 251–306.

12. Moncion T., Amar P., Hutzler G. Automatic characterization of emergent phenomena in complex systems // Journal of Biological Physics and Chemistry. 2010. Vol. 10. pp. 16--23.

13. Zeigler B. P., Muzy A. Some modeling & simulation perspectives on emergence in system-of-systems // Spring Simulation Multi-conference (SpringSim'16). 2016. pp. 1-5.

14. Chen C. C., Nagl S. B., Clack C. D. Specifying, detecting and analysing emergent behaviours in multi-level agent-based simulations // Summer Computer Simulation Conference 2007, SCSC'07. Vol. 2. pp. 969-976.

15. Alfimtsev A. N. Multi-agent reinforcement learning. BMSTU Publ., 2021. 224 p.

16. Vinyals O. et al. Grandmaster level in StarCraft II using multi-agent reinforcement learn-

ing // Nature. 2019. Vol. 575. №. 7782. pp. 350-354.

17. Tonghan Wang, Heng Dong, Victor Lesser, Chongjie Zhang. ROMA: Multi-Agent Reinforcement Learning with Emergent Roles, 2020, ICML(2020). DOI: https://doi.org/10.48550/arXiv.2003.08039

18. Tonghan Wang, Tarun Gupta, Anuj Mahajan, Bei Peng, Shimon Whiteson, Chongjie Zhang. RODE: Learning Roles to Decompose Multi-Agent Tasks, 2020, DOI: https://doi.org/10.48550/arXiv.2010.01523

19. Havrylov S., Titov I. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols // Advances in neural information processing systems. 2017. Vol. 30. pp. 1-11.

20. Heechang Ryu, Hayong Shin, Jinkyoo Park. Multi-Agent Actor-Critic with Hierarchical Graph Attention Network, 2020, AAAI(2020). https://ojs.aaai.org/index.php/AAAI/article/view/6214

21. Yali Du, Lei Han, Meng Fang, Ji Liu, Tianhong Dai, Dacheng Tao. LIIR: Learning Individual Intrinsic Reward in Multi-Agent Reinforcement Learning, 2019, NeurlIPS (2019).

22. Joel Z. Leibo, Edgar Duéñez-Guzmán, Alexander Sasha Vezhnevets, John P. Agapiou, Peter Sunehag, Raphael Koster, Jayd Matyas, Charles Beattie, Igor Mordatch, Thore Graepel. Scalable Evaluation of Multi-Agent Reinforcement Learning with Melting Pot, 2021, International Conference on Machine Learning 2021 (pp. 6187-6199).

23. Siqi Liu, Guy Lever, Josh Merel, Saran Tunyasuvunakool, Nicolas Heess, Thore Graepel. Emergent Coordination Through Competition, 2019, ICLR(2019). DOI: https://doi.org/10.48550/arXiv.1902.07151

24. Michael Bradley Johanson, Edward Hughes, Finbarr Timbers, Joel Z. Leibo. Emergent Bartering Behaviour in Multi-Agent Reinforcement Learning, 2022, DOI: https://doi.org/10.48550/arXiv.2205.06760.

25. John P. Agapiou, Alexander Sasha Vezhnevets, Edgar A. Duéñez-Guzmán, Jayd Matyas, Yiran Mao, Peter Sunehag, Raphael Köster, Udari Madhushani, Kavya Kopparapu, Ramona Comanescu, DJ Strouse, Michael B. Johanson, Sukhdeep Singh, Julia Haas, Igor Mordatch, Dean Mobbs, Joel Z. Leibo. Melting Pot 2.0, 2022, https://doi.org/10.48550/arXiv.2211.13746

Информация об авторах

Александр Николаевич Алфимцев – д. т. н., профессор, профессор кафедры «Информационные системы и телекоммуникации», Московский государственный технический университет им. Н. Э. Баумана (национальный исследовательский университет), г. Москва, e-mail: alfim@bmstu.ru.

Алексей Русланович Питикин – аспирант, преподаватель кафедры «Информационные системы и телекоммуникации», Московский государственный технический университет им. Н. Э. Баумана (национальный исследовательский университет), г. Москва, e-mail: pitikinar@bmstu.ru.

Authors

Alexander Nikolaevich Alfimtsev – Doctor of Technical Science, Professor, Department of Systems of Information and Telecommunications, Bauman Moscow State Technical University, Moscow, e-mail: alfim@bmstu.ru.

Alexey Ruslanovich Pitikin – postgraduate, Department of Systems of Information and Telecommunications, Bauman Moscow State Technical University, Moscow, e-mail: pitikinar@bmstu.ru.

Для цитирования

Алфимцев А.Н., Питикин А.Р. Эмерджентные свойства мультиагентного обучения с подкреплением // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2023. – №1(17). – С.1-10– DOI:

10.26731/2658-3704.2023.1(17).1-10 – Режим доступа: http://ismm-irgups.ru/toma/117-2023, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 31.03.2023)

For citations

Alfimtsev A.N., Pitikin A.R. Emergency properties of multi-agent reinforcement learning // Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: elektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2023. No. 1(17). P. 1-10. DOI: 10.26731/2658-3704.2023.1(17).1-10 [Accessed 31/03/23]