

17. Chowdhery A, et al. Palm: Scaling language modeling with pathways, *Journal of Machine Learning Research*, 2023, Vol. 24, No. 240, pp. 1-113.
18. Lapkin YA.D., Matveev V.D. Sravnitel'nyy analiz bol'shikh yazykovykh modeley dlya avtonomnykh robototekhnicheskikh sistem [Comparative analysis of large language models for autonomous robotic systems], *Ekstremal'naya robototekhnika* [Experimental Robotics], 2025, No. 1 (36), pp. 110-115. EDN FXTGG.
19. Gemma Team: Aishwarya Kamath, et al. Google DeepMind. Gemma 3 Technical Report, *arXiv:2503.19786*, 2025.
20. Tianyi Zhang, et al. BERTScore: Evaluating Text Generation with BERT, *arXiv:1904.09675*, 2019.
21. Rviz User's Guide, *RViz 1.9.0 documentation*. Available at: https://docs.ros.org/en/jade/api/rviz/html/user_guide/ (accessed 10 February 2026).

Кобзарь Дарья Сергеевна – Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики; e-mail: d.kobzar@rtc.ru; г. Санкт-Петербург, Россия; тел.: +78125520110; математик.

Матвеев Виктор Дмитриевич – Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики; e-mail: v.matveev@rtc.ru; г. Санкт-Петербург, Россия; тел.: +78125520110; м.н.с.

Лапкин Ярослав Дмитриевич – Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики; e-mail: ya.lapkin@rtc.ru; г. Санкт-Петербург, Россия; тел.: +78125520110; техник.

Богданов Ростислав Романович – Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики; e-mail: r.bogdanov@rtc.ru; г. Санкт-Петербург, Россия; тел.: +78125520110; техник.

Изыумов Антон Сергеевич – Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики; e-mail: a.izyumov@rtc.ru; г. Санкт-Петербург, Россия; тел.: +78125520110; техник.

Kobzar Darya Sergeevna – Russian State Scientific Center for Robotics and Technical Cybernetics; e-mail: d.kobzar@rtc.ru; St Petersburg, Russia; phone: +78125520110; mathematician.

Matveev Viktor Dmitrievich – Russian State Scientific Center for Robotics and Technical Cybernetics; e-mail: v.matveev@rtc.ru; St Petersburg, Russia; phone: +78125520110; junior research fellow.

Lapkin Yaroslav Dmitrievich – Russian State Scientific Center for Robotics and Technical Cybernetics; e-mail: ya.lapkin@rtc.ru; St Petersburg, Russia; phone: +78125520110; technician.

Bogdanov Rostislav Romanovich – Russian State Scientific Center for Robotics and Technical Cybernetics; e-mail: r.bogdanov@rtc.ru; St Petersburg, Russia; phone: +78125520110; technician.

Izyumov Anton Sergeevich – Russian State Scientific Center for Robotics and Technical Cybernetics; e-mail: a.izyumov@rtc.ru; St Petersburg, Russia; phone: +78125520110; technician.

УДК 004.896

DOI 10.18522/2311-3103-2026-2-185-201

М.Ю. Медведев, В.Х. Пшихопов, А.Р. Гайдук, И.М. Медведев, Е.Ю. Косенко

МЕТОДЫ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В ЗАДАЧАХ АДАПТИВНОГО УПРАВЛЕНИЯ НЕЛИНЕЙНЫМИ ДИНАМИЧЕСКИМИ ОБЪЕКТАМИ

Актуальность проблемы адаптивного управления нелинейными динамическими объектами обеспечивается постоянно возрастающими требованиями к качеству и условиям функционирования технических систем. Автоматизация и роботизация приводят к возрастанию сложности решаемых задач и необходимости адаптироваться к структурной и параметрической неопределенности, внешним возмущениям. В последние несколько лет приобретает популярность применение методов обучения с подкреплением для синтеза адаптивных систем управления. Указанные методы демонстрируют эффективность при управлении неопределенными динамическими объектами. Однако, существует две базовые проблемы применения методов машинного обучения в

системах управления динамическими объектами. Во-первых, необходимость обеспечения асимптотической устойчивости желаемой траектории замкнутой системы ограничивает применение методов глубокого обучения. Во-вторых, в процессе обучения необходимо, чтобы интеллектуальный регулятор не вырабатывал управления, приводящие к выходу переменных состояния за заданные ограничения. Целью данной статьи является обзор и анализ последних достижений в области применения методов обучения с подкреплением для синтеза адаптивных систем управления нелинейными динамическими объектами. Отдельное внимание уделяется методам Actor-Critic, которые по своей структуре близки к адаптивным системам управления с самонастраивающимися параметрами. На основе проведенного анализа предлагается структура адаптивной системы с двухкомпонентным управлением, включающим номинальный и адаптивный регуляторы. Предлагается алгоритм адаптивного управления, отличающийся использованием модифицированного алгоритма Actor-Critic, новой формой дельта-ошибки и отсутствием особенности в окрестности нуля. Предлагаемый алгоритм позволяет снизить число настраиваемых параметров. В статье приводятся условия устойчивости по Ляпунову нулевого положения равновесия замкнутой системы и пример синтеза и моделирования предлагаемого алгоритма адаптивного управления.

Обучение с подкреплением; адаптивное управление; нелинейный объект; алгоритм Actor-Critic; устойчивость.

M.Yu. Medvedev, V.Kh. Pshikhov, A.R. Gaiduk, I.M. Medvedev, E.Yu. Kosenko

REINFORCEMENT LEARNING METHODS IN ADAPTIVE CONTROL OF NONLINEAR DYNAMIC OBJECTS

The relevance of the problem of adaptive control of nonlinear dynamic objects is ensured by the ever-increasing demands on the quality and operating conditions of technical systems. Automation and robotics lead to an increase in the complexity of the problems being solved and the need to adapt to structural and parametric uncertainty and external disturbances. In recent years, the use of reinforcement learning methods for the synthesis of adaptive control systems has gained popularity. These methods demonstrate effectiveness in controlling uncertain dynamic objects. However, there are two fundamental problems with the application of machine learning methods to control systems for dynamic objects. First, the need to ensure asymptotic stability of the desired trajectory of a closed-loop system limits the application of deep learning methods. Second, during the learning process, it is necessary that the intelligent controller does not generate controls that lead to state variables exceeding specified limits. The purpose of this article is to review and analyze recent advances in the application of reinforcement learning methods to the synthesis of adaptive control systems for nonlinear dynamic objects. Particular attention is given to Actor-Critic methods, which are structurally similar to adaptive control systems with self-adjusting parameters. Based on the analysis, the structure of an adaptive system with two-component control, including nominal and adaptive controllers, is proposed. An adaptive control algorithm is proposed, distinguished by the use of a modified Actor-Critic algorithm, distinguished by a new form of the delta error and the absence of a singularity in the neighborhood of zero. The proposed algorithm allows for a reduction in the number of adjustable parameters. The article presents conditions for Lyapunov stability of the zero-equilibrium position of a closed-loop system and an example of the synthesis and modeling of the proposed adaptive control algorithm.

Reinforcement learning; adaptive control; nonlinear object; Actor-Critic algorithm; robustness.

Введение. Проблема адаптивного управления всегда являлась одной из центральных в теории автоматического управления. Одним из общих подходов к решению данной проблемы является метод адаптивного управления с самонастраивающимися параметрами [1, 2]. В рамках данного метода показано, что для линейных объектов обеспечивается асимптотическая устойчивость нулевого положения замкнутой системы. Для нелинейных систем линейная структура регулятора, в общем случае, может обеспечить асимптотическую устойчивость заданного положения в локальной области. В нелинейном случае наиболее часто обеспечивается устойчивость по Ляпунову, которая гарантирует нахождение ошибки в некоторой ограниченной области.

Для нелинейных систем основная проблема состоит в выборе структуры регулятора в условиях неопределенности модели объекта. Появление искусственных нейронных сетей, которые эффективно решают задачи многомерной аппроксимации, дало очередной импульс развитию адаптивных систем с настраиваемыми параметрами. Одним из рас-

пространственных способов применения нейронных сетей – параметрическая настройка параметров базового регулятора или аппроксимация модели объекта [3–5]. В таких системах алгоритм управления и адаптации / идентификации синтезируются отдельными контурами, что упрощает настройку и анализ.

В последние десятилетия перспективным направлением для синтеза адаптивного управления становится обучение с подкреплением [6]. В рамках данного подхода настройка параметров осуществляется в результате оценки эффективности того или иного действия с помощью функции вознаграждения. Совместное применение методов обучения с подкреплением и глубоких нейронных сетей показывает высокую эффективность в задачах управления нелинейными динамическими объектами. Однако существует несколько проблем, ограничивающих применением такого подхода. Первая проблема заключается в необходимости обеспечения асимптотической устойчивости желаемой траектории движения замкнутой системы. Данная проблема требует строго математического обоснования, что ограничивает способы параметризации алгоритма управления. Вторая проблема заключается в том, что в ходе настройки регулятор должен обеспечить нахождение замкнутой системы в заданной, безопасной для функционирования, области. Третья проблема заключается в большом количестве настраиваемых параметров, включающих выбор числа слоев и числа нейронов в нейронной сети, шага и гиперпараметров обучения, ограничений на приращение градиента настраиваемых параметров и др.

В данной работе исследуется широко распространенный метод Actor-Critic [6] в задаче управления нелинейными динамическими объектами. На основе проведенного анализа предлагается алгоритм адаптивного управления, включающий номинальный и адаптивный регулятор. В качестве номинального регулятора используется управление, синтезированное на основе квазилинейной математической модели [7, 8]. Данный метод отличается тем, что он не накладывает ограничений на структуру управляемого объекта и может применяться как для аффинных, так и для неаффинных по управлению объектов. В качестве адаптивного управления предложен модифицированный алгоритм Actor-Critic, в котором устранена неопределенность в окрестности нуля, введен параметр, позволяющий автоматически стабилизировать процесс настройки, предложена и исследована структура δ -ошибки.

Обзор работ в области адаптивного управления на базе обучения с подкреплением. Проблема адаптивного управления нелинейными неопределенными объектами с использованием методов обучения с подкреплением активно исследуется в настоящее время. Обзор работ показывает, что метод ценности состояния применяется намного реже, что связано с необходимостью реализации длительной процедуры предварительного обучения [9, 10]. Наиболее часто используемым подходом является применение методов градиента стратегии [11–15].

Анализ работ показывает, что наиболее актуальной темой является проблема адаптивного управления нелинейными объектами [11–20, 22, 24, 26, 27]. Для нелинейной постановки характерно использование моделей блочного вида [11, 12, 17, 18] и аффинных по управлению моделей [12–18, 23, 24].

Так в статье [11] для дискретной модели, при выполнении условий дифференцируемости и управляемости, предложен метод адаптивного управления, базирующийся на структуре Actor-Critic. Решается задача оптимизации критерия качества в виде взвешенной суммы ошибки слежения за уставкой. Для оценки значения критерия качества используется нейронная сеть Critic, а для аппроксимации оптимального управления – нейронная сеть Actor. Обе нейронные сети являются однослойными и используют радиально-базисные функции для аппроксимации. В работе доказано свойство абсолютной ограниченной устойчивости (Uniformly Ultimate Boundedness Stability – UUB). Данное свойство гарантирует, что через заданное время переменные состояния системы войдут в некоторую ограниченную область и будут оставаться там в дальнейшем.

В работе [12] рассматривается нелинейный объект второго порядка блочной структуры, аффинный по управлению. Вектор переменных состояния считается измеряемым и ограничен заданной областью. Нелинейности непрерывны и дифференцируемы. Также

присутствует ограниченное внешнее возмущение, приведенное ко входу объекта. Для синтеза управления в статье используется архитектура Actor-Critic. Для аппроксимации неизвестной нелинейности, общей функции стоимости и производной задающего сигнала используются нейронные сети, выходы которых определяют адаптивное управление. Показано свойство UUB и приведены два численных примера.

В работе [17] рассматриваются непрерывные уравнения движения надводного судна. В уравнениях динамики учтена неопределенность демпфирующей матрицы и введена немоделируемая динамика. Предложен адаптивный закон управления, который использует нейронные сети для идентификации неопределенной динамики и оценки оптимального значения критерия качества. Вводятся барьерные функции, которые ограничивают область, в которой находятся переменные состояния объекта. Доказывается свойство UUB в замкнутой системе управления. Приводятся результаты моделирования, подтверждающие эффективность предложенного метода управления.

В работе [18] рассматривается задача управления колесным роботом, который описывается уравнениями твердого тела. В модель объекта вводится неопределенность, обусловленная действием сил трения и матрицей инерции. В результате получена нелинейная аффинная по управлению модель с возмущениями. Для решения задачи оптимального по интегральному критерию качества управления используется структура Actor-Critic. Как и в других статьях, алгоритм Critic использует однослойную нейронную сеть для оценки оптимального значения критерия качества. Адаптивное управление также аппроксимируется однослойной нейронной сетью (Actor). Предлагаются градиентные алгоритмы настройки нейронных сетей, использующие метод обучения с подкреплением. В результате анализа устойчивости показано свойство UUB. Приводятся результаты численного моделирования.

В работах [12–18, 23, 24], кроме блочной структуры, рассматривается случай, когда управление входит в уравнения объекта линейно.

В статье [13] рассмотрен объект с приведенным ко входу ограниченным возмущением. Нелинейность является липшицевой, а норма вектора входа ограничена. Предлагается управление, использующее структуру Actor-Critic. Для аппроксимации неопределенных нелинейностей объекта и алгоритмов Actor и Critic используются однослойные нейронные сети, параметры которых настраиваются градиентным методом. Показано свойство UUB и приведен пример моделирования системы мостового крана с грузом, подвешенным на тележке, которая движется по рельсу. Особенностью статьи является синтез управления, которое изменяется по событиям.

В работе [14] делаются предположения, что нелинейности удовлетворяют условию Липшица и ограничены по норме. Для синтеза управления используется оценка значения критерия качества однослойной нейронной сетью. Алгоритм настройки параметров данной нейронной сети базируется на процедуре адаптивного динамического программирования. При этом синтезированное управление также изменяется по событиям. В статье показано свойство UUB и приводятся примеры синтеза регуляторов для линейных объектов четвертого порядка.

В статье [15] рассматривается сеть взаимосвязанных нелинейных объектов, описываемых непрерывной моделью. Предполагаются единственность положения равновесия, липшицевость системной нелинейности и матрицы взаимосвязей объекта, а также ограниченность нормы вектора входа. Для аппроксимации оптимального значения критерия качества используется однослойная нейронная сеть, параметры которой настраиваются из условий UUB. Приводится пример синтеза управления для системы из двух объектов второго порядка. Кроме того, что рассматривается управляемое событиями управление, условие изменения управления является динамическим.

В статье [16] рассмотрена нелинейная система с неопределенным вектором состояния и ограниченным по норме возмущением. Неопределенный вектор удовлетворяет условиям Липшица. Адаптивное управление строится с использованием структуры Actor-Critic, причем каждый из указанных алгоритмов реализован однослойной нейронной сетью с радиально-базисными функциями. В данной статье также доказано

свойство UUB. Эффективность предлагаемого метода демонстрируется на примере системы из двух перевернутых маятников, соединенных пружиной. Приводятся результаты численных расчетов.

Авторы работы [23] рассматриваются марковский процесс принятия решений, описываемый нелинейной дискретной системой. Исследуются два алгоритма обучения с подкреплением, оптимизирующие сумму обесцененных вознаграждений за все время функционирования системы. Первый алгоритм – офф-лайн Actor-Critic алгоритм, а второй – онлайн алгоритм оптимизации доверительного региона. Нелинейная неопределенная динамика объекта моделируется Гауссовским процессом, который находится в заданной области. Для обеспечения безопасности процесса обучения вводятся барьерные функции. С использованием данных функций предлагается управление, которое настраивается на основе метода обучения с подкреплением. Доказаны теоремы о том, что безопасная область является инвариантом для замкнутой системы.

В статье [24] также используются барьерные функции для обеспечения безопасности процесса обучения. Вводится номинальная модель динамики и синтезируется базовый регулятор с использованием барьерных функций Ляпунова. Далее предлагается реализация адаптивного управления с использованием барьерных функций и базового регулятора. Эффективность предложенных алгоритмов демонстрируется на примере обучения двуногого робота.

В работах [11, 19, 20, 22, 26, 27] рассматриваются неаффинные по управлению объекты.

В статье [19] рассмотрена дискретная модель, для которой синтезируется адаптивное управление, использующее структуру Actor-Critic. При этом используются двухслойные нейронные сети с линейной функцией активации во входном слое и нелинейной функцией активации в выходном слое. В статье показана асимптотическая устойчивость замкнутой системы, однако это свойство показано при условии, что все переменные состояния ограничены. В частности, приведен пример неаффинной системы второго порядка, в которой все переменные ограничены функциями тангенс гиперболический.

Авторы работы [20] рассматривают нелинейную дискретную модель общего вида. Предполагается, что правая часть может быть представлена в виде суммы известного и неизвестного векторов. Также предполагается выполнение условий Липшица для правых частей уравнений объекта. Дополнительно делается достаточно сильное предположение о том, что норма разности между правой частью объекта и некоторым средним её значением с заданной вероятностью меньше известного положительного числа. Для этих условий находится функция Ляпунова и ограниченная область устойчивости. Таким образом гарантируется безопасность процесса обучения.

В работе [22] ставится задача найти адаптивное управление, оптимизирующее заданный интегральный критерий качества. Предлагается базовый алгоритм, обеспечивающий асимптотическую устойчивость положения равновесия системы. Для неопределенной системы вводятся две нейронные сети. Рекуррентная нейронная сеть предложена для идентификации объекта, а для оценки критерия качества – однослойная нейронная сеть. На основе метода обучения с подкреплением предлагаются алгоритмы настройки коэффициентов нейронных сетей. В статье показано свойство UUB и ограниченность ошибок оценивания.

В статье [26] рассматривается нелинейный объект, представленный непрерывной моделью с чистыми обратными связями. Нелинейные функции правой части или уравнения выхода? считаются неопределенными. Кроме того, введены аддитивные ограниченные по амплитуде возмущения. Вводится адаптивное управление, использующее однослойные нейронные сети с радиально-базисными функциями активации. Проведен анализ устойчивости, который показал ограниченность сигналов в замкнутой системе. Приводятся результаты моделирования на примере звена робота, управляемого двигателем постоянного тока.

Авторы статьи [27] ставят задачу оптимизировать интегральный критерий качества в условиях, когда вектор правых частей не известен, но является гладким. Для синтеза адаптивного управления, аналогично работе [22], используется рекуррентная нейронная

сеть для идентификации объекта управления, а для оценки оптимального значения критерия качества – однослойная нейронная сеть Critic. Предложены градиентные алгоритмы настройки весовых коэффициентов нейронных сетей и показана ограниченность ошибок замкнутой системы.

В статьях [21, 28] модель объекта представляется в виде марковского процесса.

В статье [21] рассматривается случайный марковский процесс принятия решений роботом. Робот характеризуется множеством состояний и множеством возможных действий. Переход из одного состояния в другое при заданном действии описывается некоторой вероятностью. Целью является поиск алгоритма, обеспечивающего переход в нулевое состояние или слежение за заданным сигналом. Делается предположение об эргодичности и стационарности случайного марковского процесса и для алгоритма обучения Actor-Critic находятся условия сходимости в среднем, которые эквивалентны условию устойчивости по Ляпунову.

В работе [28] также рассматривается марковский процесс принятия решений, который характеризуется вероятностью перехода из одного состояния в другой при заданном действии. Формулируется проблема нахождения алгоритма, оптимизирующего сумму обесцененных вознаграждений. Предлагается структура системы управления, в которой используется управление с прогнозирующей моделью и блоком обеспечения выполнения ограничений, использующий обучение с подкреплением. Предлагаемые алгоритмы демонстрируются на примере управления системой второго порядка.

Также рассматриваются постановки задач, в которых модель является линейной [25, 29].

В работе [25] рассматривается дискретная линейная стационарная система с управлением по состоянию. Ставится задача определить вектор параметров управления, минимизирующий квадратичный функционал качества. Для решения поставленной задачи в условиях неопределенности предлагается алгоритм, базирующийся на Q-методе обучения с подкреплением. В статье доказывается сходимость алгоритма настройки и устойчивость замкнутой системы.

В статье [29] также рассматривается неопределенная линейная стационарная система. Ставится задача синтезировать управление, минимизирующее сумму вознаграждений за весь период функционирования системы. Для решения задачи, как и в [25], используется Q-метод обучения с подкреплением. На примере решения задачи квадратичной оптимизации показано схождение к оптимальному решению и асимптотическая устойчивость замкнутой системы.

Отметим, что в большинстве работ ставится задача управления состоянием объектов [13–16, 19, 20, 22, 23, 28–30]. Реже рассматривается задача управления выходом [7, 11, 12, 17, 25, 26].

В работах [30, 33] исследуются процессы сходимости процесса обучения глубоких нейронных сетей.

В работе [30] исследуется глубокая нейронная сеть, используемая для управления нелинейной аффинной по управлению дискретной системой. Ставится задача минимизации суммарного вознаграждения в течение всего времени функционирования системы. Для этого предлагается метод глобального эвристического программирования. Для аппроксимации указанного алгоритма используется многослойная нейронная сеть прямого распространения. Проводится анализ сходимости, в результате которого показано свойство UUB.

В работе [33] исследуется проблема адаптивного управления с использованием искусственных нейронных сетей. Предлагается метод синтеза нейронного регулятора, который гарантирует устойчивость желаемого положения замкнутой системы управления для объектов с существенными нелинейностями, такими как насыщение, упор и другие.

В работе [31] рассматривается проблема управления объектом с транспортным запаздыванием. Методами численного моделирования показана эффективность метода градиента политики (алгоритма управления) для компенсации транспортного запаздывания в системе первого порядка.

В статье [32] авторы рассматривают задачу перехвата подвижного объекта, который движется по прямолинейной или круговой траектории. Для ее решения используется метод Actor-Critic. При этом для решения проблемы сходимости при одновременном исследовании и управлении используется копирование сетей Actor и Critic с добавлением шума в управление, формируемое алгоритмом Actor.

Результаты проведенного обзора сведены в табл. 1, которая демонстрирует ключевые направления и методы исследований.

Таблица 1

Результаты анализа методов управления

Постановка задачи, метод	Статья
Блочная форма	[11], [12], [17], [18], [26]
Аффинные по управлению объекты	[12], [13], [14], [15], [16], [17], [18], [23], [24], [25], [29]
Объекты общего вида	[19], [20], [21], [22], [27]
Структура Actor-Critic	[11], [12], [13], [16], [18], [19], [21], [23]
Структура Critic + НС идентификатор	[17], [22], [27]
Только Critic	[14], [15]
Однослойные НС	[11], [12], [13], [14], [15], [16], [17], [18], [22], [27], [28]
Многослойные НС	[19], [30], [33]
Доказательство ограниченности ошибки, свойство UUB	[11], [12], [13], [14], [15], [16], [17], [18], [23], [24], [25], [26], [27], [28], [30]
Барьерные функции	[23], [24]
Обновляемое по событиям управление	[13], [14], [15], [22], [27]

Из табл. 1 следует, что наиболее часто рассматриваются модели с ограничениями по структуре (блочные, аффинные по управлению). Такие модели могут ограничивать класс рассматриваемых объектов. Например, неаффинность по управлению характерна для многих летательных и подводных аппаратов [34, 35].

Также заметно активное использование структуры метода обучения с подкреплением Actor-Critic, что объясняется ее близостью к структуре адаптивных систем управления с самонастраивающимися параметрами.

Отметим, что в большинстве случаев используются однослойные нейронные сети, которые позволяют провести анализ устойчивости.

Также отметим, что в нелинейных системах применение методов обучения с подкреплением обеспечивает ограниченность всех сигналов. Свойство UUB гарантирует нахождение сигналов в заданной области, но только по завершению процесса обучения. Свойство асимптотической устойчивости обеспечивается при дополнительных предположениях и ограничениях. В этом методы обучения с подкреплением сходны по результатам с известными алгоритмами адаптивного управления.

Постановка задачи, предварительные результаты и методика исследования.

Рассматривается нелинейный неаффинный объект, описываемый уравнениями вида [7]

$$\dot{x} = \zeta(x, u, v) + \Delta(x, u, v), \quad y = \xi(x, u), \quad (1)$$

где $x = [x_1 \ x_2 \ \dots \ x_n]$ – вектор переменных состояния размерности n ; u – управление; v – возмущение; y – управляемый выход объекта; $\zeta(x, u, v)$ – дифференцируемая по всем переменным, управлению и возмущению нелинейная вектор-функция размерности n ; $\xi(x, u)$ – дифференцируемая по всем переменным и управлению скалярная функция; $\Delta(x, u, v)$ – неизвестная нелинейная вектор-функция размерности n , отражающая параметрическую неопределенность и немоделируемую динамику системы.

Предполагается, что вектор x и задающее воздействие $r(t)$ являются измеряемыми. Также предполагается, что $\zeta(0, 0, 0) = 0, \xi(0, 0) = 0$. Отметим, что модель (1) является неаффинной как по управлению, так и по возмущению.

Ставится задача определить закон управления вида

$$u = u_1(r, x) + u_2(e_t), \quad (2)$$

где $u_1(r, x)$ – номинальное управление, синтезируемое по модели (1) при $\Delta(x, u, v) = 0$; $u_2(e)$ – адаптивное управление, компенсирующее ошибку $e_t = r - x$, обусловленную действием возмущения $\Delta(x, u, v)$.

Управление $u_1(r, x)$ должно обеспечивать асимптотическую устойчивость нулевого положения $x = 0$ объекта (1) при $\Delta(x, u, v) = 0$. Дополнительно при $r = r_0 1(t)$ ошибка должна стремиться к нулю $e \rightarrow 0$ при $t \rightarrow \infty$. При этом время переходных процессов в номинальной системе не должно превышать заданного значения.

В данной работе управление $u_1(r, x)$ синтезируется методом А.Р. Гайдука на основе квазилинейных моделей [7, 8, 36]. Согласно этому методу, модель (1) при $\Delta(x, u, v) = 0$ представляется квазилинейной моделью вида

$$\dot{x} = A(x)x + b(x, u)u + h(x, u, v)v, \quad y = c^T(x)x + d(x, u)u, \quad (3)$$

где

$$A(x) = \begin{bmatrix} a_{11}(x) & a_{12}(x) & \dots & a_{1n}(x) \\ a_{21}(x) & a_{22}(x) & \dots & a_{2n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(x) & a_{n1}(x) & \dots & a_{nn}(x) \end{bmatrix}, \quad b(x, u) = \begin{bmatrix} b_1(x, u) \\ b_2(x, u) \\ \vdots \\ b_n(x, u) \end{bmatrix}, \quad (4)$$

$$h(x, u, v) = \begin{bmatrix} h_1(x, u, v) \\ h_2(x, u, v) \\ \vdots \\ h_n(x, u, v) \end{bmatrix}, \quad c^T(x) = [c_1(x) \ c_2(x) \ \dots \ c_n(x)]. \quad (5)$$

В работе [7] приведены выражения для вычисления коэффициентов матрицы и векторов в выражениях (4), (5). Отметим, что квазилинейная модель (3) точно описывает исходную модель (1) и не накладывает ограничений на структуру объекта.

В работе [7] подробно описана процедура синтеза и условиях разрешимости задачи синтеза управления. К таким условиям относятся выполнение критерия управляемости по состоянию и выполнение критерия управляемости по выходу.

В настоящей статье исследуется проблема синтеза адаптивного управления $u_2(e)$. Согласно проведенному анализу распространенным подходом к синтезу управления является использование метода обучения с подкреплением на основе архитектуры Actor-Critic. В работе [6] представлен базовый алгоритм Actor-Critic. В статье [37] для линейной системы блочного вида на основе базового алгоритма Actor-Critic предложена структура адаптивного регулятора в виде:

$$u_2(e_t) = u_{max} \tanh(k_t e_t), \quad (6)$$

где u_{max} – заданное положительное число, k_t – коэффициент, настраиваемый согласно Q-методу обучения с подкреплением.

В данной работе в силу нелинейности исходного объекта нет оснований принимать линейную структуру регулятора, поэтому предлагается использовать аппроксимацию управления радиально-базисными функциями. В этом случае адаптивное управление принимает вид

$$u_2(e_t) = u_{max} \tanh(w_t^T \Phi_t), \quad (7)$$

$$\Phi_t = [\phi_1, \phi_2, \dots, \phi_{Na}]^T, \quad (8)$$

$$\phi_i = \exp(-\beta_i(e_t - c_i)^2), \quad i = \overline{1, Na}, \quad (9)$$

$$e_t = r(t) - y, \quad (10)$$

где Φ_t – вектор радиально-базисных функций (функций Гаусса); Na – число радиально-базисных функций (РБФ); β_i, c_i – положительные параметры.

Базовый алгоритм настройки вектора параметров w_t согласно [6] имеет вид

$$w_{t+1} = \begin{cases} w_t - \text{satur}_{\varepsilon_{lim}} \left[\frac{\alpha_w \delta_t^2 \nabla u_2}{u_2} \right], & |u_2| > \varepsilon_{grad}, \\ w_t, & |u_2| \leq \varepsilon_{grad}, \end{cases} \quad (11)$$

$$\nabla u_2 = u_{\max} \Phi_{it} / \cosh^2(w_t^T \Phi_t). \quad (12)$$

где $\text{satur}_{\varepsilon_{lim}}$ – функция ограничения; δ_t – дельта-ошибка; α_w – шаг обучения; ε_{grad} – малый положительный параметр, вводимый для обхода сингулярности в точке $u_2 = 0$; ε_{lim} – положительный параметр, ограничивающий величину изменения настраиваемых параметров на одном шаге.

В данной работе предлагается модификация алгоритма настройки (11), (12), в которой устраняется сингулярность в точке $u_2 = 0$ и уменьшается число настраиваемых параметров. Также исследуются различные варианты определения ошибки δ_t .

Для сравнения различных алгоритмов используется нелинейная неаффинная система следующего вида [7]

$$\begin{cases} \dot{x}_1 = x_2 + a_1 x_1 \cos(x_1 - 1), \\ \dot{x}_2 = -a_2 x_2 - a_3 \cos(x_2 - 1) - a_4 \sin(x_1 + \beta + u_1) + \beta + u_1, \end{cases} \quad (13)$$

где $a_1 = a_{10} + d_{a1}$, $a_{10} = 0.5$, $d_{a1} \in [-0.6 \ 0.6]$, $a_2 = a_{20} + d_{a2}$, $a_{20} = 0.2$, $d_{a2} \in [-0.4 \ 0.4]$, $a_3 = a_{30} + d_{a3}$, $a_{30} = 0.1$, $d_{a3} \in [-0.2 \ 0.2]$, $a_4 = a_{40} + d_{a4}$, $a_{40} = 4.9$, $d_{a4} \in [-6.0 \ 6.0]$.

В качестве номинального управления выбрано $u_1(r, x)$ используется управление из примера 4 работы [7].

Для того, чтобы устранить неопределенность в окрестности $u_2 \approx 0$ введем алгоритм настройки коэффициентов в виде

$$w_{t+1} = w_t - \alpha_w \delta_t \nabla u_2 \text{sign}(u_2). \quad (14)$$

Отметим, что алгоритм (14) требует только настройки параметра α_w в отличие от исходного алгоритма (11), в котором необходимо выбирать дополнительно ε_{grad} и ε_{lim} .

Отметим, что устойчивость системы (1) с номинальным управлением доказана в работе [7]. Так как алгоритм адаптивного управления (7)–(10), (12), (14) имеет дискретную природу, то представим модель (3) в дискретной форме

$$x_{t+1} = x_t + dt(A(x_t)x_t + b(x_t, u_{2t})u_{2t} + h(x_t, u_{2t}, v_t)v_t). \quad (15)$$

Рассмотрим следующую квадратичную функцию

$$V_t = x_t^T x_t + w_t^T w_t. \quad (16)$$

Вычислим приращение

$$\Delta V_t = V_{t+1} - V_t = x_{t+1}^T x_{t+1} + w_{t+1}^T w_{t+1} - x_t^T x_t - w_t^T w_t.$$

Принимая во внимание уравнения (14), (15) получаем

$$\begin{aligned} \Delta V_t = & [x_t + dt(A(x_t)x_t + b(x_t, u_{2t})u_{2t} + h(x_t, u_{2t}, v_t)v_t)]^T [x_t + dt(A(x_t)x_t + \\ & b(x_t, u_{2t})u_{2t} + h(x_t, u_{2t}, v_t)v_t)] + [w_t - \alpha_w \delta_t u_{\max} \Phi_{it} / \cosh^2(w_t^T \Phi_t) \text{sign}(u_2)]^T [w_t - \\ & \alpha_w \delta_t u_{\max} \Phi_{it} / \cosh^2(w_t^T \Phi_t) \text{sign}(u_2)] - x_t^T x_t - w_t^T w_t. \end{aligned}$$

Для стандартной дельта-ошибки $\delta_t = e_t^2 + \gamma e_t^2 - e_{t-1}^2$ при достаточно малом шаге dt и $r(t) = 0$, $\delta_t \approx -x_t^2$. Также учтем, что $\text{sign}(u_2) = \text{sign}(w_t^T \Phi_t)$. Тогда последнее выражение принимает вид

$$\Delta V \approx -2\alpha_w u_{\max}^2 x_t^T x_t \frac{|w_t^T \Phi_t|}{\cosh^2(w_t^T \Phi_t)} + \frac{\alpha_w^2 (x_t^T x_t)(x_t^T x_t) u_{\max}^4 \Phi_t^T \Phi_t}{\cosh^4(w_t^T \Phi_t)}. \quad (17)$$

В силу того, что $0 < \phi_{it} \leq 1$, можно заменить $\Phi_t^T \Phi_t$ на число Na . Также рассмотрим область $|x_t| \leq x_{\max}$. Тогда получим условие отрицательности приращения (17)

$$n \frac{\alpha_w x_{\max}^2 u_{\max}^2 Na}{\cosh^2(w_t^T \Phi_t)} < |w_t^T \Phi_t|. \quad (18)$$

Из выражения (18) следует, что в начальный момент времени при малых значениях элементов вектора w_t неравенство (18) не выполняется. Однако в ходе настройки значения элементов вектора w_t растут. Это приводит к уменьшению левой части и увеличению правой части неравенства (18). В некоторый момент времени оно начнет выполняться, и функция ΔV станет отрицательно полуопределенной. Таким образом, в замкнутой системе выполняется условие UUB.

Результаты. Рассмотрим результаты численного моделирования системы (13), (2), (6)–(12). Эта система соответствует случаю, когда адаптивное управление синтезируется согласно исходному алгоритму Actor-Critic. Дельта-ошибку сформируем согласно Q-методу обучения: $\delta_t = e_t^2 + \gamma e_t^2 - e_{t-1}^2$, где γ – коэффициент обесценивания.

Система управления с номинальным управлением u_1 сохраняет устойчивость системы, но присутствует статическая ошибка, обусловленная параметрической неопределенностью, как показано на рис. 1.

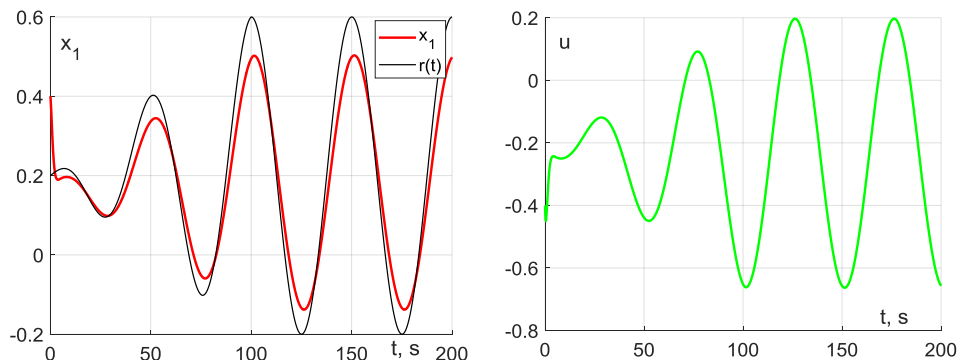


Рис. 1. Переходные процессы в системе с номинальным управлением $u = u_1(r, x)$

На рис. 1 задающий сигнал определяется выражением

$$r(t) = \begin{cases} 0.2 + 0.004t \sin(0.04\pi t + 0.5\pi), & t \leq 100, \\ 0.2 + 0.4 \sin(0.04\pi t + 0.5\pi), & t > 100. \end{cases} \quad (19)$$

Ошибка слежения для пиковых значений задающего сигнала составляет 0,1.

На рис. 2 представлены результаты моделирования системы (13), (2), (6)–(12) при $\alpha_w = 10^{-5}$, $\varepsilon_{lim} = 0.002$, $\varepsilon_{grad} = 10^{-6}$, $u_{\max} = 0.5$. На рис. 2 хорошо заметны моменты перехода адаптивного управления u_2 через ноль. Они приводят к резким непредсказуемым изменениям выходной переменной и ошибкам слежения.

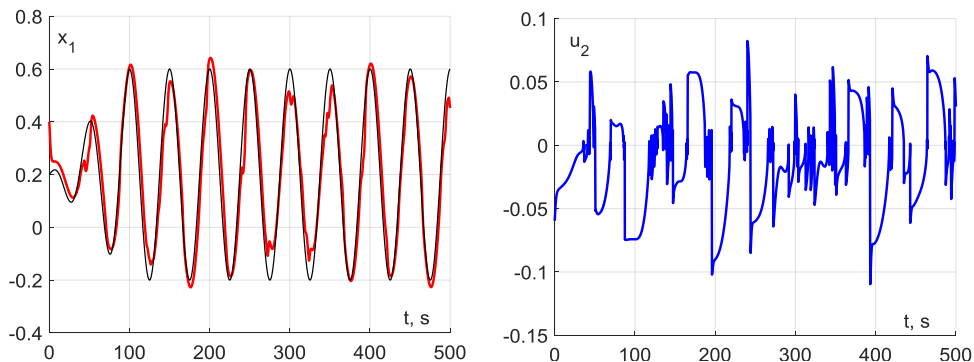


Рис. 2. Переходные процессы в системе (13), (2), (6)–(12)

Таким образом, исходный алгоритм настройки (12) обладает существенными недостатками, связанными с наличием особенности в точке $u_2 = 0$. Исследования показывают, что можно добиться некоторого снижения ошибки слежения за задающим сигналом. Однако, в целом, отмеченная проблема сохраняется при разных настройках параметров.

В частности, модификация дельта-ошибки к виду $\delta_t = e_t^2 + \alpha_{w1}(e_t^2 - e_{t-1}^2)$ и настройка параметров позволяет получить переходные процессы, представленные на рис. 3.

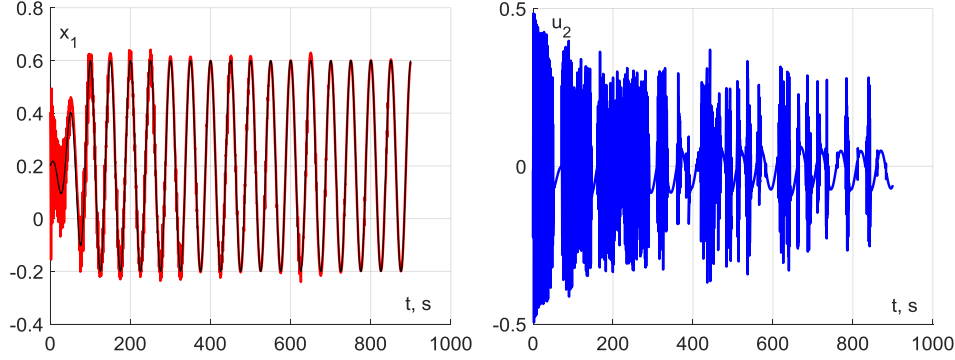


Рис. 3. Переходные процессы в системе (13), (2), (6)–(12) с ошибкой $\delta_t = e_t^2 + \alpha_{w1}(e_t^2 - e_{t-1}^2)$ при $\alpha_w = 0.01$, $\alpha_{w1} = 10000$, $\varepsilon_{lim} = 0.01$, $\varepsilon_{grad} = 10^{-7}$

Из рис. 3 видим, что ошибка слежения с течением времени убывает до значений меньше 0,01, однако наблюдаются высокочастотные колебания адаптивного управления.

Теперь исследуем систему управления с предложенным модифицированным алгоритмом настройки параметров, которая описывается уравнениями системы (13), (2), (6)–(10), (12), (14). На рис. 4 представлены результаты моделирования.

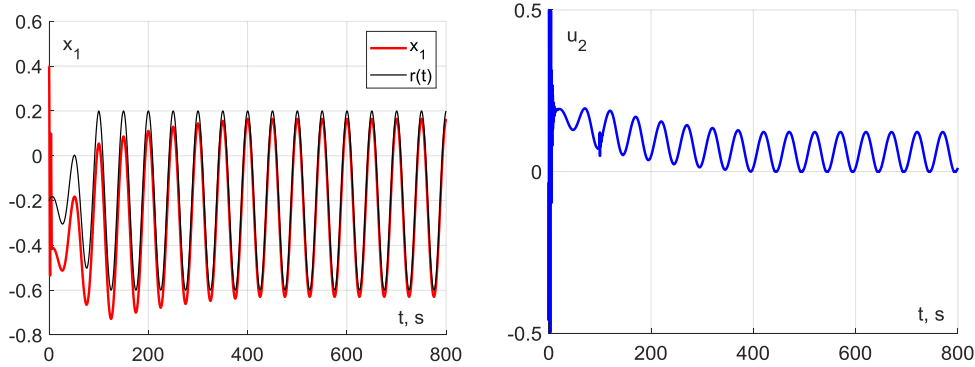


Рис. 4. Переходные процессы в системе (13), (2), (6) – (10), (12), (14) с ошибкой $\delta_t = e_t^2 + \alpha_{w1}(e_t^2 - e_{t-1}^2)$ при $\alpha_w = 0.01$, $\alpha_{w1} = 10000$

Проведенное моделирование показывает большую длительность настройки вектора параметров w_t . Для уменьшения длительности настройки предлагается использовать дельта-ошибку вида $\delta_t = e_t^2 - e_{t-1}^2$ и обучающие импульсы для настройки системы.

Результаты моделирования при задающем воздействии вида

$$r(t) = \begin{cases} \pm 0.5, & t \leq 30, \\ 0.2 + 0.002(t - 30) \sin(0.04\pi(t - 30) + 0.5\pi), & t > 30, \end{cases} \quad (20)$$

представлены на рис. 5.

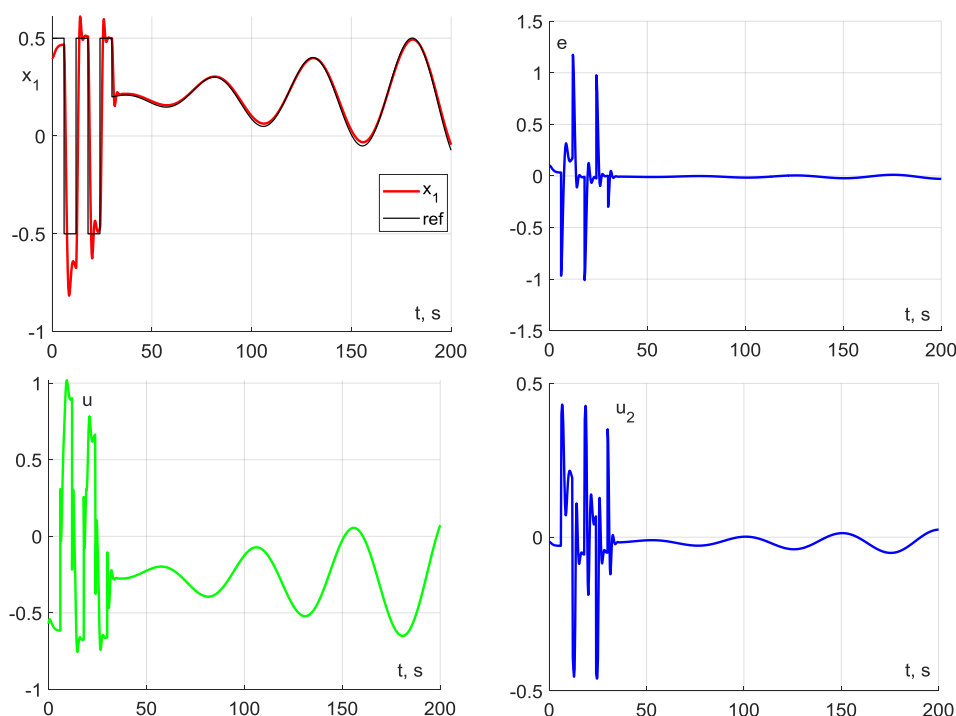


Рис. 5. Переходные процессы в системе (13), (2), (6)–(10), (12), (14) с ошибкой $\delta_t = e_t^2 - e_{t-1}^2$ при $\alpha_w = 1.0$

При моделировании число радиально-базисных функций взято равным 10.

Из рис. 5 видим, что для настройки системы достаточно несколько обучающих импульсов. В примере использовано 5 переходных процессов длительностью 6 секунд. Отметим, что обучающие импульсы должны иметь амплитуду, сравнимую с диапазоном возможного изменения ошибки.

Заключение. В статье проведен обзор и анализ базирующихся на методах обучения с подкреплением работ в области адаптивного управления нелинейными динамическими объектами. На основе проведенного анализа предложено для нелинейных неаффинных объектов формировать управление в виде совокупности номинального и адаптивного управления. Синтез адаптивного управления проведен на основе Q-метода обучения с использованием архитектуры Actor-Critic. Для аппроксимации неизвестной структуры адаптивного управления использована аппроксимация сетью радиально-базисных функций Гауссовского типа. Предложен модифицированный алгоритм настройки параметров сети, в котором, в отличие от оригинального Q-метода со структурой Actor-Critic, отсутствует особенность в нулевой точке.

Дальнейшие направления исследований связаны с оптимизацией режимов настройки адаптивной системы, исследованием различных критериев качества, используемых в адаптивном управлении, рассмотрением объектов с векторным входом и выходом. Необходимость решения указанных задач обусловлена тем, что не все объекты допускают подачу импульсов большой амплитуды. Также в данной статье в качестве критерия использовано мгновенное значение ошибки слежения за задающим сигналом. Во многих приложениях необходимо оптимизировать интегральные критерии, включающие также ограничения на производную ошибки.

Исследование выполнено за счет гранта Российского научного фонда № 25-61-00017, «Интеллектуальные методы траекторного управления робототехническими комплексами в условиях параметрических и внешних возмущений», <https://rscf.ru/project/25-61-00017/> на базе ФГАОУ ВО «Южный федеральный университет».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Zemlyakov S., Rutkovsky V.* Synthesizing self-adjusting control systems with a standard model // *Automation and Remote Control.* – 1966. – Vol. 27 (3). – P. 407-414.
2. *Parks P.* Liapunov Redesign of Model Reference Adaptive Control Systems // *IEEE Transactions on Automatic Control.* – 1966. – Vol. 11 (3). – P. 362-367.
3. *Стафейчук Б.Г., Шакирова А.Я.* Исследование адаптивной системы автоматического регулирования с применением нейросетевых технологий на имитационной модели реактора // *Вестник ПНИПУ. Химическая технология и биотехнология.* – 2019. – № 2.
4. *Еременко Ю.И., Глуценко А.И., Фомин А.В.* О применении нейросетевого настройщика параметров ПИ-регулятора на тепловых объектах горно-металлургической отрасли в режиме отработки возмущений // *ГИАБ.* – 2017. – № 12.
5. *Санталов А.А.* Нейросетевая настройка адаптивного ПИД-регулятора мощности гидроагрегата // *Вестник УЛГТУ.* – 2021. – № 3 (95).
6. *Саттон Р.С., Барто Э.Дж.* Обучение с подкреплением: Введение. – 2-е изд.: пер. с англ. А.А. Слинкина. – М.: ДМК Пресс, 2020. – 552 с.
7. *Gaiduk A.R., Medvedev M.Y. and Pshikhov V.K.* Design Method of Quasilinear Nonaffine Nonlinear Control Systems of General Structure // *IEEE Transactions on Automation Science and Engineering.* – 2025. – Vol. 22. – P. 20208-20220.
8. *Гайдук А.Р., Медведев М.Ю., Пишихов В.Х., Гусцов В.Г.* Синтез неаффинных нелинейных систем управления на основе квазилинейных моделей // *Мехатроника, автоматизация, управление.* – 2025. – 26(5). – С. 223-232. – <https://doi.org/10.17587/mau.26.223-232>.
9. *Карапеев А.Н., Косенко Е.Ю., Медведев М.Ю., Пишихов В.Х.* Исследование интеллектуального адаптивного алгоритма управления на базе метода обучения с подкреплением // *Известия ЮФУ. Технические науки.* – 2025. – № 2. – С. 162-175.
10. *Медведев М.Ю., Пишихов В.Х., & Евдокимов И.Д.* Алгоритм робастного управления одномерным динамическим объектом на основе табличного Q-метода обучения с подкреплением // *Информатика и автоматизация.* – 2025. – 24(3). – P. 717-744. – <https://doi.org/10.15622/ia.24.3.1>.
11. *Xu B., Yang C. and Shi Z.* Reinforcement Learning Output Feedback NN Control Using Deterministic Learning Technique // *IEEE Transactions on Neural Networks and Learning Systems.* – 2014. – Vol. 25 (3). – P. 635-641. – DOI: 10.1109/TNNLS.2013.2292704.
12. *Mu C., Ni Z., Sun C., and He H.* Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems // *IEEE Transactions on Cybernetics.* – 2016. – Vol. 47 (6). – P. 1460-1470. – DOI: 10.1109/TCYB.2016.2548941.
13. *Wang A., Liao X., Dong T.* Event-driven optimal control for uncertain nonlinear systems with external disturbance via adaptive dynamic programming // *Neurocomputing.* – 2018. – Vol. 281. – P. 188-195. – DOI: 10.1016/j.neucom.2017.12.010.
14. *Biao Luo, Yin Yang, Derong Liu, and Huai-Ning Wu.* Event-Triggered Optimal Control with Performance Guarantees Using Adaptive Dynamic Programming // *IEEE Transactions on Neural Networks and Learning Systems.* – 2020. – Vol. 31(1). – P. 76-88. – DOI: 10.1109/TNNLS.2019.2899594.
15. *Yang X., Xu M. and Wei Q.* Dynamic Event-Sampled Control of Interconnected Nonlinear Systems Using Reinforcement Learning // *IEEE Transactions on Neural Networks and Learning Systems.* – 2024. – Vol. 35(1). – P. 923-937. – DOI: 10.1109/TNNLS.2022.3178017.
16. *Zhang H., Zhao X., Wang H., Zong G. and Xu N.* Hierarchical Sliding-Mode Surface-Based Adaptive Actor-Critic Optimal Control for Switched Nonlinear Systems With Unknown Perturbation // *IEEE Transactions on Neural Networks and Learning Systems.* – 2024. – Vol. 35(2). – P. 1559-1571. – DOI: 10.1109/TNNLS.2022.3183991.
17. *Dong C., Chen L. and Dai S.-L.* Performance-Guaranteed Adaptive Optimized Control of Intelligent Surface Vehicle Using Reinforcement Learning // *IEEE Transactions on Intelligent Vehicles.* – 2024. – Vol. 9 (2). – P. 3581-3592. – DOI: 10.1109/TIV.2023.3338486.
18. *Phuong Nam Dao, Minh Hiep Phung.* Nonlinear robust integral based actor-critic reinforcement learning control for a perturbed three-wheeled mobile robot with mecanum wheels // *Computers and Electrical Engineering.* – 2025. – Vol. 121. 109870. – DOI:10.1016/j.compeleceng.2024.109870.
19. *Ding Wang, Xin P., Ren J., and Qiao J.* Adaptive Critic Tracking Design for Data-Based Nonaffine Predictive Control // *IEEE Transactions on Automation Science and Engineering.* – 2024. – Vol. 21 (4). – P. 5534-5545. – DOI: 10.1109/TASE.2023.3313159.
20. *Berkenkamp F., Turchetta M., Schoellig A., Krause A.* Safe model-based reinforcement learning with stability guarantees // *Advances in Neural Information Processing Systems.* – 2017. – Vol. 30. – P. 908-918. – DOI: 10.48550/arXiv.1705.08551.

21. Han M., Zhang L., Wang J., Pan W. Actor-critic reinforcement learning for control with stability guarantee // *IEEE Robotics and Automation Letters*. – 2020. – Vol. 5 (4). – P. 6217-6224. – DOI: 10.1109/LRA.2020.3011351.
22. Zhou Z., Liu A., and Wang D. Improved adaptive-critic-based dynamic event-triggered control for non-affine systems // *Mechatronics Tech.* – 2024. – Vol. 1 (1). – DOI: 10.55092/mt20230002.
23. Cheng R., Orosz G., Murray R.M., Burdick J.W. End-to end safe reinforcement learning through barrier functions for safety critical continuous control tasks // *The Thirty-Third AAAI Conference on Artificial Intelligence*. – 2019. – P. 3387-3395.
24. Choi J., Castaneda F., Tomlin C.J., Sreenath K. Reinforcement learning for safety-critical control under model uncertainty, using control Lyapunov functions and control barrier functions // *Conference Robotics: Science and Systems*. – 2020.
25. Xue W., Lian B., Fan J., Kolaric P., Cha T., Lewis F. Inverse Reinforcement Q-Learning Through Expert Imitation for Discrete-Time Systems // *IEEE Transactions on Neural Networks and Learning Systems*. – 2021. – P. 1-14. – DOI: 10.1109/TNNLS.2021.3106635.
26. Yu Zhang, Niu B., Zhao X., Duan P., Wang H., and Gao B. Global Predefined-Time Adaptive Neural Network Control for Disturbed Pure-Feedback Nonlinear Systems with Zero Tracking Error // *IEEE Transactions on Neural Networks and Learning Systems*. – 2023. – Vol. 34(9). – P. 6328-6338.
27. Wang D., Zhou Z., Li M., Ren J., and Qia J. Event-based robust performance guarantee for nonaffine plants via system identification // *International Journal Robust Nonlinear Control*. – 2023. – Vol. 33. – P. 53650-5387. – DOI: 10.1002/rnc.6646.
28. Zanon M. Gros S. Safe Reinforcement Learning Using Robust MPC // *IEEE Transactions on Automatic Control*. – 2020. – P. 1-1. – DOI: 10.1109/TAC.2020.3024161.
29. Rizvi S. A.A., Lin Z. Output Feedback Q-Learning Control for the Discrete-Time Linear Quadratic Regulator Problem // *IEEE Transactions on Neural Networks and Learning Systems*. – 2018. – P. 1-14. – DOI: 10.1109/TNNLS.2018.2870075.
30. Kim J.W., Oh T.H., Son S.H., Jeong D.H., Lee J.M. Convergence analysis of the deep neural networks based globalized dual heuristic programming // *Automatica*. – 2020. – Vol. 122. – DOI: 10.1016/j.automatica.2020.109222.
31. Боровик В.С., Шидловский С.В. Обучение с подкреплением в системах управления объектами с транспортным запаздыванием // *Автометрия*. – 2021. – Т. 57(3). – С. 48-57. – DOI: 10.15372/AUT20210306.
32. Галеев А.А., Медведев А.И., Насонов И.А. Нейросетевой алгоритм перехвата машиной Дубинса целей, движущихся по известным траекториям // *Автоматика и телемеханика*. – 2023. – № 3. – С. 3-21. – DOI: 10.31857/S0005231023030017.
33. Ханкин Д.Л., Феофилов С.В. Синтез устойчивых нейросетевых регуляторов для объектов с ограничителями в условиях неполной информации // *Мехатроника, автоматизация, управление*. – 2024. – Vol. 25(7). – P. 345-353. – <https://doi.org/10.17587/mau.25.345-353>.
34. Pshikhov V., Krukhmalev V., Medvedev M., and Neydorf R. Estimation of Energy Potential for Control of Feeder of Novel Cruiser/Feeder MAAT System // *SAE Technical Paper*. – 2012. – Vol. 2012-01-2099. – doi: 10.4271/2012-01-2099.
35. Пилюхов В.Х., Медведев М.Ю., Гайдук А.Р., Нейдорф Р.А., Беляев В.Е., Федоренко Р.В., Костюков В.А., Крехмалев В.А. Система позиционно-траекторного управления роботизированной воздухоплавательной платформой: алгоритмы управления // *Мехатроника, автоматизация и управление*. – 2013. – № 7. – С. 13-20.
36. Гайдук А.Р., Пилюхов В.Х., Медведев М.Ю., Гусцов В.Г. Непрерывное управление нелинейными неаффинными объектами // *Известия ЮФУ. Технические науки*. – 2024. – № 1 (237). – С. 122-133.
37. Медведев М.Ю., Пилюхов В.Х., Медведев И.М. Алгоритмы адаптивного управления каскадными динамическими объектами на базе метода обучения с подкреплением // *Мехатроника, автоматизация и управление*. – 2026. – № 3.

REFERENCES

1. Zemlyakov S., Rutkovsky V. Synthesizing self-adjusting control systems with a standard model, *Automation and Remote Control*, 1966, Vol. 27(3), pp. 407-414.
2. Parks P. Liapunov Redesign of Model Reference Adaptive Control Systems, *IEEE Transactions on Automatic Control*, 1966, Vol. 11(3), pp. 362-367.
3. Stafeychuk B.G., Shakirova A.Ya. Issledovanie adaptivnoy sistemy avtomaticheskogo regulirovaniya s primeneniem neyrosetevykh tekhnologiy na imitatsionnoy modeli reaktora [Study of an adaptive automatic control system using neural network technologies on a simulation model of a reactor], *Vestnik PNIPU. Khimicheskaya tekhnologiya i biotekhnologiya* [Bulletin of PNIPU. Chemical technology and biotechnology], 2019, No. 2.

4. Eremenko Yu.I., Glushchenko A.I., Fomin A.V. O primeneniі neyrosetevogo nastroyshchika parametrov PI-regulyatora na teplovykh ob'ektakh gorno-metallurgicheskoy otrasli v rezhime otrabotki vozmushcheniy [On the use of a neural network tuner for PI controller parameters at thermal facilities in the mining and metallurgical industry in the disturbance processing mode], *GIAB* [GIAB], 2017, No. 12.
5. Santalov A.A. Neyrosetevaya nastroyka adaptivnogo PID-regulyatora moshchnosti gidroagregata [Neural network tuning of an adaptive PID controller for the power of a hydroelectric unit], *Vestnik UIGTU* [Bulletin of UISTU], 2021, No. 3 (95).
6. Sutton R.S., Barto E.Dzh. Obucheniye s podkreplenіem: Vvedeniye [Reinforcement Learning: An Introduction. 2nd ed.: transl. from Engl. A.A. Slinkina. Moscow: DMC Press, 2020, 552 p.
7. Gaiduk A.R., Medvedev M.Y. and Pshikhov V.K. Design Method of Quasilinear Nonaffine Nonlinear Control Systems of General Structure, *IEEE Transactions on Automation Science and Engineering*, 2025, Vol. 22, pp. 20208-20220.
8. Gayduk A.R., Medvedev M.Yu., Pshikhov V.Kh., Gistsov V.G. Sintez neaffinnykh nelineynykh sistem upravleniya na osnove kvazilineynykh modeley [Design of nonaffine nonlinear control systems based on quasilinear models], *Mekhatronika, avtomatizatsiya, upravleniye* [Mekhatronika, Avtomatizatsiya, Upravleniye], 2025, 26(5), pp. 223-232. Available at: <https://doi.org/10.17587/mau.26.223-232>.
9. Karapeev A.N., Kosenko E.Yu., Medvedev M.Yu., Pshikhov V.Kh. Issledovaniye intellektual'nogo adaptivnogo algoritma upravleniya na baze metoda obucheniya s podkreplenіem [Study of an intelligent adaptive control algorithm based on the reinforcement learning method], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2025, No. 2, pp. 162-175.
10. Medvedev M.Yu., Pshikhov V.Kh., & Evdokimov I.D. Algoritm robastnogo upravleniya odnomernym dinamicheskim ob'ektom na osnove tablitsnogo Q-metoda obucheniya s podkreplenіem [Robust control algorithm for single input single output dynamic object based on table-based Q-method of reinforcement learning], *Informatika i avtomatizatsiya* [Informatics and Automation], 2025, 24 (3), pp. 717-744. Available at: <https://doi.org/10.15622/ia.24.3.1>.
11. Xu B., Yang C. and Shi Z. Reinforcement Learning Output Feedback NN Control Using Deterministic Learning Technique, *IEEE Transactions on Neural Networks and Learning Systems*, 2014, Vol. 25 (3), pp. 635-641. DOI: 10.1109/TNNLS.2013.2292704.
12. Mu C., Ni Z., Sun C., and He H. Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems, *IEEE Transactions on Cybernetics*, 2016, Vol. 47 (6), pp. 1460-1470. DOI: 10.1109/TCYB.2016.2548941.
13. Wang A., Liao X., Dong T. Event-driven optimal control for uncertain nonlinear systems with external disturbance via adaptive dynamic programming, *Neurocomputing*, 2018, Vol. 281, pp. 188-195. DOI: 10.1016/j.neucom.2017.12.010.
14. Biao Luo, Yin Yang, Derong Liu, and Huai-Ning Wu. Event-Triggered Optimal Control with Performance Guarantees Using Adaptive Dynamic Programming, *IEEE Transactions on Neural Networks and Learning Systems*, 2020, Vol. 31 (1), pp. 76-88. DOI: 10.1109/TNNLS.2019.2899594.
15. Yang X., Xu M. and Wei Q. Dynamic Event-Sampled Control of Interconnected Nonlinear Systems Using Reinforcement Learning, *IEEE Transactions on Neural Networks and Learning Systems*, 2024, Vol. 35(1), pp. 923-937. DOI: 10.1109/TNNLS.2022.3178017.
16. Zhang H., Zhao X., Wang H., Zong G. and Xu N. Hierarchical Sliding-Mode Surface-Based Adaptive Actor-Critic Optimal Control for Switched Nonlinear Systems With Unknown Perturbation, *IEEE Transactions on Neural Networks and Learning Systems*, 2024, Vol. 35 (2), pp. 1559-1571. DOI: 10.1109/TNNLS.2022.3183991.
17. Dong C., Chen L. and Dai S.-L. Performance-Guaranteed Adaptive Optimized Control of Intelligent Surface Vehicle Using Reinforcement Learning, *IEEE Transactions on Intelligent Vehicles*, 2024, Vol. 9(2), pp. 3581-3592. DOI: 10.1109/TIV.2023.3338486.
18. Phuong Nam Dao, Minh Hiep Phung. Nonlinear robust integral based actor-critic reinforcement learning control for a perturbed three-wheeled mobile robot with mecanum wheels, *Computers and Electrical Engineering*, 2025, Vol. 121. 109870. DOI:10.1016/j.compeleceng.2024.109870.
19. Ding Wang, Xin P., Ren J., and Qiao J. Adaptive Critic Tracking Design for Data-Based Nonaffine Predictive Control, *IEEE Transactions on Automation Science and Engineering*, 2024, Vol. 21(4), pp. 5534-5545. DOI: 10.1109/TASE.2023.3313159.
20. Berkenkamp F., Turchetta M., Schoellig A., Krause A. Safe model-based reinforcement learning with stability guarantees, *Advances in Neural Information Processing Systems*, 2017, Vol. 30, pp. 908-918. DOI: 10.48550/arXiv.1705.08551.
21. Han M., Zhang L., Wang J., Pan W. Actor-critic reinforcement learning for control with stability guarantee, *IEEE Robotics and Automation Letters*, 2020, Vol. 5 (4), pp. 6217-6224. DOI: 10.1109/LRA.2020.3011351.

22. Zhou Z., Liu A., and Wang D. Improved adaptive-critic-based dynamic event-triggered control for non-affine systems, *Mechatronics Tech.*, 2024, Vol. 1 (1). DOI: 10.55092/mt20230002.
23. Cheng R., Orosz G., Murray R.M., Burdick J.W. End-to end safe reinforcement learning through barrier functions for safety critical continuous control tasks, *The Thirty-Third AAAI Conference on Artificial Intelligence*, 2019, pp. 3387-3395.
24. Choi J., Castaneda F., Tomlin C.J., Sreenath K. Reinforcement learning for safety-critical control under model uncertainty, using control Lyapunov functions and control barrier functions, *Conference Robotics: Science and Systems*, 2020.
25. Xue W., Lian B., Fan J., Kolaric P., Cha T., Lewis F. Inverse Reinforcement Q-Learning Through Expert Imitation for Discrete-Time Systems, *IEEE Transactions on Neural Networks and Learning Systems*, 2021, pp. 1-14. 10.1109/TNNLS.2021.3106635.
26. Yu Zhang, Niu B., Zhao X., Duan P., Wang H., and Gao B. Global Predefined-Time Adaptive Neural Network Control for Disturbed Pure-Feedback Nonlinear Systems with Zero Tracking Error, *IEEE Transactions on Neural Networks and Learning Systems*, 2023, Vol. 34 (9), pp. 6328-6338.
27. Wang D., Zhou Z., Li M., Ren J., and Qia J. Event-based robust performance guarantee for nonaffine plants via system identification, *International Journal Robust Nonlinear Control*, 2023, Vol. 33, pp. 53650-5387. DOI: 10.1002/rnc.6646.
28. Zanon M. Gros S. Safe Reinforcement Learning Using Robust MPC, *IEEE Transactions on Automatic Control*, 2020, pp. 1-1. 10.1109/TAC.2020.3024161.
29. Rizvi S. A.A., Lin Z. Output Feedback Q-Learning Control for the Discrete-Time Linear Quadratic Regulator Problem, *IEEE Transactions on Neural Networks and Learning Systems*, 2018, pp. 1-14. 10.1109/TNNLS.2018.2870075.
30. Kim J.W., Oh T.H., Son S.H., Jeong D.H., Lee J.M. Convergence analysis of the deep neural networks based globalized dual heuristic programming, *Automatica*, 2020, Vol. 122. DOI: 10.1016/j.automatica.2020.109222.
31. Borovik V.S., Shidlovskiy S.V. Obucheniye s podkrepleniem v sistemakh upravleniya ob'ektami s transportnym zapazdyvaniem [Reinforcement learning in plant control systems with transport lag], *Avtometriya [Autometry]*, 2021, Vol. 57 (3), pp. 48-57. DOI: 10.15372/AUT20210306.
32. Galyaev A.A., Medvedev A.I., Nasonov I.A. Neyrosetevoy algoritm perekhvata mashinoy Dubinsa tseley, dvizhushchikhsya po izvestnym traektoriyam [Neural network algorithm for intercepting targets moving along known trajectories by a Dubins' car], *Avtomatika i telemekhanika [Automation and Telemekhanics]*, 2023, No. 3, pp. 3-21. DOI: 10.31857/S0005231023030017.
33. Khapkin D.L., Feofilov S.V. Sintez ustoychivyykh neyrosetevykh regulyatorov dlya ob'ektov s ogranichitelyami v usloviyakh nepolnoy informatsii [Synthesis of stable neural network controllers for objects with constraints under conditions of incomplete information], *Mekhatronika, avtomatizatsiya, upravlenie [Mekhatronika, Avtomatizatsiya, Upravlenie]*, 2024, Vol. 25 (7), pp. 345-353. Available at: <https://doi.org/10.17587/mau.25.345-353>.
34. Pshikhov V., Krukhmalev V., Medvedev M., and Neydorf R. Estimation of Energy Potential for Control of Feeder of Novel Cruiser/Feeder MAAT System, *SAE Technical Paper*, 2012, Vol. 2012-01-2099. doi: 10.4271/2012-01-2099.
35. Pshikhov V.Kh., Medvedev M.Yu., Gayduk A.R., Neydorf R.A., Belyaev V.E., Fedorenko R.V., Kostyukov V.A., Krukhmalev V.A. Sistema pozitsionno-traektnogo upravleniya robotizirovannoy vozdukhoplavatel'noy platformoy: algoritmy upravleniya [Position-trajectory control system of a robotic aeronautical platform: control algorithms], *Mekhatronika, avtomatizatsiya i upravlenie [Mekhatronika, Avtomatizatsiya, Upravlenie]*, 2013, No. 7, pp. 13-20.
36. Gayduk A.R., Pshikhov V.Kh., Medvedev M.Yu., Gistsov V.G. Nepreryvnoye upravlenie nelineynymi neaffinnymi ob'ektami [Continuous control of nonlinear non-affine objects], *Izvestiya YuFU. Tekhnicheskie nauki [Izvestiya SFedU. Engineering Sciences]*, 2024, No. 1 (237), pp. 122-133.
37. Medvedev M.Yu., Pshikhov V.Kh., Medvedev I.M. Algoritmy adaptivnogo upravleniya kaskadnymi dinamicheskimi ob'ektami na baze metoda obucheniya s podkrepleniem [Algorithms for adaptive control of cascade dynamic objects based on the reinforcement learning method], *Mekhatronika, avtomatizatsiya i upravlenie [Mekhatronika, Avtomatizatsiya, Upravlenie]*, 2026, No. 3

Медведев Михаил Юрьевич – НИИ робототехники и процессов управления Южного федерального университета; e-mail: medvmihal@sfedu.ru; г. Таганрог, Россия; тел.: 88634371694; д.т.н.; ведущий научный сотрудник.

Пшихов Вячеслав Хасанович – НИИ робототехники и процессов управления Южного федерального университета; e-mail: pshichop@rambler.ru; г. Таганрог, Россия; тел.: 88634371694; д.т.н.; профессор; директор.

Гайдук Анатолий Романович – НИИ робототехники и процессов управления Южного федерального университета; e-mail: gaiduk_2003@mail.ru; г. Таганрог, Россия; тел.: 88634371694; д.т.н.; профессор; ведущий научный сотрудник.

Медведев Илья Михайлович – Южный федеральный университет; e-mail: imedvede@sfedu.ru; г. Таганрог, Россия; студент.

Косенко Евгений Юрьевич – НИИ робототехники и процессов управления Южного федерального университета; e-mail: ekosenko@sfedu.ru; г. Таганрог, Россия; тел.: 88634371694; к.т.н.; старший научный сотрудник.

Medvedev Mikhail Yurjevich – R&D Institute of Robotics and Control Systems, Southern Federal University; e-mail: medvmihal@sfedu.ru; Taganrog, Russia; phone: 88634371694; dr. of eng. sc.; leading researcher.

Pshikhov Viacheslav Khasanovich – R&D Institute of Robotics and Control Systems, Southern Federal University; e-mail: pshichop@rambler.ru; Taganrog, Russia; phone: 88634371694; dr. of eng. sc.; professor; director.

Gaiduk Anatoly Romanovich – R&D Institute of Robotics and Control Systems, Southern Federal University; e-mail: gaiduk_2003@mail.ru; Taganrog, Russia; phone: 88634371694; dr. of eng. sc.; professor; leading researcher.

Medvedev Ilya Mikhailovich – Southern Federal University; e-mail: imedvede@sfedu.ru; Taganrog, Russia; student.

Kosenko Evgeniy Yurjevich – R&D Institute of Robotics and Control Systems, Southern Federal University; e-mail: ekosenko@sfedu.ru; Taganrog, Russia; phone: 88634371694; cand. of eng. sc.; senior researcher.

УДК 004.896+629.58+001.891.57

DOI 10.18522/2311-3103-2026-2-201-212

В.С. Быкова, А.И. Машошин

АЛГОРИТМ УПРАВЛЕНИЯ АВТОНОМНЫМ НЕОБИТАЕМЫМ ПОДВОДНЫМ АППАРАТОМ ПРИ ПОИСКЕ НАЗНАЧЕННОГО ДОННОГО ОБЪЕКТА ПРИ КОМПЛЕКСНОМ ИСПОЛЬЗОВАНИИ РАЗЛИЧНЫХ СРЕДСТВ МОНИТОРИНГА ДНА

Поиск назначенных донных объектов является одной из наиболее сложных задач, решаемых АНПА, что обусловлено рядом факторов, основными из которых являются: разнообразие объектов поиска (затонувшие подводные лодки, надводные корабли, самолёты, вертолёты, мины, подводные трубопроводы и кабели связи, различная подводная инфраструктура), необходимостью комплексного использования для поиска различных средств мониторинга дна, отличающихся физическими принципами функционирования, разрешающей способностью и поисковой производительностью. Целью работы, результаты которой приведены в статье, была разработка алгоритма управления АНПА при поиске назначенного донного объекта, удовлетворяющего этим требованиям, и его проверка с использованием цифрового полигона и цифрового двойника АНПА. В качестве критерия выбора средства мониторинга дна в каждом конкретном случае была принята вероятность правильного отнесения обнаруженного донного объекта к объекту поиска. В результате логика поиска назначенного донного объекта выглядит следующим образом. Осуществляется поиск донных объектов средством с максимальной производительностью поиска. При обнаружении донного объекта определяется вероятность его принадлежности объекту поиска. Если она больше заданного высокого порога, принимается решение об обнаружении назначенного донного объекта. Если же она меньше заданного низкого порога, принимается решение, что обнаружен посторонний объект. В остальных случаях принимается решение о необходимости обследования объекта средством с большей разрешающей способностью. Описана технология классификации донных объектов на основе обучения искусственной нейронной сети с использованием синтезированного обучающего материала. Приведены результаты проверки эффективности разработанного алгоритма с использованием цифрового полигона и цифрового двойника АНПА. Выполненное моделирование разработанного алгоритма показало, что комплексное использование средств мониторинга дна повышает вероятность успешного решения задачи и сокращает время её решения.

АНПА; система управления АНПА; поиск назначенных донных объектов; алгоритм классификации донных объектов; искусственная нейронная сеть; цифровой полигон; цифровой двойник.