DOI: 10.34220/issn.2222-7962/2020.1/19
УДК 519*6

ПРИМЕНЕНИЕ МЕТОДА ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В РОБОТИЗИРОВАННЫХ И АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ ЛЕСНОЙ ПРОМЫШЛЕННОСТИ

А.А. Толстых[1]
кандидат технических наук **Д.С. Ступников**[2]
кандидат технических наук **С.В. Малюков**[2]
кандидат технических наук **А.С. Лукьянов**[1]
кандидат технических наук **Ю.С. Лунёв**[1]

1 – ФГКОУ ВО «Воронежский институт Министерства внутренних дел Российской Федерации», г. Воронеж, Российская Федерация
2 – ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова», г. Воронеж, Российская Федерация

В настоящее время на большинстве крупных предприятий активно используются промышленные роботы и другие автоматизированные решения. Это позволяет в значительной степени повысить производительность и качество выполняемых работ. В данной статье был дан краткий обзор современных промышленных роботов, их принцип работы, основные узлы и системы. Был разработан и протестирован алгоритм обучения с подкреплением. Задача построения алгоритма обучения с подкреплением была разделена на два этапа: моделирование среды и описание и оптимизация функции стоимости. Так как промышленные робототехнические системы работают в реальном мире, модель окружения должна отражать основные физические законы. Поэтому в качестве физической среды для тестирования была выбрана библиотека физического окружения pyBullet. После моделирования манипулятора в выбранной физической среде перед ним была поставлена тривиальная задача – касание захватом манипулятора заданного объекта. В качестве агента, взаимодействующего со средой, использовалась искусственная нейронная сеть. Входами служили координаты объекта и существующие углы поворотов шарнирных сочленений робота. Выходами – угол поворота сочленений на данном шаге. Данная сеть обучалась методом обратного распространения ошибки, модификацией Adam. Система обучалась около 12 часов. При тестировании устойчивости системы (случайное положение цилиндра) успех достигается в 95 % случаев. В дальнейшем планируется тестирование полученных моделей на стендовых образцах.

Ключевые слова: промышленный робот, нейронная сеть, алгоритм, обучение с подкреплением, автоматизация, робот-манипулятор

APPLICATION OF LEARNING REINFORCEMENT METHOD IN ROBOTIZED AND AUTOMATED FORESTRY SYSTEMS

A.A. Tolstykh[1]
PhD (Engineering) **D.S. Stupnikov**[2]
PhD (Engineering) **S.V. Malyukov**[2]
PhD (Engineering) **A.S. Lukyanov**[1]
PhD (Engineering) **Yu.S. Lunev**[1]

1 – FSBEI HE "Voronezh Institute of the Ministry of Internal Affairs of the Russian Federation", Voronezh, Russian Federation
2 – FSBEI HE "Voronezh State University of Forestry and Technologies named after G.F. Morozov", Voronezh, Russian Federation

Abstract

Currently, most large enterprises are actively using industrial robots and other automated solutions. This allows a significant increase in productivity and quality of work performed. This article gave a brief overview of modern industrial robots, their operating principle, basic components and systems. A reinforcement learning algorithm was developed and tested. The task of constructing a learning algorithm with reinforcement was divided into two stages: modeling the environment and description and optimization of the cost function. Since industrial robotic systems operate in the real world, the environment model should reflect basic physical laws. Therefore, the pyBullet library of the physical environment was chosen as the physical environment for testing. After modeling the manipulator in the selected physical medium, it was given the trivial task of touching a given object with the capture of the manipulator. An artificial neural network was used as an agent interacting with the environment. The inputs were the coordinates of the object and the existing angles of rotation of the articulated joints of the robot. Outputs - angle of rotation of joints at this step. This network was trained using the back propagation method, Adam modification. The system was trained for about 12 hours. Success is achieved in 95 % of cases when testing the stability of the system (random position of the cylinder). In future, it is planned to test the obtained models on bench samples.

Keywords: industrial robot, neural network, algorithm, reinforcement learning, automation, robot-manipulator

Введение

В настоящее время практически все крупные предприятия любой промышленной сферы стараются использовать множество различных автоматизированных решений. При использовании подобных средств человеческий фактор практически полностью исключается из производственного процесса. В лесной промышленности подобная тенденция тоже прослеживается, начиная от полуавтоматизированных лесозаготовительных комплексов по типу харвестера от John Deere, заканчивая автоматизированными станками с ЧПУ и манипуляторами для погрузки хлыстов и сортимента. Использование такого рода техники позволяет в значительной степени повысить производительность и качество выполняемых работ, а также сократить количество используемой техники [2].

В данной статье хотелось бы затронуть вопрос современных промышленных роботов, их возможностей, структуры, а также программной составляющей для их управления. Также приведем некоторые исследования в данной области.

Материал и методы исследования

Зачастую промышленный робот представляется в антропоморфной форме (аналог человеческой руки). Подобная структура является наиболее популярной и неспроста. Это позволяет сделать конструкцию более универсальной и

иметь несколько степеней свободы (обычно от 4 и более).

В промышленной сфере большим спросом пользуются полностью автоматизированные роботы-манипуляторы, которые выполняют определенный спектр поставленных задач при помощи алгоритмов и разного рода датчиков. Такие роботы могут сами принимать решения, но все они будут в рамках прописанных алгоритмом.

Однако специфика и спектр различных применений промышленных роботов подразумевает разработку и создание промышленных роботов с использованием нейронных сетей, способных моделировать свое виртуальное пространство, в котором могут ориентироваться и принимать решения о последующих действиях. Такие роботы могут обучаться сами, по мере поступления опыта [1]. Яркими примерами являются промышленные роботы фирмы KUKA и AMAZON (рис. 1).

Конструкцию современного промышленного робота можно разделить на две основные части: механическую и аппаратную.

Механическая часть включает в себя звенья манипулятора, систему приводов и т. д. Аппаратная часть включает в себя систему управления приводами механической части, а также информационно-сенсорную систему, которая обеспечивает робот необходимой вводной информацией. Объем и сложность аппаратной части во многом зависит от области использования того или иного робота.

Функциональная схема робота-манипулятора показана на рис. 2.

Исполнительным механизмом манипулятора обычно является открытая кинематическая цепь. При этом каждое звено цепи соединяется последовательно между собой вращательными либо поступательными сочленениями.

а)

б)

Рис. 1. Антропоморфные роботы – манипуляторы: а) образец компании KUKA; Источник: компания KUKA. URL: https://www.kuka.com/; б) образец компании AMAZON. Источник: Amazon полностью заменит работу человека роботом на промышленных предприятиях. URL: https://versiya.info/tehnika-i-tehnologii/101175/amp

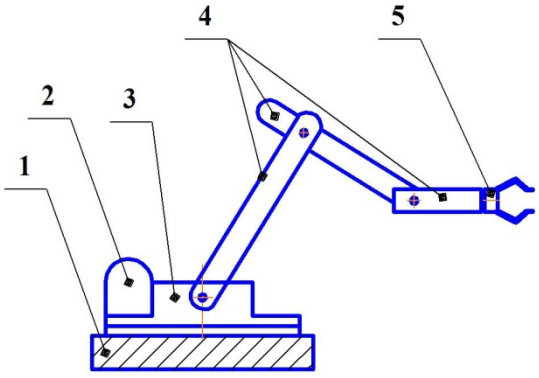

Рис. 2. Функциональная схема робота-манипулятора: 1 – неподвижная опора; 2 – электродвигатель; 3 – опорное вращательное звено; 4 – возвратно-поступательные звенья; 5 – захват
Источник: собственная разработка авторов

Количество степеней свободы, а также рабочая область промышленного робота будет зависеть от взаимного расположения и комбинации звеньев (рис. 3).

В большинстве случаев в исполнительном механизме робота-манипулятора первые три звена осуществляют транспортные функции. Они обеспечивают смещение рабочего органа в необходимое положение. Все остальные сочленения выполняют ориентирующую функцию, направляя рабочий орган согласно поставленной задаче [3].

Ниже представлены четыре категории промышленных роботов, которые делятся в зависимости от вида первых трёх сочленений:

- роботы, у которых все три начальных сочленения являются поступательными (они работают в декартовой системе координат);
- роботы, у которых среди начальных сочленений два поступательных и одно вращательное (они работают в цилиндрической системе координат);
- роботы, у которых среди начальных сочленений одно поступательных и два вращательных (они работают в сферической системе координат);
- роботы, у которых все три начальных сочленения являются вращательными (они работают в угловой, или вращательной, системе координат).

Разделение степеней подвижности у некоторых промышленных роботов на переносные и ориентирующие не предусмотрено. В качестве примера можно привести роботов с числом степеней свободы более шести (избыточная кинематика).

Рабочий орган – устройство, которое предназначено для реализации конкретной производственной задачи. Он размещается на последнем звене робота-манипулятора. В качестве рабочего органа могут выступать как универсальные устройства по типу захватов, так и профильные инструменты.

Схват – устройство, захватывающее и удерживающее объект посредством относительного перемещения частей данного устройства. Он является одним из самых универсальных видов захватного устройства. Схват по конструкции напоминает человеческую кисть: захват объектов производится при помощи механических «пальцев» [1, 3].

Электрические, пневматические или гидравлические двигатели применяют в качестве привода. Электрические приводы способствуют выполнению более точных операций. При этом гидравлические приводы используют для более тяжелых работ, где необходимо развивать высокое быстродействие или большое усилие. В свою очередь, пневматические приводы применяют на малогабаритных роботах для выполнения простых цикличных операций.

Основным элементом аппаратной части является силовой преобразователь – драйвер двигателя. Для управления электродвигателями постоянного тока используют некоторое количество схем. Самой функциональной из всех является Н-мост. Общая схема Н-моста изображена на рис. 4.

Показанная схема содержит четыре ключа. Они включены попарно последовательно. Между парами располагается двигатель (якорная цепь). Два нижних ключа подключаются к отрицательной шине источника питания, два верхних ключа – к положительной шине источника питания. Для включения двигателя необходимо, чтобы были включены два ключа, допустим S1 и S4, в данном случае ток будет протекать от источника питания через ключ S1, далее через якорь двигателя и через ключ S4, а два другие ключа должны быть закрыты. Для того чтобы реверсировать движение тока, в якоре двигателя необходимо закрыть ключи S1 и S4, а ключи S2 и S3 открыть [3].

В наши дни предъявляются все большие требования к универсальности алгоритмов, используемых для управления промышленными робототехническими системами. В настоящее время наиболее перспективным подходом является применение обучения с подкреплением [4, 5]. Рассмотрим подробнее данный подход.

В литературе [4] используются термины «агент» и «среда» – для обозначения робототехнической системы и внешних факторов соответственно. Вся концепция подхода строится на утверждении, что существует функция стоимости [5], зависящая от

предыдущих действий агента и состояния среды, которая может быть рассчитана в каждый момент времени, и ее максимизация влечет за собой выполнение поставленной перед робототехнической системой задачи.

На рис. 5 приведена схема процесса обучения с подкреплением.

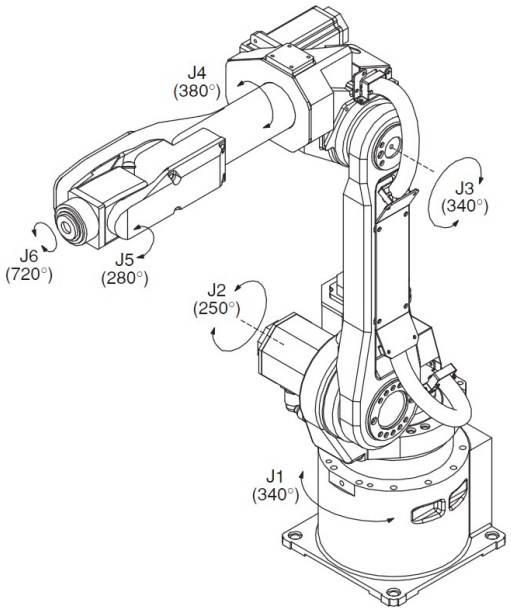

Рис. 3. Схема робота-манипулятора с обозначением степеней свободы [2, 3]

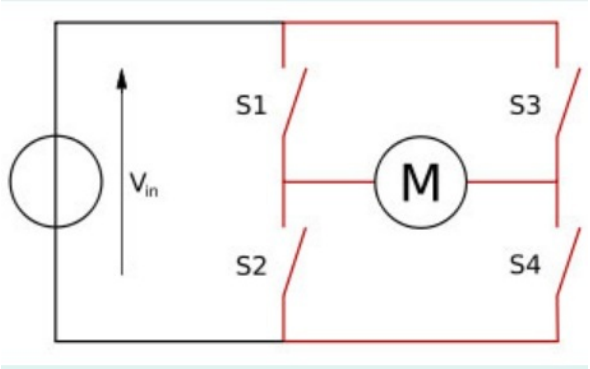

Рис. 4. Общая схема Н-моста [2, 3]

Рис. 5. Схема обучения с подкреплением
Источник: собственная разработка авторов

Результаты исследования и их обсуждение

Технически наиболее сложным является математическое описание функции стоимости, в то время как задача её оптимизации может быть решена с помощью численных методов оптимизации. Рассмотрим подробнее механизм определения функции стоимости.

Введем к уже используемым обозначениям переменную V_t, определяющую взвешенную сумму ранее полученных наград к шагу *t* [4]. Она определяется как

$$V_t = \sum_{i=0}^{t} \gamma^i R(S_t),$$

где γ – коэффициент, обеспечивающий снижение значения последних действий. Подобная формулировка необходима для увеличения количества действий, приводящих к положительному результату. Данное выражение справедливо для случая, когда действия последовательны, то есть у агента нет выбора. Для случая выбора из *K* действий на каждом шаге

$$V_t\big|_{S_1} = \sum_{k=2}^{K} P\left(s_k \middle| s_1\right)\left(R_k + \gamma V\big|_{S_k}\right),$$

где $V_t\big|_S$ – ранее полученные награды к шагу *t*, при текущем состоянии среды *S*; $P\left(s_k \middle| s_1\right)$ – вероятность перехода среды в состояние S_k при выборе действия *k*. Заключительным изменением данной формулировки является введение «политики» [6]. Под политикой (π) понимается стратегия принятия решения выбора конкретного действия в текущем состоянии среды

$$V_\pi\big|_S = \sum_a \pi(a \mid s) \sum_{s' \in S} \sum_{r \in R} p\left(s', r \mid s, a\right)\left(r + \gamma V_\pi\big|_{S'}\right),$$

где *a* – все доступные для агента действия; π – текущая политика.

Задача построения алгоритма обучения с подкреплением делится на два этапа: моделирование среды и описание и оптимизация функции стоимости. Так как промышленные робототехнические системы работают в реальном мире, модель окружения должна отражать основные физические законы. Проведя анализ доступных физических окружений – библиотек, использующих просчет физики для инженерных целей, – были выделены две библиотеки: MuJoCo [7, 8] и pyBullet [9]. Данные библиотеки фундаментально не отличаются между собой – только разными подходами к описанию логики и ценовой политикой. Было принято решение использовать в качестве физического окружения pyBullet ввиду его бесплатной модели распространения. Для реализации среды был смоделирован промышленный робот, соответствующая физическая среда и объекты, на которые робот должен воздействовать (рис. 6). Таким образом, первый этап построения модели обучения с подкреплением можно считать выполненным.

Была выбрана тривиальная задача – касание захватом манипулятора заданного объекта (на рис. 6 зеленый цилиндр). В качестве функции стоимости выступало следующее выражение:

$$R_t = \frac{\sqrt{\left(C_x - T_x\right)^2 + \left(C_y - T_y\right)^2 + \left(C_y - T_y\right)^2}}{t},$$

где *C* – координаты цели (цилиндра); *T* – координаты захвата робота; *t* – текущий шаг. Деление на *t* обусловлено условием минимизации количества действий робота: чем больше шагов пройдено, тем меньше награда на текущем шаге. Функция стоимости определяется как максимизация награды $V = \max_W(R)$, где *W* – параметры модели.

В качестве агента, взаимодействующего со средой, использовалась искусственная нейронная сеть [10]. Использовался простейший персептрон с 2 скрытыми слоями. Входами служили координаты объекта и существующие углы поворотов шарнирных сочленений робота. Выходами – угол поворота сочленений на данном шаге. Архитектура искусственной нейронной сети приведена на рис. 7.

Данная архитектура была выбрана из эмпирических соображений [11], подбор гиперпараметров не производился.

Данная сеть обучалась методом обратного распространения ошибки, модификацией Adam [10]. Система обучалась около 12 часов. При тестировании устойчивости системы (случайное положение цилиндра) успех достигается в 95 % случаев. Следует отметить, что для применения системы требуется 4 операции матричного перемно-

жения, 4 операции матричного сложения и 3 операции поэлементного применения нелинейной функции. Данное количество операций уже учитывает моделируемые физические условия. Увеличение конечной точности системы может быть достигнуто за счет подбора гиперпараметров и увеличения времени обучения.

На рис. 8 приведены графики изменения ошибки обучения в зависимости от эпохи. Под эпохой обучения понимается одно изменение весов ИНС [10].

Из рисунка видно, что после 500 эпох обучения появляется участок с флуктуациями ошибки. Это связано с тем, что ИНС обучается на случайно сгенерированных примерах, обобщая полученную информацию. После 1500 эпохи целесообразно прекратить обучение, так как достигнута квазиоптимальная конфигурация ИНС.

Выводы

По результатам проведенных теоретических исследований было выявлено, что данный метод показывает высокий процент реализации поставленной задачи (95 %). В дальнейшем планируется тестирование данных моделей на стендовых образцах.

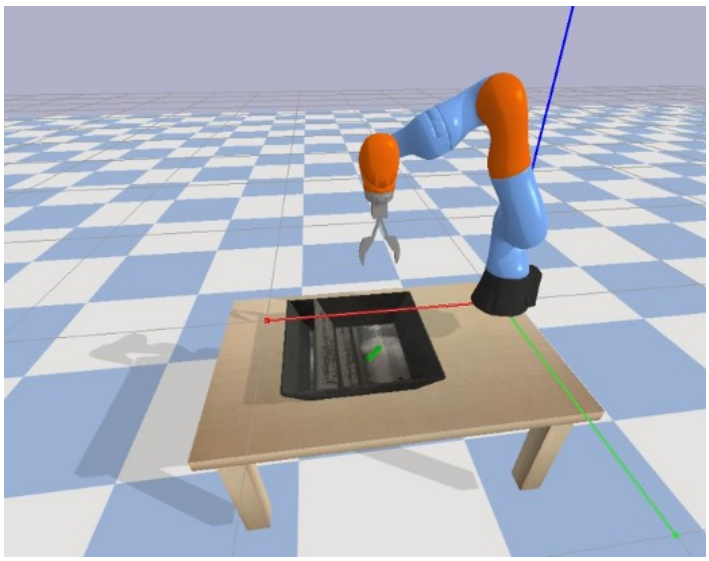

Рис. 6. Рендеринг физического окружения pyBullet для робота-манипулятора
Источник: собственная разработка авторов в рендеринге физического окружения pyBullet

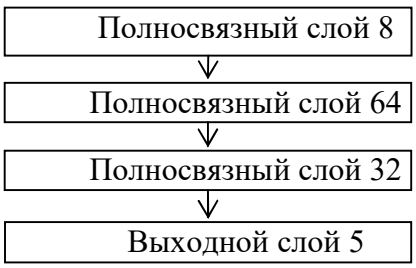

Рис. 7. Архитектура агента, выполненного в виде искусственной нейронной сети
Источник: собственная разработка авторов

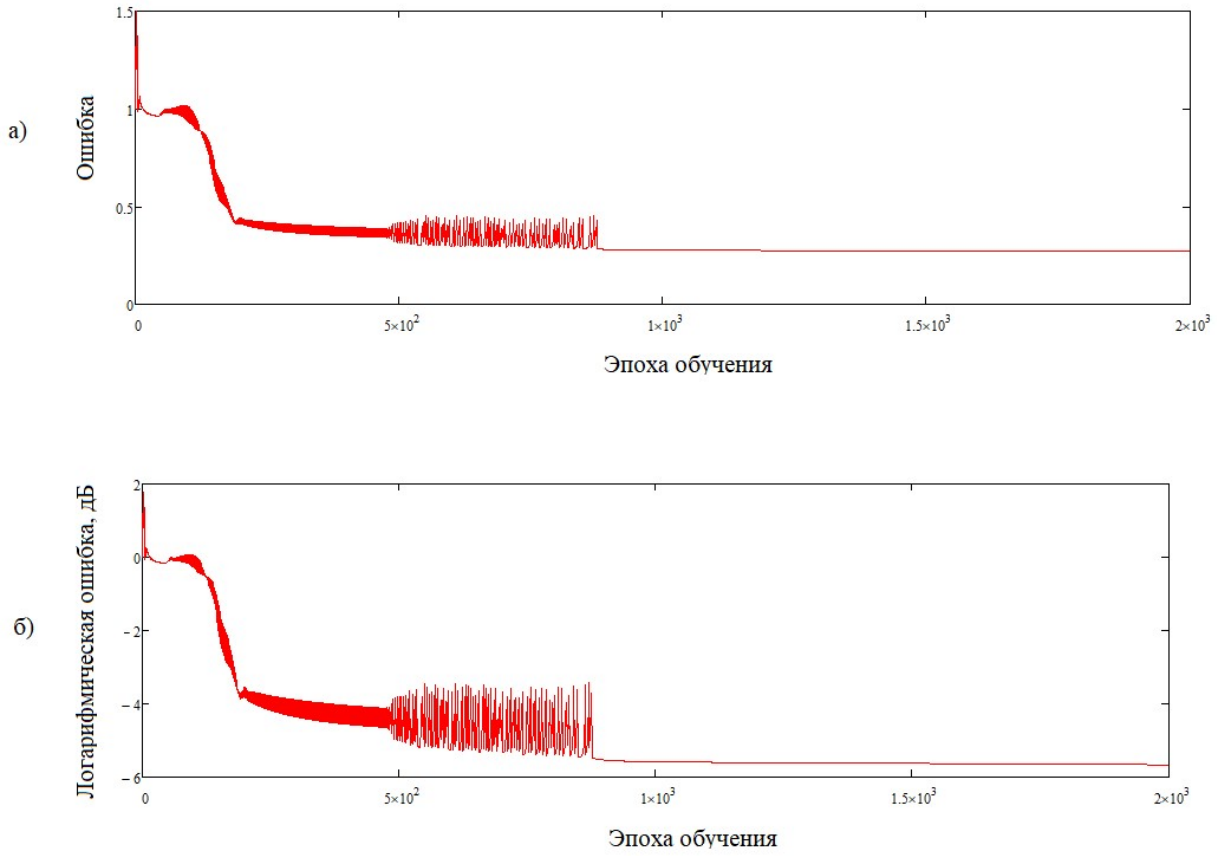

Рис. 8. Графики изменения ошибки обучения в зависимости от эпохи
Источник: собственная разработка авторов

Библиографический список

1. Лекции. Классификации систем координат. – URL: http://baumanki.net/lectures/ (дата обращения 17.10.2019).

2. Анализ современного состояния применения роботов в промышленности. – URL: https://ru.wikipedia.org/wiki (дата обращения 17.10.2019).

3. Схиртладзе, А. Г. Классификация и структура промышленных роботов / А. Г. Схиртладзе, В. И. Выходец, Н. И. Никифоров. – URL: http://www.metal-working.ru/ (дата обращения 17.10.2019).

4. Саттон, Р. С. Обучение с подкреплением / Р. С. Саттон, Э. Г. Барто. – Москва : Бином. Лаборатория знаний, 2017. – 399 с.

5. Dueling Network Architectures for Deep Reinforcement Learning / Z. Wang, T. Schaul, M. Hessel [et al.] // CoRR. – abs/1511.06581. – 2016. – P. 1–15.

6. Policy Gradient Methods for Reinforcement Learning with Function Approximation / R. S. Sutton, D. A. McAllester, S. P. Singh, M. Yishay // Advances in Neural Information Processing Systems. – 2000. – No. 12. – P. 1057–1063.

7. MuJoCo advanced physics simulation. – URL: http://www.mujoco.org/ (дата обращения: 18.10.2019).

8. Todorov, E. A physics engine for model-based control / E. Todorov, T. Erez, Y. Tassa // IROS. – 2012. – P. 5026–5033. – DOI: 10.1109/IROS.2012.6386109.

9. Bullet Real-Time Physics Simulation. – URL: https://pybullet.org/wordpress/ (дата обращения: 18.10.2019).

10. Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль. – Москва : «ДМК Пресс», 2017. – 652 с.

11. Толстых, А.А. Выбор архитектуры искусственной нейронной сети на основе сравнения эффективности методов распознавания изображений / А. А. Толстых, А. Н. Голубинский // Вестник Воронежского института МВД России. – 2018. – № 1. – С. 27–37.

References

1. *Lektsii. Klassifikatsii sistem koordinat* [Lectures. Classifications of coordinate systems]. Available at: http://baumanki.net/lectures/ (Accessed 17 November 2019). (In Russian).

2. *Analiz sovremennogo sostojanija primenenija robotov v promyshlennosti* [Analysis of the current state of application of robots in industry]. Available at: https://ru.wikipedia.org/wiki (Accessed 17 November 2019). (In Russian).

3. Shirtladze A.G., Vyhodec V.I., Nikiforov N.I. *Klassifikacija i struktura promyshlennyh robotov* [Classification and structure of industrial robots]. Available at: http://www.metal-working.ru/ (Accessed 17 November 2019). (In Russian).

4. Satton R.S., Barto Je.G. *Obuchenie s podkrepleniem* [Training with reinforcement]. Moscow, 2017, 399 p. (In Russian).

5. Wang Z., Schaul T., Hessel M. et al. (2016) Dueling Network Architectures for Deep Reinforcement Learning. *CoRR*, abs/1511.06581, pp. 1-15.

6. Sutton R.S., McAllester D.A., Singh S.P., Yishay M. (2000) Policy Gradient Methods for Reinforcement Learning with Function Approximation. *Advances in Neural Information Processing Systems*,12, pp. 1057-1063.

7. MuJoCo advanced physics simulation. Available at: http://www.mujoco.org/ (Accessed 18 November 2019).

8. Todorov E., Erez T., Tassa Y. (2012) A physics engine for model-based control. *IROS*, pp. 5026-5033. DOI: 10.1109/IROS.2012.6386109.

9. Bullet Real-Time Physics Simulation. Available at: https://pybullet.org/wordpress/ (Accessed 18 November 2019).

10. Gudfellou Ya., Bendzhio I., Kurvill A. *Glubokoe obuchenie* [Deep Learning]. Moscow, 2017, 652 p. (In Russian).

11. Tolstyh A.A., Golubinsky A.N. (2018) *Vybor arhitektury iskusstvennoj nejronnoj seti na osnove sravnenija jeffektiv-nosti metodov raspoznavanija izobrazhenij* [The choice of artificial neural network architecture based on a comparison of the efficiency of image recognition methods]. *Vestnik Voronezhskogo instituta MVD Rossii* [Bulletin of the Voronezh Institute of the Ministry of Internal Affairs of Russia], no. 1, pp. 27-37 (In Russian).

Сведения об авторах

Толстых Андрей Андреевич – преподаватель кафедры тактико-специальной подготовки Воронежского института МВД России, г. Воронеж, Российская Федерация; e-mail: tolstykh.aa@yandex.ru

Ступников Дмитрий Сергеевич – кандидат технических наук, преподаватель кафедры механизации лесного хозяйства и проектирования машин ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова», г. Воронеж, Российская Федерация; e-mail: Neiti1992@mail.ru.

Малюков Сергей Владимирович – кандидат технических наук, доцент кафедры механизации лесного хозяйства и проектирования машин ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова», г. Воронеж, Российская Федерация; e-mail: malyukovsergey@yandex.ru.

Лукьянов Александр Сергеевич – кандидат технических наук, старший преподаватель кафедры инфокоммуникационных систем и технологий Воронежского института МВД России, г. Воронеж, Российская Федерация; e-mail: las92@yandex.ru.

Лунёв Юрий Станиславович – кандидат технических наук, старший преподаватель кафедры автоматизированных информационных систем Воронежского института МВД России, г. Воронеж, Российская Федерация; e-mail: xalt@mail.ru.

Information about authors

Tolstykh Andrey Andreevich – Lecturer at the Department of Tactical and Special Training, Federal State Public Educational Establishment of Higher Training "Voronezh Institute of the Ministry of the Interior of the Russian Federation", Voronezh, Russian Federation; e-mail: tolstykh.aa@yandex.ru.

Stupnikov Dmitry Sergeevich – PhD (Engineering), Lecturer of the Department of Forestry Mechanization and Machine Design, FSBEI HE "Voronezh State University of Forestry and Technologies named after G.F. Morozov", Voronezh, Russian Federation; e-mail: Neiti1992@mail.ru.

Malyukov Sergey Vladimirovich – PhD (Engineering), Associate Professorof the Department of Forestry Mechanization and Machine Design, FSBEI HE "Voronezh State University of Forestry and Technologies named after G.F. Morozov", Voronezh, Russian Federation; e-mail: malyukovsergey@yandex.ru.

Lukyanov Aleksandr Sergeevich – PhD (Engineering), Senior Lecturer, Department of Infocommunication Systems and Technologies, Federal State Public Educational Establishment of Higher Training "Voronezh Institute of the Ministry of the Interior of the Russian Federation", Voronezh, Russian Federation; e-mail: las92@yandex.ru.

Lunev Yury Stanislavovich – PhD (Engineering), Senior Lecturer, Department of Automated Information Systems, Federal State Public Educational Establishment of Higher Training "Voronezh Institute of the Ministry of the Interior of the Russian Federation", Voronezh, Russian Federation; e-mail: xalt@mail.ru.