ТЕХНИЧЕСКИЕ НАУКИ

УДК 004.852.
DOI: 10.35330/1991-6639-2021-3-101-32-44
MSC: 68T05

КЛАССИФИКАЦИЯ ЗАДАЧ МУЛЬТИАГЕНТНОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

В.И. ПЕТРЕНКО

Федеральное государственное автономное образовательное учреждение высшего образования
«Северо-Кавказский федеральный университет»
355017, Ставропольский край, г. Ставрополь, ул. Пушкина, 1
E-mail: info@ncfu.ru

С появлением глубокого одноагентного обучения с подкреплением (ООП) мультиагентное обучение с подкреплением (МОП) получило новый толчок к развитию в виде глубокого МОП (ГМОП). Активное развитие методов данной области в течение последних нескольких лет актуализирует вопросы их систематизации и классификации. Существующие работы в качестве признаков классификации используют механизмы, применяемые в соответствующих методах ГМОП. Однако применимость того или иного метода определяется не только классом метода, но и классом задачи МОП. Целью данной работы являются формализация и классификация задач МОП. Для достижения цели выполнены математическая формализация и обобщение существующих классификаций задач ООП. Рассмотрены и математически формализованы особенности, возникающие при переходе от задачи ООП к задаче МОП. Выделены существенные признаки и выполнена классификация задач МОП на основе теоретико-множественного подхода. Использование теоретико-множественного подхода позволило выявить классы задач МОП, обобщаемые в других подобных работах, однако обладающие специфическими свойствами, что может быть использовано при разработке более эффективных методов решения таких задач МОП. Ожидается, что предложенные формализм и классификация задач МОП будут полезны исследователям в качестве инструмента постановки задачи и определения места исследования в общей структуре методов и задач МОП, а также разработчикам для обоснованного выбора методов МОП на основе класса решаемой задачи.

Ключевые слова: мультиагентное обучение с подкреплением, мультиагентные системы, классификация.

Поступила в редакцию 27.05.2021

Для цитирования. Петренко В.И. Классификация задач мультиагентного обучения с подкреплением // Известия Кабардино-Балкарского научного центра РАН. 2021. № 3 (101). С. 32-44.

1. Введение

Классические методы мультиагентного обучения с подкреплением (МОП, англ. *multiagent reinforcement learning – MARL*) зародились и активно исследовались еще в конце 20 века. Тем не менее в области МОП в последнее время наблюдается активный рост количества публикаций и разрабатываемых методов. Новая волна интереса исследователей и разработчиков к МОП вызвана развитием искусственных нейронных сетей (ИНС), а также успехами их применения в глубоком одноагентном обучении с подкреплением (ООП, англ. *reinforcement learning – RL*). Методы глубокого ООП демонстрируют эффективность выполнения интеллектуальных задач, сопоставимую с уровнем человека [1].

Универсальность и эффективность подхода глубокого ООП обусловило его расширение до глубокого МОП (ГМОП, англ. *multiagent deep reinforcement learning – MDRL*) для применения в мультиагентных системах (МАС).

Интерес к МАС обусловлен следующими причинами: применение МАС из более простых агентов вместо одного более сложного агента является экономически более эффективным [2]; децентрализованное решение задач с помощью МАС характеризуется более высокой эффективностью по сравнению с аналогичными централизованными методами [3]. Важными задачами, возникающими при использовании МАС, являются задачи управления поведением [4–6], коллективного принятия решений [2], распределения ресурсов [7], обеспечения безопасности [8] и надежности [9]. Многие из перечисленных задач могут быть решены с помощью методов ГМОП [10–18].

Разнообразие методов ГМОП привело к появлению ряда работ по классификации методов ГМОП [19–22]. Данные классификации основаны на используемых в методах механизмах, однако применимость того или иного метода ГМОП определяется не только его классом, но и классом задачи МОП.

Как следует из анализа, приведенного в разделе 6, наиболее полной моделью задачи МОП является децентрализованный частично наблюдаемый мультиагентный марковский процесс принятия решений. Несмотря на использование данной модели в различных работах, реализация взаимосвязей между элементами модели может сильно различаться. Различные задачи МОП могут отличаться такими важными признаками, как наличие коммуникации между агентами, кооперация или конкуренция между агентами, полная или частичная наблюдаемость среды и др. Данные признаки определяют эффективность и применимость различных методов МОП. Поэтому для применения методов МОП на практике необходимо учитывать не только их класс, но и класс задачи МОП. Как следует из раздела 6, данная проблема слабо освещена в существующих публикациях. С целью восполнения данного пробела в работе предложена классификация задач МОП, основанная на теоретико-множественном анализе математической модели задачи МОП.

Работа построена следующим образом. В разделе 2 приведена математическая модель задачи ООП. В разделе 3 приведена классификация задач ООП. В разделе 4 рассмотрены и формализованы системные связи, возникающие при переходе от задачи ООП к задаче МОП. В разделе 5 описана предлагаемая классификация задач МОП на основе теоретико-множественного анализа возникающих системных связей. В разделе 6 выполнено сравнение данной работы с аналогами, описаны отличия и научный вклад данной работы.

2. Математическая модель задачи ООП

ООП нацелено на формирование методом проб и ошибок у агента поведения, необходимого для достижения заданной цели. Согласно схеме обучения с подкреплением (рис. 1), агент (англ. *agent*) взаимодействует со средой (англ. *environment*), наблюдая среду в момент времени t в виде некоторого наблюдения o (англ. *observation*), на основе которого он предпринимает действие a (англ. *action*) и получает награду r (англ. *reward*), зависящую от результатов воздействия на среду. Получаемая награда r в соответствии с алгоритмом обучения с подкреплением увеличивает или уменьшает вероятность совершения действия a при тех же условиях в будущем.

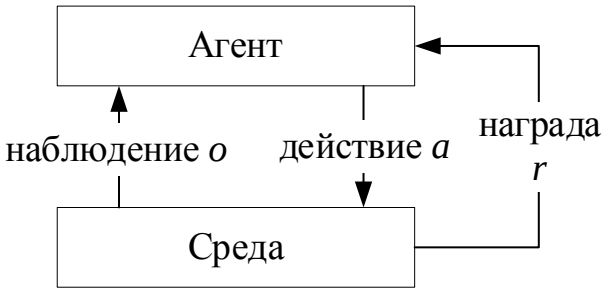

Рис. 1. *Схема ООП*

Для моделирования ООП используется дискретная рекуррентная математическая модель задачи ООП, описываемая для момента времени t следующими свойствами:

$$(S, A, \tau, O, \sigma, R, \rho), \tag{1}$$

где

$$S = \{s\} \tag{2}$$

есть множество возможных состояний среды;

$$A = \{a\} \tag{3}$$

есть множество возможных действий агента;

$$\tau(s_t, a_t, s_{t+1}) : S \times A \times S \rightarrow [0,1] \tag{4}$$

есть функция перехода (англ. *transition*), возвращающая вероятность перехода среды из состояния s_t в состояние s_{t+1} при совершении агентом действия a_t;

$$O = \{o\} \tag{5}$$

есть множество возможных наблюдений среды, доступное агенту;

$$\sigma(s_t, o_t): S \times O \rightarrow [0,1] \tag{6}$$

есть функция наблюдения, возвращающая вероятность получения агентом наблюдения o_t в состоянии среды s_t;

$$R = \{r\} \tag{7}$$

есть множество возможных значений вознаграждения агента;

$$\rho(s_t, a_t, s_{t+1}): S \times A \times S \rightarrow R \tag{8}$$

есть функция вознаграждения, получаемого агентом при совершении в состоянии среды s_t действия a_t и переходе среды в состояние s_{t+1}.

Вычислительная схема используемой модели задачи ООП представлена на рисунке 2. Все блоки на схеме, кроме блока s_t, выполняются без задержки. Блок s_t хранит текущее состояние среды s_t и имеет внутреннюю задержку, симулирующую дискретизацию времени. На схеме выражение $p(x)$ обозначает плотность вероятности величины x, блок «rnd», на вход которого подается величина $p(x)$, обозначает выборку случайного значения величины x с заданным распределением $p(x)$.

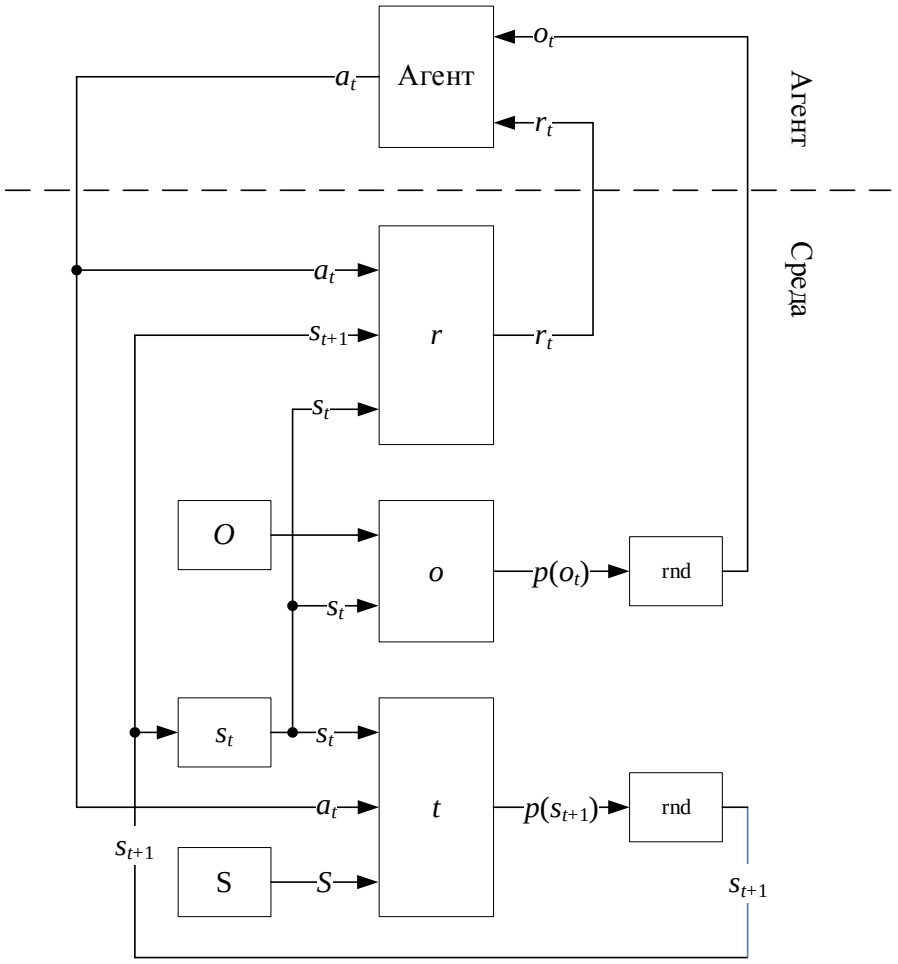

Рис. 2. *Вычислительная схема математической модели задачи ООП*

3. Классификация задач ООП

Используемые далее классы задач ООП определяются конкретной реализацией элементов математической модели (1).

По виду множеств S, A, O, R задачи ООП делятся на классы с дискретными/непрерывными множествами состояний, действий, наблюдений и награды. Несмотря на то, что в большинстве работ множество R возможных значений вознаграждения агента считается эквивалентным множеству $\mathbb{R}$ рациональных чисел: $R \equiv \mathbb{R}$, в строгом понимании данная эквивалентность не выполняется для рассматриваемых задач.

По характеру распределения, возвращаемого функцией перехода τ, различают детерминированные и недетерминированные (стохастические) задачи ООП. В детерминированных задачах ООП

$$t(s_t, a_t, s_{t+1}) : S \times A \times S \rightarrow \{0,1\}, \tag{9}$$

т. е. при совершении агентом действия a_t в состоянии среды s_t среда либо гарантированно перейдет в какой-либо вариант состояния s_{t+1}, либо гарантированно не перейдет в него. В стохастических задачах ООП

$$t(s_t, a_t, s_{t+1}) : S \times A \times S \rightarrow [0,1], \tag{10}$$

т. е. в общем случае при совершении агентом действия a_t в состоянии среды s_t возможен переход среды в один из нескольких вариантов состояния s_{t+1}.

По неизменности распределения, возвращаемого функцией перехода τ для любых моментов времени t_1, t_2, различают стационарные задачи ООП, в которых выполняется равенство

$$\tau_{t_1}(s_t, a_t, s_{t+1}) = \tau_{t_2}(s_t, a_t, s_{t+1})\ \forall t_1, t_2, \quad (11)$$

где t_1 и t_2 – произвольные моменты времени, и нестационарные, в которых это равенство не выполняется.

Состояние среды s_t и наблюдаемое состояние среды o_t являются упорядоченными множествами параметров. По соотношению между состоянием среды s_t и наблюдаемым состоянием среды o_t различают задачи ООП с полной наблюдаемостью:

$$O \equiv S \quad (12)$$

и задачи ООП с частичной наблюдаемостью:

$$\exists s\ o \subset s\ \forall o.$$

По виду функции наблюдения различают задачи ООП с наблюдением среды без шума:

$$\mathcal{O}(s, o) = \{0; 1\}\ \forall s, o \quad (13)$$

и задачи ООП с зашумленным наблюдением среды:

$$\exists s, o\ \mathcal{O}(s, o) \neq \{0; 1\}. \quad (14)$$

По виду функции вознаграждения задачи ООП бывают эпизодическими, в которых агент получает награду только за достижение одного из состояний, принадлежащих множеству конечных (англ. *end*) состояний S_e:

$$\mathcal{r}(s_t, a_t, s_{t+1}) \neq 0 \Leftrightarrow s_{t+1} \in S_e \subset S, \quad (15)$$

после чего цель агента считается достигнутой, либо ее достижение становится невозможным. Примерами конечных состояний могут являться победа/поражение в шахматах, достижение БПЛА заданной позиции или его крушение. В случае получения агентом награды не только в конечном состоянии $s \in S_e$ задача ООП классифицируется как продолжительная. Также возможна комбинированная функция награды.

4. Математическая модель задачи МОП

Обозначим переход от скалярной величины к вектору как векторизацию. При переходе от задачи ООП к задаче МОП в математической модели происходит переход от скалярных величин o_t, a_t, r_t к векторам $\boldsymbol{o_t}, \boldsymbol{a_t}, \boldsymbol{r_t}$ (рис. 3). На рисунке 3 векторизованные величины и новые связи обозначены линиями зеленого цвета.

Переход от задачи ООП к задаче МОП обусловливает следующие изменения. Множество состояний среды S экспоненциально возрастает с ростом количества агентов, т. к. в состояние среды, помимо состояния x среды без агентов, включается состояние s_i физической составляющей каждого i-го агента:

$$S = \{s\} = \{\langle x, s_1, \dots s_n | s_i \in S_i, i = \overline{1, n}\rangle\}, \quad (16)$$

где S_i – множество возможных состояний i-го агента.

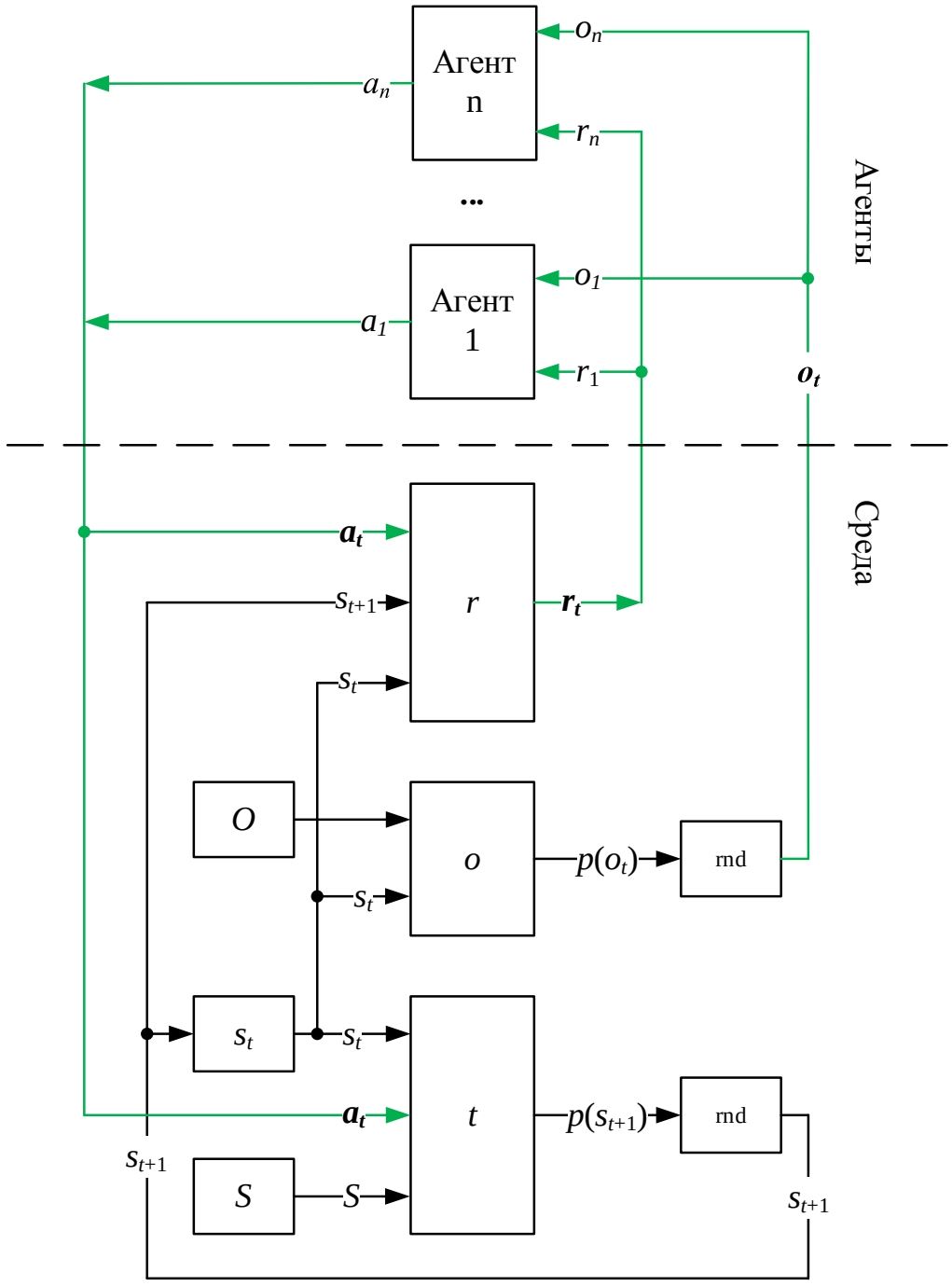

Рис. 3. *Вычислительная схема математической модели задачи МОП*

Множество доступных действий A одного агента становится семейством $\boldsymbol{A}$ множеств A_i доступных действий i-го агента:

$$\boldsymbol{A} = \{a\} = \{\langle a_i | a_i \in A_i, i = \overline{1,n}\rangle\}. \quad (17)$$

Аналогичным образом при переходе от задачи ООП к задаче МОП происходит переход от множеств O и R к семействам множеств $\boldsymbol{O}$ и $\boldsymbol{R}$.

Мощность множества $A_i = \{a_i\}$ возможных действий отдельного i-го агента увеличивается за счет дополнительных возможностей по взаимодействию агентов между собой.

Количество составляющих отдельного элемента o_i множества O_i возможных наблюдений среды i-го агента увеличивается, т. к. в большинстве случаев агенты наблюдают не только состояние x среды без агентов, но и состояния других агентов $s_i, i = \overline{1,n}$.

Функция перехода подвергается векторизации:

$$t(s_t, \boldsymbol{a_t}, s_{t+1}) : S \times \boldsymbol{A} \times S \rightarrow [0,1], \quad (18)$$

где $\boldsymbol{a_t} = (a_1^t, \dots, a_n^t), a_i^t \in A_i$ – коллективное действие всех $i = \overline{1,n}$ агентов, оказываемое на среду в момент времени t, где n – количество агентов. Аналогичным образом векторизуются функции σ и r.

5. Классификация задач МОП

Обучение с подкреплением включает в себя две стадии: стадию обучения и стадию функционирования. Чтобы разграничить свойства, присущие задаче МОП, и свойства, присущие методу МОП, в данной работе при классификации задач МОП учитываются только те свойства, которые реализуются на стадии функционирования. Например, если агенты на стадии функционирования не обладают сведениями о действиях, которые собираются предпринять другие агенты, то задача МОП будет классифицирована как задача без знаний агентов о планируемых действиях других агентов (описание класса приводится далее). В то же время данная информация может дополнительно использоваться методом МОП на стадии обучения, как, например, в работе [10]. Данный пример иллюстрирует тот факт, что особенности, присутствующие на стадии обучения и отсутствующие на стадии функционирования, следует отнести к особенностям метода, а не задачи МОП.

Целью данной работы является классификация задач МОП с точки зрения особенностей их математической модели, поэтому при классификации не рассматриваются такие свойства задачи МОП, как область применения или численность агентов МАС.

Векторизация связей в математической модели при переходе от задачи ООП к задаче МОП обусловливает возникновение следующих новых классов (табл. 1).

Таблица 1

Классы задач МОП

Признак классификации	**Классы задач МОП**
Одновременность выполнения действий a_i	С последовательными действиями, с одновременными действиями, комбинированные
Индивидуализация награды агентов	С общей наградой, с индивидуальной наградой
Информированность агентов о планах других агентов	Со знаниями агентов о планируемых действиях других агентов, без знаний агентов о планируемых действиях других агентов
Наличие в наблюдении информации, генерируемой другими агентами	С коммуникацией, без коммуникации
Доступность информации обо всех других агентах	Централизованные, децентрализованные
Конкуренция за получение награды	Конкурентные, кооперативные, смешанные
Физическая гомогенность агентов МАС	Гомогенные, гетерогенные
Ролевая гомогенность агентов МАС	Гомогенные, гетерогенные

По признаку одновременности выполнения действий a_i задачи МОП подразделяются на одновременные, где действия всех агентов выполняются одновременно, и последовательные, где действия всех агентов выполняются по очереди, а также комбинированные.

По признаку индивидуализации награды задачи МОП классифицируются на задачи с общей наградой (рис. 4 а), для которых выполняется условие

$$r_i = r_j \; \forall r_i, r_j \in \boldsymbol{r}, \tag{19}$$

и задачи с индивидуальной наградой (рис. 4 б), для которых условие (19) не выполняется.

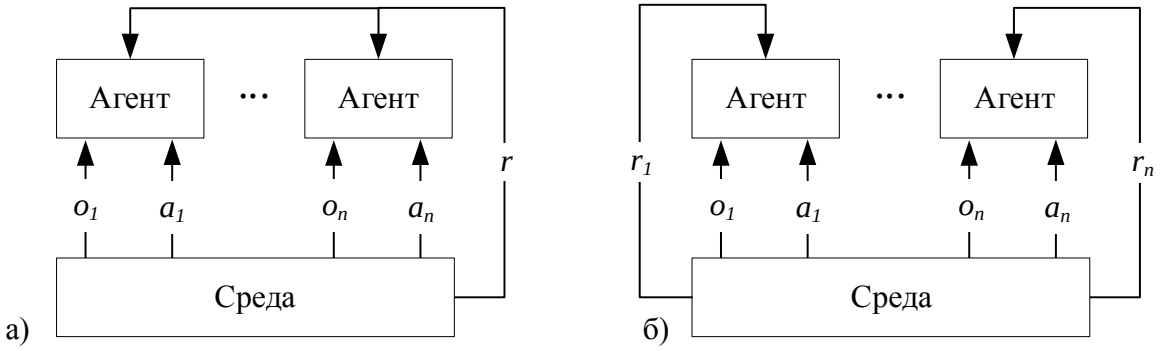

Рис. 4. *Классификация задачи МОП по признаку индивидуальной награды*

По наличию в наблюдении o_i информации о действиях, которые планируют предпринять другие агенты, задачи МОП классифицируются как задачи со знаниями агентов о планируемых действиях других агентов и задачи без знаний агентов о планируемых действиях других агентов.

По наличию в наблюдении o_i информации, генерируемой другими агентами, задачи МОП подразделяются на задачи с коммуникацией и задачи без коммуникации.

По доступности в наблюдении o_i информации обо всех других агентах задачи МОП делятся на централизованные и децентрализованные.

По признаку конкуренции за получение награды задачи МОП можно разделить на кооперативные, конкурентные и смешанные. В кооперативных задачах агенты стремятся максимизировать общее долгосрочное вознаграждение. Соревновательные задачи могут рассматриваться как игры с нулевой суммой. В качестве примера задачи МОП со смешанным получением награды можно рассматривать командно-конкурентные задачи.

По наличию гомогенности множеств доступных действий и наблюдений задачи МОП подразделяются на задачи с физически гомогенными и физически гетерогенными агентами. Дополнительно по гомогенности функции награды задачи МОП могут подразделяться на гомогенные и гетерогенные с точки зрения ролей, выполняемых агентами.

6. Анализ релевантных работ

Несмотря на то, что в различных работах используются схожие математические модели задачи МОП, в каждой работе разнятся детали реализации элементов кортежа из выражения (1). В данном разделе приведен анализ работ, легший в основу модели и классификации задач МОП, приведенных в разделах 4 и 5.

В процессе развития методов ОП модель задач ООП, а затем МОП претерпевала эволюционное изменение. Модель марковского процесса принятия решений (МППР) была расширена до мультиагентного марковского процесса принятия решений (ММППР), именуемого в различных работах также как марковские/стохастические игры. Модель ММППР используется в работах [15, 16, 20–30]. В дальнейшем модель ММППР была расширена до децентрализованного частично наблюдаемого ММППР (ДММППР). Модель ДММППР используется значительно реже [11, 12].

Непосредственно моделирование и классификация задач МОП рассматриваются в работах [19–22, 28]. В работах [11–13, 15, 16, 23–27, 31] рассматриваются только частные реализации модели задач МОП в рамках разработки специфических методов их решения.

В работе [19] основной акцент сделан на классификации методов МОП. Задачи МОП в работе [19] классифицируются только по признаку индивидуализации награды агентов. В работе [20] в качестве признаков классификации рассматривается конкуренция за получение награды. В работе [21] задачи МОП классифицируются по критерию наличия в наблюдении информации, генерируемой другими агентами. В работе [22] задачи МОП классифицируются по критерию индивидуализации награды агентов и доступности информации обо всех других агентах. В работе [28] рассматриваются следующие признаки классификации задач МОП: индивидуализация награды агентов, информированность агентов о планах других агентов, физическая и ролевая гомогенность агентов MAC.

В результате проведенного анализа обзорных и специфических работ по МОП и ГМОП было выявлено, что в большинстве публикаций классификация методов ГМОП выполняется на основе используемых в методах принципов, без привязки к классам решаемых задач МОП. Вопросы классификации задач МОП затрагиваются лишь поверхностно. Таким образом, новизна работы заключается в более глубокой классификации задач МОП на основе теоретико-множественного анализа их математической модели.

Заключение

В работе на основе математической модели и классификации задач ООП были рассмотрены качественные и количественные изменения, возникающие при переходе от задачи ООП к задаче МОП. Предложена математическая модель задачи МОП, обобщающая модели, используемые в существующих работах. Предложена классификация задач МОП на основе теоретико-множественного анализа математической модели МОП.

Проведенный сравнительный анализ показал, что предложенная классификация задач МОП является более подробной, чем классификации, найденные в аналогичных работах. Ожидается что предложенные модель и классификация задач МОП будут полезны исследователям и разработчикам при постановке исследования или выборе способа решения поставленных задач.

ЛИТЕРАТУРА

1. *Mnih V. et al.* Human-level control through deep reinforcement learning // Nature. Nature Publishing Group, 2015. Vol. 518. № 7540. P. 529–533.

2. *Petrenko V.I., Tebueva F.B., Ryabtsev S.S., Gurchinsky M.M., Struchkov I. V.* Consensus achievement method for a robotic swarm about the most frequently feature of an environment // IOP Conference Series: Materials Science and Engineering. 2020. Vol. 919, № 4.

3. *Kovács G., Yussupova N., Rizvanov D.* Resource management simulation using multi-agent approach and semantic constraints // Pollack Period. 2017. Vol. 12, № 1.

4. *Пшихопов В.Х., Медведев М.Ю.* Групповое управление движением мобильных роботов в неопределенной среде с использованием неустойчивых режимов // Труды СПИИРАН. 2018. Том 60. № 5. С. 39–63.

5. *Тугенгольд А.К., Лукьянов Е.А.* Интеллектуальные функции и управление автономными технологическими мехатронными объектами. Ростов-на-Дону: Донской государственный технический университет, 2013. 203 с.

6. *Mironov K. V., Pongratz M. U.* Applying neural networks for prediction of flying objects trajectory // Vestn. UGATU. 2013. № 6.

7. *Даринцев О.В., Мигранов А.Б.* Распределенная система управления группами мобильных роботов // Вестник УГАТУ. 2017. Том 2 № 76.

8. *Петренко В.И., Тебуева Ф.Б., Гурчинский М.М., Рябцев С.С.* Анализ технологий обеспечения информационной безопасности мультиагентных робототехнических систем с роевым интеллектом // Наука и бизнес: пути развития. 2020. № 4 (106). С. 96–99.

9. *Yusupova N., Rizvanov D., Andrushko D.* Cyber-Physical Systems and Reliability Issues // Proceedings of the 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020). Atlantis Press, 2020. P. 133–137.

10. *Lowe R. et al.* Multi-agent actor-critic for mixed cooperative-competitive environments // Advances in Neural Information Processing Systems. 2017. Vol. 2017-December.

11. *Wang H., Liu Z., Yi J., Pu Z.* Multiagent hierarchical cognition difference policy for multiagent cooperation // Algorithms. 2021. Vol. 14. № 3.

12. *Silva F.L. Da, Nishida C.E.H., Roijers D.M., Costa A.H.R.* Coordination of Electric Vehicle Charging through Multiagent Reinforcement Learning // IEEE Trans. Smart Grid. 2020. Vol. 11. № 3.

13. *Cui J., Liu Y., Nallanathan A.* Multi-Agent Reinforcement Learning-Based Resource Allocation for UAV Networks // IEEE Trans. Wirel. Commun. 2020. Vol. 19. № 2.

14. *Shamsoshoara A., Khaledi M., Afghah F., Razi A., Ashdown J.* Distributed cooperative spectrum sharing in UAV networks using multi-agent reinforcement learning // arXiv. 2018.

15. *Qie H. et al.* Joint Optimization of Multi-UAV Target Assignment and Path Planning Based on Multi-Agent Reinforcement Learning // IEEE Access. 2019. Vol. 7.

16. *Fang X. et al.* Multi-agent reinforcement learning approach for residential microgrid energy scheduling // Energies. 2019. Vol. 13. № 1.

17. *Пшенокова И.А., Сундуков З.А.* Разработка имитационной модели сценарного прогнозирования поведения интеллектуального агента на основе инварианта рекурсивной мультиагентной нейрокогнитивной архитектуры // Известия Кабардино-Балкарского научного центра РАН. 2020. № 6(98). С. 80–90.

18. *Пшенокова И.А., Нагоева О.В., Гуртуева И.А., Айран А.А.* Алгоритм обучения интеллектуальной системы принятия решений на основе мультиагентных нейрокогнитивных архитектур // Известия Кабардино-Балкарского научного центра РАН. 2020. № 3(95). С. 23–31.

19. *Hernandez-Leal P., Kartal B., Taylor M.E.* A survey and critique of multiagent deep reinforcement learning // Auton. Agent. Multi. Agent. Syst. 2019. Vol. 33. № 6.

20. *Buşoniu L., Babuška R., De Schutter B.* A comprehensive survey of multiagent reinforcement learning // IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews. 2008. Vol. 38. № 2.

21. *Hernandez-Leal P., Kaisers M., Baarslag T., De Cote E.M.* A survey of learning in multiagent environments: Dealing with non-stationarity // arXiv. 2017.

22. *Zhang K., Yang Z., Başar T.* Multi-agent reinforcement learning: A selective overview of theories and algorithms // arXiv. 2019.

23. *Hao J., Huang D., Cai Y., Leung H. fung.* The dynamics of reinforcement social learning in networked cooperative multiagent systems // Eng. Appl. Artif. Intell. 2017. Vol. 58.

24. *Da Silva F.L., Reali Costa A.H.* A survey on transfer learning for multiagent reinforcement learning systems // J. Artif. Intell. Res. 2019. Vol. 64.

25. *Nguyen T.T., Nguyen N.D., Nahavandi S.* Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications // IEEE Trans. Cybern. 2020. Vol. 50. № 9.

26. *Yang Y. et al.* Q-value path decomposition for deep multiagent reinforcement learning // arXiv. 2020.

27. *Shamsoshoara A., Khaledi M., Afghah F., Razi A., Ashdown J.* Distributed Cooperative Spectrum Sharing in UAV Networks Using Multi-Agent Reinforcement Learning // 2019 16th IEEE Annual Consumer Communications and Networking Conference, CCNC 2019. 2019.

28. *Tuyls K., Weiss G.* Multiagent learning: Basics, challenges, and prospects // AI Magazine. 2012. Vol. 33. № 3.

29. *Matignon L., Laurent G.J., Le Fort-Piat N.* Independent reinforcement learners in cooperative Markov games: A survey regarding coordination problems // Knowledge Engineering Review. 2012. Vol. 27. № 1.

30. *Littman M.L.* Markov games as a framework for multi-agent reinforcement learning Michael // Thromb. Res. 2007. Vol. 120. № 1.

31. *Tampuu A. et al.* Multiagent cooperation and competition with deep reinforcement learning // PLoS One. 2017. Vol. 12. № 4.

REFERENCES

1. Mnih V. et al. Human-level control through deep reinforcement learning // Nature. Nature Publishing Group, 2015. Vol. 518. № 7540. Pp. 529–533.

2. Petrenko V.I., Tebueva F.B., Ryabtsev S.S., Gurchinsky M.M., Struchkov I. V. Consensus achievement method for a robotic swarm about the most frequently feature of an environment // IOP Conference Series: Materials Science and Engineering. 2020. Vol. 919. № 4.

3. Kovács G., Yussupova N., Rizvanov D. Resource management simulation using multi-agent approach and semantic constraints // Pollack Period. 2017. Vol. 12. № 1.

4. Pshikhopov V. Kh., Medvedev M. Yu. *Gruppovoe upravlenie dvizheniem mobil'nyh robotov v neopredelennoj srede s ispol'zovaniem neustojchivyh rezhimov* [Group motion control of mobile robots in an uncertain environment using unstable modes]. Proceedings of SPIIRAS. 2018. Vol. 60. № 5. Pp. 39–63.

5. Tugengold A. K., Lukyanov E. A. *Intellektual'nye funkcii i upravlenie avtonomnymi tekhnologicheskimi mekhatronnymi ob"ektami* [Intelligent functions and control of autonomous technological mechatronic objects]. Rostov-on-Don: Don State Technical University. 2013. Pp. 203.

6. Mironov K. V., Pongratz M. U. Applying neural networks for prediction of flying objects trajectory // Vestn. UGATU. 2013. № 6.

7. Darintsev O. V., Migranov A. B. *Raspredelennaya sistema upravleniya gruppami mobil'nyh robotov* [Distributed control system for groups of mobile robots]. Vestnik USATU. 2017. Vol. 2. № 76.

8. Petrenko V.I., Tebueva F.B., Gurchinsky M.M., Ryabtsev S.S. *Analiz tekhnologij obespecheniya informacionnoj bezopasnosti mul'tiagentnyh robototekhnicheskih sistem s roevym intellektom* [Analysis of information security technologies for multi-agent robotic systems with swarm intelligence]. Science and business development paths. 2020. No. 4 (106). Pp. 96–99.

9. Yusupova N., Rizvanov D., Andrushko D. Cyber-Physical Systems and Reliability Issues // Proceedings of the 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020). Atlantis Press, 2020. Pp. 133–137.

10. Lowe R. et al. Multi-agent actor-critic for mixed cooperative-competitive environments // Advances in Neural Information Processing Systems. 2017. Vol. 2017-December.

11. Wang H., Liu Z., Yi J., Pu Z. Multiagent hierarchical cognition difference policy for multiagent cooperation // Algorithms. 2021. Vol. 14. № 3.

12. Silva F.L. Da, Nishida C.E.H., Roijers D.M., Costa A.H.R. Coordination of Electric Vehicle Charging through Multiagent Reinforcement Learning // IEEE Trans. Smart Grid. 2020. Vol. 11. № 3.

13. Cui J., Liu Y., Nallanathan A. Multi-Agent Reinforcement Learning-Based Resource Allocation for UAV Networks // IEEE Trans. Wirel. Commun. 2020. Vol. 19. № 2.

14. Shamsoshoara A., Khaledi M., Afghah F., Razi A., Ashdown J. Distributed cooperative spectrum sharing in UAV networks using multi-agent reinforcement learning // arXiv. 2018.

15. Qie H. et al. Joint Optimization of Multi-UAV Target Assignment and Path Planning Based on Multi-Agent Reinforcement Learning // IEEE Access. 2019. Vol. 7.

16. Fang X. et al. Multi-agent reinforcement learning approach for residential microgrid energy scheduling // Energies. 2019. Vol. 13. № 1.

17. Pshenokova I.A., Sundukov Z.A. *Razrabotka imitatsionnoy modeli stsenarnogo prognozirovaniya povedeniya intellektual'nogo agenta na osnove invarianta rekursivnoy mul'ti-agentnoy neyrokognitivnoy arkhitektury* [Development of a simulation model for predicting the behavior of an intelligent agent based on an invariant of a recursive multi-agent neurocognitive architecture] // *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [News of the Kabardino-Balkarian Scientific Center of the RAS]. 2020. № 6(98). Pp. 80–90.

18. Pshenokova I.A., Nagoeva O.V., Gurtueva I.A., Airan A.A. *Algoritm obucheniya intel-lektual'noy sistemy prinyatiya resheniy na osnove mul'tiagentnykh neyrokognitivnykh arkhitektur* [Learning algorithm for an intelligent decision making system based on multi-agent neurocognitive architectures] // *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [News of the Kabardino-Balkarian Scientific Center of the RAS]. 2020. № 3(95). Pp. 23–31.

19. Hernandez-Leal P., Kartal B., Taylor M.E. A survey and critique of multiagent deep reinforcement learning // Auton. Agent. Multi. Agent. Syst. 2019. Vol. 33. № 6.

20. Buşoniu L., Babuška R., De Schutter B. A comprehensive survey of multiagent reinforcement learning // IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews. 2008. Vol. 38. № 2.

21. Hernandez-Leal P., Kaisers M., Baarslag T., De Cote E.M. A survey of learning in multiagent environments: Dealing with non-stationarity // arXiv. 2017.

22. Zhang K., Yang Z., Başar T. Multi-agent reinforcement learning: A selective overview of theories and algorithms // arXiv. 2019.

23. Hao J., Huang D., Cai Y., Leung H. fung. The dynamics of reinforcement social learning in networked cooperative multiagent systems // Eng. Appl. Artif. Intell. 2017. Vol. 58.

24. Da Silva F.L., Reali Costa A.H. A survey on transfer learning for multiagent reinforcement learning systems // J. Artif. Intell. Res. 2019. Vol. 64.

25. Nguyen T.T., Nguyen N.D., Nahavandi S. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications // IEEE Trans. Cybern. 2020. Vol. 50. № 9.

26. Yang Y. et al. Q-value path decomposition for deep multiagent reinforcement learning // arXiv. 2020.

27. Shamsoshoara A., Khaledi M., Afghah F., Razi A., Ashdown J. Distributed Cooperative Spectrum Sharing in UAV Networks Using Multi-Agent Reinforcement Learning // 2019 16th IEEE Annual Consumer Communications and Networking Conference, CCNC 2019. 2019.

28. Tuyls K., Weiss G. Multiagent learning: Basics, challenges, and prospects // AI Magazine. 2012. Vol. 33. № 3.

29. Matignon L., Laurent G.J., Le Fort-Piat N. Independent reinforcement learners in cooperative Markov games: A survey regarding coordination problems // Knowledge Engineering Review. 2012. Vol. 27. № 1.

30. Littman M.L. Markov games as a framework for multi-agent reinforcement learning Michael // Thromb. Res. 2007. Vol. 120. № 1.

31. Tampuu A. et al. Multiagent cooperation and competition with deep reinforcement learning // PLoS One. 2017. Vol. 12. № 4.

CLASSIFICATION OF MULTI-AGENT REINFORCEMENT LEARNING PROBLEMS

V.I. PETRENKO

Federal State Autonomous Educational Institution for Higher Education
"North-Caucasus Federal University"
355017, Stavropol region, Stavropol, 1 Pushkin str.
E-mail: info@ncfu.ru

With the advent of deep single-agents reinforcement learning (SARL), multi-agent reinforcement learning (MARL) has received a new impetus for development in the form of deep multi-agent reinforcement learning (MDRL). The active development of methods in this area over the past few years has actualized the issues of their systematization and classification. Existing works use the mechanisms used in the corresponding MDRL methods as classification signs. However, the applicability of a particular method is determined not only by the class of the method, but also by the class of the MARL problem. The purpose of this work is to formalize and classify MARL tasks. To achieve the goal, the mathematical formalization and generalization of the existing classifications of SARL tasks is carried out. The peculiarities arising in the transition from the SARL problem to the MARL problem are considered and mathematically formalized. The essential features are highlighted and the classification of MARL tasks is carried out on the basis of the set-theoretic approach. The use of the set-theoretic approach made it possible to identify classes of MARL problems, generalized in other similar works, but possessing specific properties, which can be used to develop more efficient methods for solving such MARL problems. It is expected that the proposed formalism and classification of MARL problems will be useful for researchers as a tool for setting a problem and determining the place of research in the general structure of MARL methods and tasks, and will also be useful for developers for a reasonable choice of MARL methods based on the class of the problem being solved.

Keywords: multi-agent reinforcement learning, multi-agent systems, classification.

Received by the editors 27.05.2021

For citation. Petrenko V.I. Classification of multi-agent reinforcement learning problems // News of the Kabardino-Balkarian Scientific Center of RAS. 2021. No. 3 (101). Pp. 32-44.

Сведения об авторе:

Петренко Вячеслав Иванович, к.т.н., доцент, зав. кафедрой организации и технологии защиты информации Северо-Кавказского федерального университета.
355017, Ставропольский край, г. Ставрополь, ул. Пушкина, 1.
E-mail: vip.petrenko@gmail.com.

Information about the author:

Petrenko Vyacheslav Ivanovich, Candidate of Technical Sciences**,** Associate Professor, Head of the Department of Organization and Technology of Information Security. Federal State Autonomous Educational Institution for Higher Education "North-Caucasus Federal University".
355017, Stavropol region, Stavropol, 1 Pushkin str.
E-mail: vip.petrenko@gmail.com.