Раздел II. Системы управления и моделирование

УДК 004.896 DOI 10.18522/2311-3103-2025-2-162-175

А.Н. Карапеев, Е.Ю. Косенко, М.Ю. Медведев, В.Х. Пшихопов

ИССЛЕДОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОГО АДАПТИВНОГО АЛГОРИТМА УПРАВЛЕНИЯ НА БАЗЕ МЕТОДА ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Предложен и исследован алгоритм адаптивного управления двигателем постоянного тока, базирующийся на применении технологии машинного обучения с подкреплением. Дан обзор и краткий анализ состояния дел в области интеллектуальных систем управления приводами. Представлено математическое описание двигателя, приведена структурная схема обучения интеллектуального агента. Предложена интеллектуальная система адаптивного управления скоростью вращения двигателя, при построении которой двигатель представляется в виде черного ящика с заданными ограничениями на вход и выход. Система управления строится на базе алгоритма Q-обучения нулевого порядка. Предполагается, что выходом интеллектуального агента является управление, подаваемое на вход двигателя. Экспериментальным путем подобрано множество таких управлений, позволяющих реализовать заданную точность поддержания частоты вращения. В интеллектуальной системе используются приближенные табличные оценки ценности каждого из управлений в зависимости от желаемой и текущей частоты вращения двигателя. В настоящей статье проведено исследование влияния дискретности представления значений состояния, используемого множества управляющих воздействий, применяемых вознаграждений, а также параметров алгоритма обучения на ошибку управления. Исследована чувствительность интеллектуальной системы управления к изменению параметров моделируемого двигателя и не измеряемому моменту сопротивления на валу двигателя. По результатам проведенного исследования сделан вывод о необходимости использования модифицированного алгоритма, в котором предполагается измерение или оценка тока статорной обмотки двигателя и использование непрерывного упарвления. В данной постановке задачи алгоритм управления обеспечивает робастность к переменным параметрам и внешнему возмущению. Также обсуждаются вопросы аппроксимации функции ценности управления с помощью полиномов и с применением нейронной сети. Показана возможность высокой точности аппроксимации с помощью нейронной сети простой структуры.

Адаптивное управление двигателем; обучение с подкреплением; робастное управление; Q-алгоритм обучения; нейронная сеть.

A.N. Karapeev, E.Y. Kosenko, M.Yu. Medvedev, V.Kh. Pshikhopov

RESEARCH OF AN INTELLIGENT ADAPTIVE CONTROL ALGORITHM BASED ON THE REINFORCEMENT LEARNING METHOD

An algorithm for adaptive control of a DC motor based on the use of machine learning technology with reinforcement is proposed and investigated. An overview and brief analysis of the state of affairs in the field of intelligent motor control systems is given. A mathematical model of the DC motor is presented, and a structural scheme for training an intellectual agent is presented. An intelligent adaptive motor speed control system is proposed. The DC motor is represented as a black box with the limited input and output. The control system is based on a zero-order Q-learning algorithm. It is assumed that the output of the intelligent agent is a control applied to the motor input. The intelligent system uses a tabular approximation of the value of each of the control action. In this article, we study the effect of the discreteness of the representation of state, the set of control effects used, the applied rewards, and the parameters of the learning algorithm on the control error. The sensitivity of the control system to the parameters of the motor and an unmeasured moment is investigated. Based on the results of the study, a modified algorithm is proposed, which assumes the measurement or evaluation of the current of the motor stator. The control algorithm provides robustness to parameters and external disturbance. Additionally, the approximation of the control value function using polynomials and using a neural network are investigated.

Adaptive control of DC motor; reinforcement learning; robust control; Q-algorithm; neural network.

Введение. В настоящее время набирают популярность методы интеллектуального планирования и управления, базирующиеся на машинном обучении. Наилучшие результаты алгоритмы, основанные на методах обучения, достигнуты в обработке изображений, текстов и сигналов [1–6]. В указанных областях результаты, которые демонстрируют алгоритмы, базирующиеся на глубоких нейронных сетях, превосходят возможности человека. Также глубокие нейронные сети успешно решают задачи планирования пути по заданной карте [7, 8].

С развитием методов обучения с подкреплением [9] глубокие нейронные сети получили возможность решать задачи планирования и управления в динамических системах [10–12]. При этом могут использоваться как относительно простые Q-методы глубокого обучения с подкреплением [13–15], так и более сложные методы, такие как детерминированный градиент политики DPG [16, 17].

Наибольшую сложность представляют собой задачи управления динамическими системами с гарантией асимптотической устойчивости или ограниченности сигнала ошибки. При этом основную проблему представляет теоретическое обоснование устойчивости замкнутой системы с нейросетевым регулятором [18]. В той связи наиболее распространенной схемой применения нейронных сетей является их сочетание с классическими регуляторами [19–21].

Текущий уровень практических задач, которые решаются методам глубокого обучения с подкреплением, говорят о возможности их применения для управления динамическими объектами [9, 22].

Актуальность настоящего исследования обусловлена следующими факторами. В настоящее время существует большое количество методов адаптивного и робастного управления динамическими объектами [23, 24]. В частности, в системах управления приводами существуют промышленные технологии автоматической настройки ПИД-регуляторов, которые обеспечивают требуемое качество функционирования в заданных режимах. Однако, существующие аналитические алгоритмы адаптивного управления базируются на математических моделях, в которых выделяются неопределенные параметры, структурные элементы и возмущения. Кроме того, адаптивный подход требует определения эталонных свойств замкнутой системы управления, что в отсутствии математической модели объекта вызывает затруднения.

Основное преимущество глубокого обучения с подкреплением – возможность разработать алгоритм управления динамическим объектом, как черным ящиком, на основе минимально доступной информации. Обычно требуется знать ограничения на вход и выход объекта, а также рабочую область пространства состояний динамической системы.

Вместе с тем, кроме проблемы обеспечения асимптотической устойчивости желаемого положения равновесия замкнутой системы, существует ряд дополнительных сложностей при применении методов обучения с подкреплением.

Первой проблемой является обоснованная настройка параметров обучения с подкреплением, включая коэффициент обучения, вознаграждения, коэффициент обесценивания [9, 22]. Обоснованный выбор указанных параметров существует только для частных случаев, которые описывают отдельные режимы и относительно простые задачи. В более сложных случаях, в которых применяются нейронные сети, дополнительно требуется разрабатывать рациональную архитектуру нейронной сети, выбрать ее настройки и гиперпараметры обучения.

Вторая проблема заключается в том, что основным инструментом обучения с подкреплением является виртуальная среда моделирования, требующая использования модели управляемого объекта [9, 22]. Иными словами, объект рассматривается в виде «черного ящика» в замкнутой системе управления, однако обучение на реальном объекте связано с рисками, обусловленными подачей на вход не соответствующих текущему состоянию управляющих воздействий. В этой связи на этапе обучения с применением виртуальной среды необходима адекватная модель объекта управления. В ней возможно варьирование параметров, структуры и возмущений, однако перенос на реальный объект, как правило требует дообучения. Такой подход не всегда возможен в силу требований безопасности и надежности.

Третья проблема заключается в необходимости обучать алгоритм управления при различных сочетаниях начальных значений переменных состояния, задающего и возмущающих воздействий, параметрических и структурных возмущений. Указанная необходимость приводит к значительному числу итераций обучения для достижения приемлемого качества управления. В свою очередь, большое число итераций обучения обуславливает необходимость применения специализированных вычислительных средств, что ограничивает использование технологий глубокого обучения с подкреплением в инженерной практике.

В настоящей статье разрабатывается и исследуется методика синтеза адаптивного управления динамическим объектом на примере двигателя, базирующаяся на методе обучения с подкреплением. Предлагается и исследуется базовая методика, в рамках которой рассматривается модель двигателя с неизвестными постоянными параметрами, которые не изменяются в ходе обучения и при функционировании системы управления. На основе исследования формулируются условия выбора параметров обучения. Далее, на основе полученных результатов, разрабатывается методика решения задачи управления более общего вида, когда параметры двигателя неизвестны, постоянны, но варьируются в ходе обучения и в замкнутой системе управления. Проводится анализ полученной методики и даются рекомендации по выбору параметров обучения.

I. Постановка задачи. Рассмотрим систему с интеллектуальным агентом, представленную на рис. 1. Система состоит из двигателя Д и интеллектуального агента А. Выходной измеряемой переменной является частота вращения двигателя ω. Входным воздействием является напряжение *u*, подаваемое на двигатель. Агенту также задается желаемая частота вращения двигателя ω0.

В качестве объекта управления рассматривается модель двигателя постоянного тока с независимым возбуждением, которая описывается уравнениями [25]:

$$\begin{aligned} J\frac{d\omega(t)}{dt} &= k_m I - M_l, \\ L\frac{dI(t)}{dt} &= \mathrm{u} - k_m\omega - rI, \end{aligned} \tag{1}$$

где ω – частота вращения вала двигателя, находящееся в интервале [ω*min*, ω*max*]; *I* – ток статорной обмотки; *u* – напряжение, подаваемое на статор, находящееся в интервале [*umin*, *umax*]; M_l – момент нагрузки; *J*, *L*, *r*, k_m – постоянные параметры.

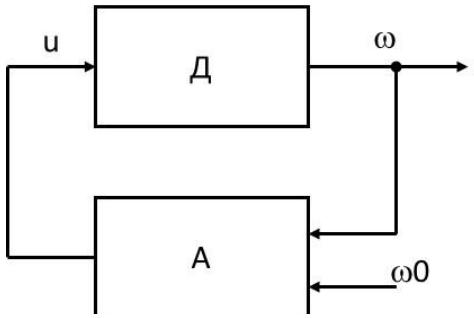


Рис. 1. Структура системы управления с интеллектуальным агентом.

Требуется разработать методику обучения с подкреплением агента А, которая должна включать обоснование алгоритма и параметров обучения, в том числе выбор подаваемых управляющих воздействий из множества $u \in \{ui\}$, *i*=1,2,…,*Nu*, вознаграждений *R* из множества $R \in \{ Rj\}$, *j*=1,2,…,*NR*,, шага обучения α, коэффициента обесценивания γ.

II. Решение задачи. Вначале будем полагать, что параметры двигателя *J*, *L*, *r*, k_m и момент нагрузки M_l являются неизвестными, но постоянными на всем протяжении обучения и дальнейшего функционирования замкнутой системы управления.

Рассмотрим вариант, при котором на вход двигателя Д подается управляющее воздействие из следующего множества $u \in \{-umax, 0, +umax\}$. Вознаграждение R на шаге будем вычислять в соответствии с выражением [9]:

$$R = -1, if\ \omega \neq \omega 0,$$
$$R = 0,\ if\ \omega = \omega 0. \quad (2)$$

Данная постановка относится к классу задач, решаемых классическими методами обучения с подкреплением [9]. Для ее решения применим подход, заключающийся в построении дискретных оценок функций ценности действий. Для этого разобьем интервал [ω*min*, ω*max*] на *N*ω дискретных интервалов. Полагая, что желаемые значения частота вращения двигателя ω0 задаются в том же диапазоне [ω*min*, ω*max*], разобьем диапазон значений ω0 также на *N*ω дискретных интервалов. Тогда наблюдаемое состояние системы в момент времени t можно описать вектором $I - St = [\omega k,\ \omega 0k]^T$. В каждом состоянии St агент может предпринять действие $A_t \in \{-umax, 0, +umax\}$. Необходимо найти алгоритм, обеспечивающий выбор действия с максимальной ценностью в каждом состоянии.

Известно, что методы на основе временных различий сходятся быстрее, чем методы Монте-Карло [9, 26]. В этой связи будем использовать метод Q-обучения нулевого порядка [9], который относится к методам обучения с разделенной стратегией и описывается выражением:

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1}, u) - Q(S_t, A_t)], \quad (3)$$

где $Q(S_t, A_t)$ – ценность действия A_t в состоянии S_t; R_{t+1} – вознаграждение на шаге t+1; α – шаг обучения; γ – коэффициент обучения.

Преимуществом метода Q-обучения является тот факт, что относясь к методам с разделенной стратегий, он допускает использование ε-жадного алгоритма [27, 28] в процессе обучения и жадного алгоритма в процессе функционирования обученного агента.

Целью обучения является максимизация вознаграждения, определяемого выражением [9]:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \quad (4)$$

Для оценки качества функционирования замкнутой системы управления будем использовать среднеквадратическое отклонение (СКО) частоты вращения ω от желаемого значения ω0 в установившемся режиме.

Проведем приближенное исследование Q-алгоритма обучения на устойчивость. Для этого представим уравнение (1) в дискретной форме, используя численный метод интегрирования прямоугольниками. В результате получим:

$$\omega_{k+1} = \omega_k + \frac{dt}{J}(k_m I_k - M_{lk}),$$
$$I_{k+1} = I_k + \frac{dt}{L}(u_k - k_m \omega_k - r I_k), \quad (5)$$

где dt – шаг интегрирования; ω_k, I_k, u_k, M_{lk} – значения переменных состояния и воздействий в момент времени k.

При исследовании устойчивости будем полагать задающее воздействие ω0 = 0 и момент сопротивления $M_l = 0$.

Введем в рассмотрение следующую положительно определенную функцию:

$$V_k = \omega_k^2 + Q_k^2(\omega_k, u_k). \quad (6)$$

где $Q_k\ (\omega_k, u_k)$ – оценка ценности действия u_k в состоянии ω_k.

Рассмотрим момент времени k+1 и разность

$$\Delta V_k = V_{k+1} - V_k = \omega_{k+1}^2 + Q_{k+1}^2(\omega_{k+1}, u_{k+1}) - \omega_k^2 - Q_k^2(\omega_k, u_k). \quad (7)$$

Подставим в (7) выражения (3), (5):

$$\Delta V_k = \left(\omega_k + \frac{dt}{J} k_m I_k\right)^2 + \left[Q_k + \alpha\left(R_{k+1} + \max_u Q'_k - Q_k\right)\right]^2 - \omega_k^2 - Q_k^2, \quad (8)$$

где $Q'_k = Q(\omega_{k+1}, u)$. Также для краткости введено обозначение $Q_k = Q_k\ (\omega_k, u_k)$.

Раскроем в выражении (8) скобки и сократим подобные:

$$\Delta V_k = 2\frac{dt}{J}k_m\omega_k I_k + \left(\frac{dt}{J}k_m I_k\right)^2 + 2\alpha Q_k\left(R_{k+1} + \max_u Q'_k - Q_k\right) + \left[\alpha\left(R_{k+1} + \max_u Q'_k - Q_k\right)\right]^2. \quad (9)$$

В силу того, что нулевое положение равновесия двигателя постоянного тока (1) является асимптотически устойчивым, и в силу ограниченных значений управления u переменные ω_k, I_k также ограничены. Тогда при достаточно малом шаге интегрирования $dt \to 0$, первые 2 слагаемый выражения (9) можно отбросить в силу их малости. В результате получим выражением

$$\Delta V_k = 2\alpha Q_k\left(R_{k+1} + \max_u Q'_k - Q_k\right) + \left[\alpha\left(R_{k+1} + \max_u Q'_k - Q_k\right)\right]^2. \quad (10)$$

Преобразуем последнее выражение:

$$\Delta V_k = \alpha\left(R_{k+1} + \max_u Q'_k - Q_k\right)\left(2Q_k + \alpha\left(R_{k+1} + \max_u Q'_k - Q_k\right)\right). \quad (11)$$

Учитывая, что вознаграждения $R_{k+1} \le 0$, значения полезности также отрицательны $Q_k \le 0$. Тогда можно записать

$$R_{k+1} + \max_u Q'_k - Q_k \le R_{k+1} - Q_k \le -Q_k. \quad (12)$$

Учитывая неравенство (12) получим

$$\Delta V_k \le -\alpha(2-\alpha)Q_k^2. \quad (13)$$

Согласно выражению (13), разность ΔV_k (7) является не положительно определенной при выполнении условий $0 < \alpha < 2$, что соответствует известным условиях сходимости алгоритмов обучения [19].

Таким образом, доказано следующее утверждение.

Утверждение. Алгоритм обучения (3) с вознаграждением (2) и управлением $u=\max_u Q_k(\omega_k, u_k)$, $u \in \{-umax, 0, +umax\}$ обеспечивает асимптотическую сходимость функции полезности $Q_k(\omega_k, u_k)$ к постоянным значениям и ограниченность частоты вращения двигателя ω_k, описываемой дискретной моделью (5).

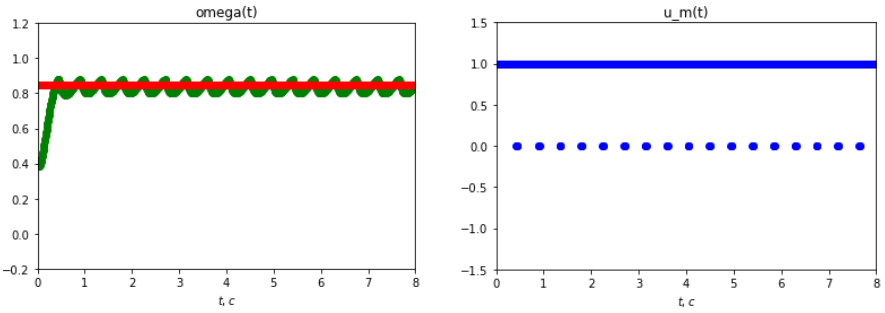
Для численного исследования зависимости СКО от параметров обучения будем осуществлять обучение агента и моделирование системы, представленной на рис. 1, с обученным агентом при различных шагах обучения и коэффициенте обучения. В табл. 1 представлены результаты такого исследования при вознаграждении, определяемом выражением (2), жадном алгоритме обучения (ε=0), числе эпизодов 80 000, 8 000 шагов обучения в каждом эпизоде. Параметры двигателя равны: J=0.2; k=1.0; Ml=0.1; L=0.1; r=1.0. Значение *umax* = 1, а частота вращения ограничена интервалом [0, 1].

Таблица 1

Зависимость СКО от шага обучения и коэффициента обесценивания для $u \in \{-umax, 0, +umax\}$

α \ γ	1.0	0.98	0.94	0.9
0.03	0.063	0.063	0.061	0.05
0.05	0.063	0.063	0.057	0.069
0.07	0.031	0.057	0.041	0.55
0.1	0.045	0.058	0.56	0.058

На рис. 2 представлен пример переходных процессов при функционировании обученного интеллектуального Q-агента. На рис. 2,а хорошо заметно смещение среднего значения выходной величины относительно уставки. Это смещение обусловлено использованием функции максимума в выражении (3), которое приводит к смещению получаемых оценок функции полезности [9]. Для решения этой проблемы можно использовать двойное Q-обучение – DQ метод [9, 29].



а – Частота ω и уставка ω0 б – управление *u*

Рис. 2. Переходные процессы в замкнутой системе (1) – (3) при u∈{-umax, 0, + umax}

Наличие колебаний на рис. 2 ожидаемо, они обусловлены релейным характером применяемого управления. Как видно из табл. 1 коэффициент обесценивания может быть принят равным γ=1. Относительно шага обучения α наблюдается минимум в районе значения α=0.07.

Исследуем теперь возможности по уменьшению ошибки управления. Вначале рассмотрим, как влияет введение увеличение числа возможных управлений в множестве *u*. Зависимость СКО от шага обучения и коэффициента обесценивания для *u*∈{-*umax*, -*umax*/2, 0, +*umax*/2, + *umax*} представлена в табл. 2.

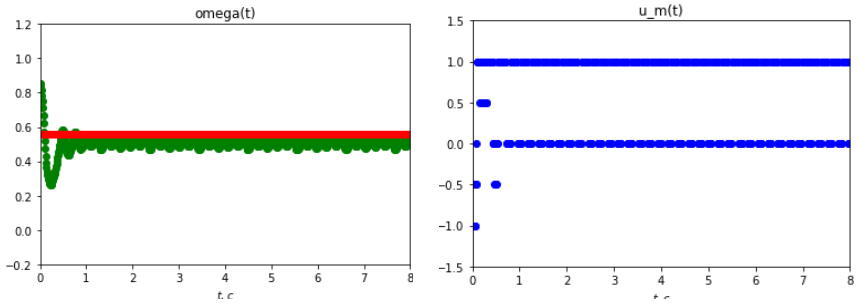
Таблица 2

Зависимость СКО от шага обучения и коэффициента обесценивания для *u*∈{-*umax*, -*umax*/2, 0, +*umax*/2, + *umax*}

γ α	1.0	0.98	0.94	0.9
0.03	0.061	0.267	0.092	0.132
0.05	0.046	0.04	0.109	0.69
0.07	0.056	0.131	0.707	0.698
0.1	0.046	0.097	0.703	0.696

Сравнение средней ошибки для различного сочетания параметров α и γ показывает, что для α={0.03, 0.05, 0.07,0.1} и γ={1.0, 0.98, 0.94, 0.9} среднее значение СКО для *u*∈{-*umax*, 0, + *umax*} составляет 0.118. Среднее значение СКО для *u*∈{-*umax*, -*umax*/2, 0, +*umax*/2, + *umax*} составляет 0.286.

На рис. 3 представлены графики изменения частоты вращения двигателя и управляющего воздействия при α=0.1 и γ=1.0. Результаты численных исследований показывают, что введение дополнительный ступеней управления позволяют снизить амплитуду колебаний частоты вращения двигателя ω относительно заданного желаемого значения ω0. Однако, при это возрастает постоянное смещение относительно желаемого значения ω0.



а – Частота ω и уставка ω0 б – управление *u*

Рис. 3. Переходные процессы при u∈{-umax, -umax/2, 0, +umax/2, + umax}

Из табл. 1 и 2 следует, что одновременное увеличение шага обучения α и уменьшение коэффициента обесценивания γ приводит к росту СКО. Это обусловлено тем, что с ростом α и уменьшением γ увеличивается число эпизодов, в которых управление, выбираемое в соответствии с максимумом

$$u_k = \max_u Q_k\ (\omega_k, \mathrm{u}_k). \quad (14)$$

не приводит к успешной стабилизации частоты вращения двигателя ω в окрестности значения ω0. В связи с этим, далее при исследовании будем выбирать небольшие значения α≤0.07 параметр γ, равный или близкий к 1.

Отметим, что при малых значениях α и значениях γ, близких к 1, наблюдается стабилизация частоты вращения двигателя в окрестности ω0. При этом, СКО управления примерно равна (1–1.5) от величины дискретизации непрерывного пространства (ω, ω0). Для результатов, представленных в табл. 1 и 2, размер ячейки равен 0.05×0.05.

Дальнейшее введение дополнительных ступеней в управляющие воздействия может приводить к уменьшению амплитуды колебаний выходной величины двигателя, однако не устранит эти колебания полностью. Кроме того, известно [9], что в дискретной форме Q-метод хорошо работает для относительно небольшого множества возможных управляющих воздействий. В противном случае необходимо переходить к задаче непрерывного управления, что требует применения методов аппроксимации функций ценности, в том числе с использованием нейронных сетей.

Основываясь на проведенном исследовании, рассмотрим три важных аспекта применения метода Q-обучения для управления динамическими объектами. Первым аспектом является способ назначения вознаграждения, вторым – способ формирования управления, который не требует рассмотрения непрерывной задачи управления, а третьим – размер ячейки при дискретизации непрерывного пространства.

III. Исследование способов дискретизации, назначения вознаграждения и формирования управления. Рассмотрим вначале влияние размера ячейки дискретизации пространства (ω, ω0), так как этот параметр наиболее сильно сказывается на вычислительных затратах на проведение исследований. При разбиении среды на 30×30 ячеек (размеры ячейки 0.333×0.333) для обучения Q-алгоритма потребовалось 240 000 эпизодов, что оказалось приемлемым с точки зрения времени, затрачиваемого на обучение алгоритма управления.

В табл. 3 представлены результаты сравнения погрешности управления для рассматриваемой задачи при дискретизации 30×30 ячеек и 20×20 ячеек, $u \in \{-umax, -umax/2, 0, +umax/2, +umax\}$, γ=1.

Таблица 3

Сравнение СКО при дискретизации 30×30 ячеек и 20×20 ячеек

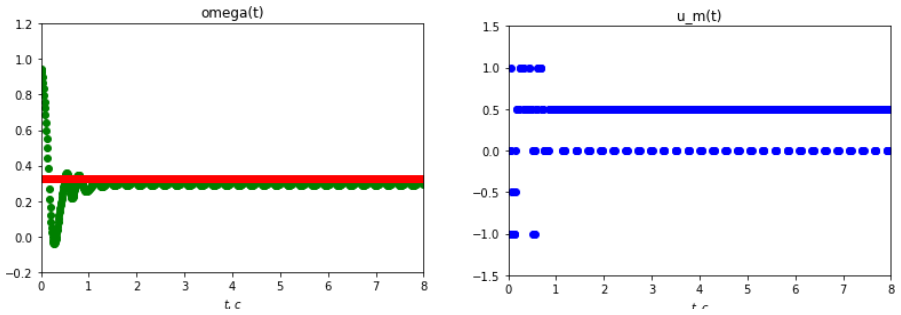
α	0.03	0.05	0.07
СКО 20×20	0.061	0.046	0.056
СКО 30×30	0.039	0.044	0.028

На рис. 4 представлены переходные процессы в замкнутой системе в одном из эпизодов при α=0.03.

Из табл. 3 следует, что среднее значение СКО при дискретизации 30×30 ячеек равно 0.037, что равно 1,12 от размера ячейки. Для дискретизации 20×20 ячеек среднее значение СКО равно 0.054, что приблизительно составляет 1,09 от размера ячейки. Таким образом, уменьшение размера ячейки привело к почти пропорциональному уменьшению ошибки управления.

С точки зрения обучения с подкреплением, особенность рассматриваемой задачи заключается в том, что в данном случае известна ошибка управления, что дает возможность сформировать на ее основе вознаграждение. Рассмотрим вариант, в которой вознаграждение на отдельном шаге равно взятому с минусом квадрату ошибки управления:

$$R_k = -(\omega 0 - \omega)^2. \quad (15)$$



а – Частота ω и уставка ω0 б – управление *u*

Рис. 4. Переходные процессы при u∈{-umax, -umax/2, 0, +umax/2, + umax} и дискретизации 30×30 ячеек

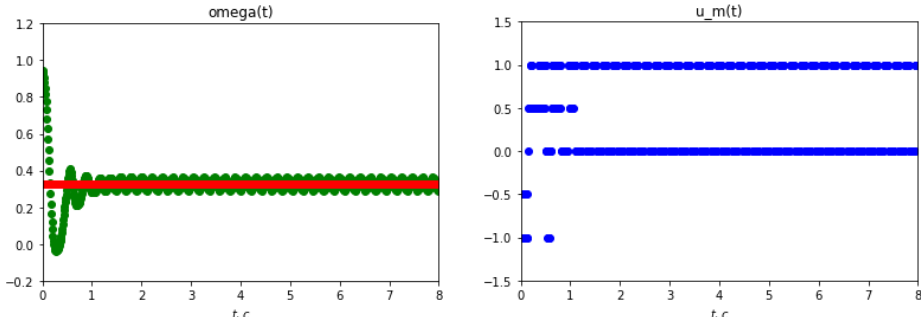
В табл. 4 представлена СКО замкнутой системы управления, обученной Q-методом с вознаграждением (15) при дискретизации 30×30 ячеек, *u*∈{*-umax, -umax*/2, 0, *+umax*/2, *+ umax*}, γ=1.

Таблица 4

СКО при дискретизации 30×30 ячеек *u*∈{*-umax, -umax*/2, 0, *+umax*/2, *+ umax*}, γ=1 и вознаграждении (15)

α	0.03	0.05	0.07
СКО	0.028	0.025	0.023

Согласно табл. 4 среднее СКО равно около 0.025, что составляет 0.76 от размера ячейки дискретизации пространства. Также отметим, что при вознаграждении (15) существенно снижается смещение СКО ошибки управления относительно 0. На рис. 5 представлены результаты моделирования замкнутой системы, обученной Q-методом с вознаграждением (15), из которых видно отсутствие указанного смещения.



а – Частота ω и уставка ω0 б – управление *u*

Рис. 5. Переходные процессы при u∈{-umax, -umax/2, 0, +umax/2, + umax} и дискретизации 30×30 ячеек γ=1 и вознаграждении (15)

Теперь рассмотрим инкрементальный способ введения управления, при котором вычисляется не абсолютное значение управления, а его приращение. Будем рассматривать множество возможных управлений, включающих 3 действия: оставить управление прежним, увеличить управление на величину *+du*, уменьшить управление на величину *- du*.

Результаты моделирования показывают, что в данном случае СКО ошибки управления составляет порядка 0,18. Большая ошибка обусловлена тем, что для данного способа формирования управления требуется строить функцию ценности действий в пространстве $(\omega, \omega 0, u)$, т.е. учитывать текущее значение управляющего воздействия.

IV. Методика разработки интеллектуальной системы управления на основе Q-метода обучения с подкреплением. На основе проведенных исследований предлагается следующая методика.

1. Исходными данными являются диапазон изменения частот вращения [ω*min*, ω*max*], диапазон изменения входных воздействий [*umin*, *umax*], требуемая точность поддержания частоты вращения *e*ω. Номер итерации устанавливается Nit=1.

2. Непрерывное пространство (ω, ω0) разбивается на ячейки, число которых определяется выражением

$$N_c = \left\lceil \frac{\omega max - \omega min,}{e\omega} \right\rceil, \quad (16)$$

где ⌈∗⌉ – операция округления к ближайшему сверху целому числу.

3. Выбирается множество управлений *u*∈{-*umax*, 0, + *umax*}.

4. Осуществляется обучение Q-методом (3) алгоритма управления (14) при γ=1 и $\alpha \in (0, 0.1]$, при вознаграждении (15).

5. Проводится исследование точности замкнутой системы.

6. Если полученное СКО меньше или равно величины *e*ω, то процесс завершается (переход на п. 10).

7. Если полученное СКО больше величины *e*ω, то в множество управлений *u* осуществляется введение дополнительных управлений величиной *umax*/(2Nit).

8. Повторяются пункты 4, 5, 6.

9. Если полученное СКО все еще больше величины *e*ω, то увеличивается число ячеек, на которые разбивается пространство (ω, ω0) и повторяются пункты 4, 5, 6.

10. Конец процедуры.

Рассмотренная методика позволяет обеспечить требуемую точность поддержания заданной частоты вращения двигателя, однако существуют колебательные движения в малой окрестности желаемой частоты из-за релейного характера управления. Устранение колебаний требует решения задачи в классе непрерывных управлений, что, как указывалось выше требует непрерывной аппроксимации функций ценности управляющих воздействий. Как правило, такая задача требует применения нейронных сетей в качестве аппроксиматоров функций ценности.

Необходимость использования интеллектуальных методов аппроксимации функций ценности в описываемой задаче обусловлена сложностью этих функций. Для примера, на рис. 6 представлен функция ценности действия *u*=-*umax*.

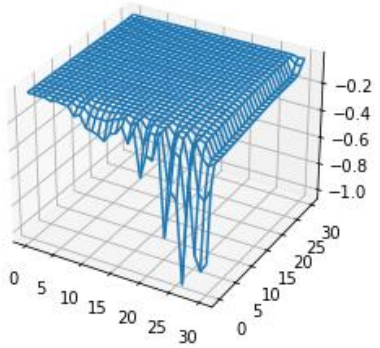


Рис. 6. Вид функции ценности Q(ω, ω0, u) при u=-umax

На рис. 6 функция ценности приведена к диапазону [0 1]. Для функций ценности в решаемой задаче характерны высокая нелинейность и большой диапазон значений, что обуславливает необходимость применения нейросетевых аппроксиматоров. В табл. 5 представлены результаты сравнения полиномиальной и нейросетевой аппроксимации функции ценности. Полиномиальная аппроксимация выполнена для порядка полинома n=3. Нейро-

сетевая аппроксимация выполнена трехслойной статической полносвязной нейронной сетью прямого распространения с числом нейронов 40, 40 и 1. Функции активации первого и второго слоев сигмоидальные, а функция активации выходного слоя – линейная. Из табл. 5 следует, что погрешность нейросетевой аппроксимации функции ценности действий почти на порядок меньше, чем погрешность полиномиальной аппроксимации.

Таблица 5

СКО при полиномиальной и нейросетевой аппроксимации

Метод аппроксимации	Полиномиальный	Нейросетевой
Ошибка	0.18	0.02

Обсуждение результатов и заключение. В работе исследован процесс формирования управления методом Q-обучения с подкреплением для динамического устойчивого объекта с одним входом и одним выходом. В качестве объекта рассмотрен двигатель постоянного тока с независимым возбуждением.

Получено доказательство сходимости процесса обучения и оценка шага обучения. Данная оценка является грубой и завышает допустимый шаг обучения α, что связано с двумя факторами. Первый из них заключается в том, что в приведенном доказательстве не учтена дискретизации пространства состояния по переменным $(\omega, \omega 0)$. Второй фактор – предположение о том, что шаг дискретизации по времени бесконечно мал. Влияние последнего предположения всегда можно снизить, выбрав достаточно малый шаг дискретизации по времени. Первой предположение требует рассмотрения гибридной дискретно-непрерывной системы.

Однако, принципиальным ограничением предложенного метода является необходимость наличия модели на этапе обучения алгоритма управления. Такая модель должна позволять обучать алгоритм управления во всех возможных режимах, при всех возможных параметрах и возмущениях. Вместе с тем, известно, что обучение на синтетических данных не позволяет достичь высокой точности, поэтому требуется проведение дообучения на реальном объекте. Очевидно, неустойчивость какого-либо состояния объекта управления будет при этом проблемой.

К достоинствам предложенного метода управления относится его простота и возможность автоматической настройки для объекта с высокой степенью неопределенности параметров и внешних возмущений. При этом настройки осуществляется до начала функционирования замкнутой системы и может занимать до нескольких минут. Аналогичный подход используется в ряде приводов Siemens для настройки ПИД-регуляторов [30].

Основным недостатком предложенного метода является то, что на выходе системы присутствуют неустранимые колебания, то не всегда допустимо с точки зрения управляемого технологического процесса. Для устранения данного недостатка необходимо переходить к постановке задачи с непрерывным управлением. В свою очередь, такая постановка задачи требует аппроксимации функций ценности и соответственно применения нейросетевых аппроксиматоров.

Для придания свойств адаптивности в динамике необходимо проводить обучение в пространстве состояний, включающем ток и текущее значение управляющего воздействия $(\omega, \omega 0, I, u)$. В этом случае можно обеспечить адаптацию к неизвестным изменяющимся в ходе функционирования объекта параметров и внешних возмущений.

Исследование выполнено за счет гранта Российского научного фонда № 24-19-00063, «Теоретические основы и методы группового управления безэкипажными подводными аппаратами», https://rscf.ru/project/24-19-00063/ на базе ФГАОУ ВО «Южный федеральный университет».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Фаворская М.Н., Пахирка А.И.* Восстановление аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей // Информатика и автоматизация. – 2024. – Т. 23 (4). – С. 1047-1076. – DOI: 10.15622/ia.23.4.5.

2. *Голубинский А.Н., Толстых А.А., Толстых М.Ю.* Автоматическая генерация аннотаций научных статей на основе больших языковых моделей // Информатика и автоматизация. – 2025. – Т. 24 (1). – С. 275-301. – https://doi.org/10.15622/ia.24.1.10.
3. *Zhang D., He R., Liao X., Li F., Chen J. and Yang G*. Face Forgery Detection Based on Fine-Grained Clues and Noise Inconsistency // IEEE Transactions on Artificial Intelligence. – 2025. – Vol. 6 (1). – P. 144-158. – DOI: 10.1109/TAI.2024.3455311.
4. *Sobo A., Mubarak A., Baimagambetov A., Polatidis N.* Evaluating LLMs for Code Generation in HRI: A Comparative Study of ChatGPT, Gemini, and Claude // Applied Artificial Intelligence. – 2024. – Vol. 39 (1). – https://doi.org/10.1080/08839514.2024.2439610.
5. *Чен Х., Игнатьева С.А., Богуш Р.П., Абламейко С.В.* Повторная идентификация людей в системах видеонаблюдения с использованием глубокого обучения: анализ существующих методов // Автоматика и телемеханика. – 2023. – № 5. – С. 61-112. – DOI: 10.31857/S0005231023050057
6. *Понимаш З.А., Потанин М.В.* Метод и алгоритм извлечения признаков из цифровых сигналов на базе нейросетей трансформер // Известия ЮФУ. Технические науки. – 2024. – № 6. – С. 52-64. – DOI: 10.18522/2311-3103-2024-6-52-64.
7. *Hamdan N., Medvedev M., Pshikhopov V.* Method of Motion Path Planning Based on a Deep Neural Network with Vector Input // Mekhatronika, Avtomatizatsiya, Upravlenie. – 2024. – Vol. 25(11). – P. 559-567. – https://doi.org/10.17587/mau.25.559-567.
8. *Gaiduk A.R., Martjanov O.V., Medvedev M.Yu., PshikhopovV.Kh., Hamdan N., Farhood A.* Neural network based control system for robots group operating in 2-d uncertain environment // Mekhatronika, Avtomatizatsiya, Upravlenie. – 2020. – Vol. 21 (8). – P. 470-479. – https://doi.org/10.17587/mau.21.470-479.
9. *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.
10. *Lei X., Zhang Z., Dong P*. Dynamic Path Planning of Unknown Environment Based on Deep Reinforcement Learning // Journal of Robotics. – V. 2018, Article ID 5781591. – 10 p. – https://doi.org/10.1155/2018/5781591.
11. *Wang B., Liu Z., Li Q., Prorok A.* Mobile Robot Path Planning in Dynamic Environments Through Globally Guided Reinforcement Learning // IEEE Robotics and Automation Letters. – 2020. – Vol. 5 (4). – P. 6932-6939. – DOI: 10.1109/LRA.2020.3026638.
12. *Srikonda S., Norris W.R., Nottage D., Soylemezoglu A.* Deep Reinforcement Learning for Autonomous Dynamic Skid Steer Vehicle Trajectory Tracking // Robotics. – 2022. – Vol. 11 (95). – https://doi.org/10.3390/robotics11050095.
13. *Van Hasselt H., Guez A. Silver D.* Deep Reinforcement Learning with Double Q-Learning[C] // Proceedings of the AAAI Conference on Artificial Intelligence. – 2016. – 30 (1).
14. *Lv L.H., Zhang S.J., Ding D.R., Wang Y.X.* Path Planning via an Improved DQN-Based Learning Policy // IEEE Access. – 2019. – Vol. 7. – P. 67319-67330.
15. *Ramaswamy A. and Hüllermeier E.* Deep Q-Learning: Theoretical Insights From an Asymptotic Analysis // IEEE Transactions on Artificial Intelligence. – 2022. – Vol. 3 (2). – P. 139-151. – DOI: 10.1109/TAI.2021.3111142
16. *Lillicrap T.P., Hunt J.J., Pritzel A., Heess N., Erez T., Tassa Y., Silver D., Wierstra D.* Continuous control with deep reinforcement learning // arXiv:1509.02971. – 2015.
17. *Fujimoto S., Hoof H.V., Meger D.* Addressing Function Approximation Error in Actor-Critic Methods // ArXiv. – 2018. – abs/1802.09477.
18. *Хапкин Д.Л., Феофилов С.В.* Синтез устойчивых нейросетевых регуляторов для объектов с ограничителями в условиях неполной информации // Мехатроника, автоматизация, управление. – 2024. – Т. 25 (7). – С. 345-353. – https://doi.org/10.17587/mau.25.345-353.
19. *Gupta M., Jin L., Homma N.* Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory. – John Wiley & Sons, Hoboken, New Jersey, 2003.
20. *Жилов Р.А.* Постройка ПИД-регулятора с использованием нейронных сетей // Известия Кабардино-Балкарского научного центра РАН. – 2022. – № 5 (109). – С. 38-47. – DOI: 10.35330/1991-6639-2022-5-109-38-47.
21. *Nguyet T.M., Dang X.B.* A neural flexible PID controller for task-space control of robotic manipulators // Frontiers in Robotics and AI. – 2023. – Vol. 9. – P. 1-10. – DOI=10.3389/frobt.2022.975850.
22. *Mnih V., Kavukcuoglu K., Silver D. et al.* Human-level control through deep reinforcement learning // Nature. – 2015. – No. 518 (7540). – P. 529-533.
23. *Пшихопов В.Х., Медведев М.Ю.* Блочный синтез робастных систем при ограничениях на управления и координаты состояния // Мехатроника, автоматизация и управление. – 2011. – № 1. – С. 2-8.
24. *Pshikhopov V., Medvedev M.* Multi-Loop Adaptive Control of Mobile Objects in Solving Trajectory Tracking Tasks // Automation and Remote Control. – 2020. – Vol. 81 (11). – P. 2078-2093.
25. – https://doi.org/10.1134/S0005117920110090.

26. *Пшихопов В.Х., Медведев М.Ю., Шевченко В.А.* Адаптивное управление с эталонной моделью приводом постоянного тока // Известия ЮФУ. Технические науки. – 2015. – № 2 (163). – С. 6-18.
27. *Еремеев А.П., Сергеев М.Д., Петров В.С.* Интеграция методов обучения с подкреплением и нечеткой логики для интеллектуальных систем реального времени // Программные продукты и системы. – 2023. – Т. 36 (4). – С. 600-606. – DOI: 10.15827/0236-235X.142.600-606.
28. *Takuya Okano, Masaki Onishi.* A Parameter Investigation of the ϵ-greedy Exploration Ratio Adaptation Method in Multi-agent Reinforcement Learning // European Workshop on Reinforcement Learning. – 2018. – Vol. 14.
29. *Shah Asif Bashir1, Farida Khursheed, Ibrahim Abdoulahi.* Adaptive ε-Greedy Exploration for Finite Systems // Gedrag & Organisatie Reniew. – 2021. – Vol. 34 (04). – DOI: 10.37896/GOR34.04/044.
30. *Van Hasselt H.* Double Q-learning // Advances in Neural Information Processing Systems. – 2011. – Vol. 23. – P. 2613-2622.
31. *Бергер Г.* Автоматизация посредством STEP 7 с использованием STL и SCL программируемых контроллеров SIMATIC S7-300/400. – Siemens AG, Нюрнберг, 2001.

REFERENCES

1. *Favorskaya M.N., Pakhirka A.I.* Vosstanovlenie aerofotosnimkov sverkhvysokogo razresheniya s uchetom semanticheskikh osobennostey [Restoration of ultra-high-resolution aerial photographs taking into account semantic features], *Informatika i avtomatizatsiya* [Computer Science and Automation], 2024, Vol. 23 (4), pp. 1047-1076. DOI: 10.15622/ia.23.4.5.
2. *Golubinskiy A.N., Tolstykh A.A., Tolstykh M.Yu.* Avtomaticheskaya generatsiya annotatsiy nauchnykh statey na osnove bol'shikh yazykovykh modeley [Automatic generation of abstracts of scientific articles based on large language models], *Informatika i avtomatizatsiya* [Computer Science and Automation], 2025, Vol. 24 (1), pp. 275-301. Available at: https://doi.org/10.15622/ia.24.1.10.
3. *Zhang D., He R., Liao X., Li F., Chen J. and Yang G*. Face Forgery Detection Based on Fine-Grained Clues and Noise Inconsistency, *IEEE Transactions on Artificial Intelligence,* 2025, Vol. 6 (1), pp. 144-158. DOI: 10.1109/TAI.2024.3455311.
4. *Sobo A., Mubarak A., Baimagambetov A., Polatidis N.* Evaluating LLMs for Code Generation in HRI: A Comparative Study of ChatGPT, Gemini, and Claude, *Applied Artificial Intelligence,* 2024, Vol. 39 (1). Available at: https://doi.org/10.1080/08839514.2024.2439610.
5. *Chen Kh., Ignat'eva S.A., Bogush R.P., Ablameyko S.V.* Povtornaya identifikatsiya lyudey v sistemakh videonablyudeniya s ispol'zovaniem glubokogo obucheniya: analiz sushchestvuyushchikh metodov [People re-identification in video surveillance systems using deep learning: analysis of existing methods], *Avtomatika i telemekhanika* [Automation and Telemetry], 2023, No. 5, pp. 61-112. DOI: 10.31857/S0005231023050057
6. *Ponimash Z.A., Potanin M.V.* Metod i algoritm izvlecheniya priznakov iz tsifrovykh signalov na baze neyrosetey transformer [Method and algorithm for extracting features from digital signals based on transformer neural networks], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2024, No. 6, pp. 52-64. DOI: 10.18522/2311-3103-2024-6-52-64.
7. *Hamdan N., Medvedev M., Pshikhopov V.* Method of Motion Path Planning Based on a Deep Neural Network with Vector Input, *Mekhatronika, Avtomatizatsiya, Upravlenie,* 2024, Vol. 25(11), pp. 559-567. Available at: https://doi.org/10.17587/mau.25.559-567.
8. *Gaiduk A.R., Martjanov O.V., Medvedev M.Yu., PshikhopovV.Kh., Hamdan N., Farhood A.* Neural network based control system for robots group operating in 2-d uncertain environment, *Mekhatronika, Avtomatizatsiya, Upravlenie,* 2020, Vol. 21 (8), pp. 470-479. Available at: https://doi.org/10.17587/mau.21.470-479.
9. *Satton R.S., Barto E.Dzh.* Obuchenie s podkrepleniem: Vvedenie [Reinforcement learning: Introduction]. 2nd ed.: transl. from Engl. A.A. Slinkina. Moscow: DMK Press, 2020, 552 p.
10. *Lei X., Zhang Z., Dong P.* Dynamic Path Planning of Unknown Environment Based on Deep Reinforcement Learning, *Journal of Robotics,* V. 2018, Article ID 5781591, 10 p. Available at: https://doi.org/10.1155/2018/5781591.
11. *Wang B., Liu Z., Li Q., Prorok A.* Mobile Robot Path Planning in Dynamic Environments Through Globally Guided Reinforcement Learning, *IEEE Robotics and Automation Letters,* 2020, Vol. 5 (4), pp. 6932-6939. DOI: 10.1109/LRA.2020.3026638.
12. *Srikonda S., Norris W.R., Nottage D., Soylemezoglu A.* Deep Reinforcement Learning for Autonomous Dynamic Skid Steer Vehicle Trajectory Tracking, Robotics, 2022, Vol. 11 (95). Available at: https://doi.org/10.3390/robotics11050095.
13. *Van Hasselt H., Guez A. Silver D.* Deep Reinforcement Learning with Double Q-Learning[C], *Proceedings of the AAAI Conference on Artificial Intelligence,* 2016, 30 (1).

14. *Lv L.H., Zhang S.J., Ding D.R., Wang Y.X.* Path Planning via an Improved DQN-Based Learning Policy, *IEEE Access,* 2019, Vol. 7, pp. 67319-67330.
15. *Ramaswamy A. and Hüllermeier E.* Deep Q-Learning: Theoretical Insights From an Asymptotic Analysis, *IEEE Transactions on Artificial Intelligence,* 2022, Vol. 3 (2), pp. 139-151. DOI: 10.1109/TAI.2021.3111142.
16. *Lillicrap T.P., Hunt J.J., Pritzel A., Heess N., Erez T., Tassa Y., Silver D., Wierstra D.* Continuous control with deep reinforcement learning, *arXiv:1509.02971,* 2015.
17. *Fujimoto S., Hoof H.V., Meger D.* Addressing Function Approximation Error in Actor-Critic Methods, *ArXiv,* 2018. abs/1802.09477.
18. *Khapkin D.L., Feofilov S.V.* Sintez ustoychivykh neyrosetevykh regulyatorov dlya ob"ektov s ogranichitelyami v usloviyakh nepolnoy informatsii [Synthesis of stable neural network controllers for objects with constraints under incomplete information], *Mekhatronika, avtomatizatsiya, upravlenie* [Mechatronics, Automation, Control], 2024, Vol. 25 (7), pp. 345-353. Available at: https://doi.org/10.17587/mau.25.345-353.
19. *Gupta M., Jin L., Homma N.* Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory. John Wiley & Sons, Hoboken, New Jersey, 2003.
20. *Zhilov R.A.* Postroyka PID-regulyatora s ispol'zovaniem neyronnykh setey [Construction of a PID controller using neural networks], *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [Bulletin of the Kabardino-Balkarian Scientific Center of the RAS], 2022, No. 5 (109), pp. 38-47. DOI: 10.35330/1991- 6639-2022-5-109-38-47.
21. *Nguyet T.M., Dang X.B.* A neural flexible PID controller for task-space control of robotic manipulators, *Frontiers in Robotics and AI,* 2023, Vol. 9. – P. 1-10. DOI=10.3389/frobt.2022.975850.
22. *Mnih V., Kavukcuoglu K., Silver D. et al.* Human-level control through deep reinforcement learning, *Nature,* 2015, No. 518 (7540), pp. 529-533.
23. *Pshikhopov V.Kh., Medvedev M.Yu.* Blochnyy sintez robastnykh sistem pri ogranicheniyakh na upravleniya i koordinaty sostoyaniya [Block synthesis of robust systems with constraints on controls and state coordinates], *Mekhatronika, avtomatizatsiya, upravlenie* [Mechatronics, Automation, Control], 2011, No. 1, pp. 2-8.
24. *Pshikhopov V., Medvedev M.* Multi-Loop Adaptive Control of Mobile Objects in Solving Trajectory Tracking Tasks, *Automation and Remote Control,* 2020, Vol. 81 (11), pp. 2078-2093. Available at: https://doi.org/10.1134/S0005117920110090.
25. *Pshikhopov V.Kh., Medvedev M.Yu., Shevchenko V.A.* Adaptivnoe upravlenie s etalonnoy model'yu privodom postoyannogo toka [Adaptive control with a reference model of a DC drive], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 2 (163), pp. 6-18.
26. *Eremeev A.P., Sergeev M.D., Petrov V.S.* Integratsiya metodov obucheniya s podkrepleniem i nechetkoy logiki dlya intellektual'nykh sistem real'nogo vremeni [Integration of reinforcement learning methods and fuzzy logic for intelligent real-time systems], *Programmnye produkty i sistemy* [Software products and systems], 2023, Vol. 36 (4), pp. 600-606. DOI: 10.15827/0236-235X.142.600-606.
27. *Takuya Okano, Masaki Onishi.* A Parameter Investigation of the ϵ-greedy Exploration Ratio Adaptation Method in Multi-agent Reinforcement Learning, *European Workshop on Reinforcement Learning,* 2018, Vol. 14.
28. *Shah Asif Bashir1, Farida Khursheed, Ibrahim Abdoulahi.* Adaptive ε-Greedy Exploration for Finite Systems, *Gedrag & Organisatie Reniew,* 2021, Vol. 34 (04). DOI: 10.37896/GOR34.04/044.
29. *Van Hasselt H.* Double Q-learning, *Advances in Neural Information Processing Systems,* 2011, Vol. 23, pp. 2613-2622.
30. *Berger G.* Avtomatizatsiya posredstvom STEP 7 s ispol'zovaniem STL i SCL programmiruemykh kontrollerov SIMATIC S7-300/400 [Automation with STEP 7 using STL and SCL of SIMATIC S7-300/400 programmable controllers]. Siemens AG, Nyurnberg, 2001.

Карапеев Александр Николаевич – ООО «Миртек»; e-mail: a.karapeev@mirtekgroup.ru; г. Таганрог, Россия; тел.: +79185564919; зам. директора по общим вопросам.

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

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

Пшихопов 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.

Karapeev Alexander Nikolaevich – Mirtek LLC; e-mail: a.karapeev@mirtekgroup.ru; Taganrog, Russia; phone: +79185564919; Deputy Director.

Kosenko Evgeniy Косенко Yurjevich – R&D Institute of Robotics and Control Systems; e-mail: ekosenko@sfedu.ru; Taganrog, Russia; phone: +78634371694; cand. of eng. sc.; senior researcher.

Medvedev Mikhail Yurjevich – R&D Institute of Robotics and Control Systems; e-mail: medvmihal@sfedu.ru; Taganrog, Russia; phone: +78634371694; dr. of eng. sc.; leading researcher.

Pshikhopov Viacheslav Khasanovich – R&D Institute of Robotics and Control Systems; e-mail: pshichop@rambler.ru; Taganrog, Russia; phone: +78634371694; dr. of eng. sc.; professor; director.

УДК 007.52 DOI 10.18522/2311-3103-2025-2-175-185

Л.А. Рыбак, Д.И. Малышев, Д.А. Дьяконов, А.А. Мамченкова

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ПЛАНИРОВАНИЯ ТРАЕКТОРИИ ДВИЖЕНИЯ ГРУППЫ МОБИЛЬНЫХ РОБОТОВ ПРИ НАЛИЧИИ СТАЦИОНАРНЫХ И ПОДВИЖНЫХ ПРЕПЯТСТВИЙ

Рассматривается метод планирования траектории движения группы мобильных роботов, обеспечивающий безопасное перемещение и исключающий возможность столкновений как между самими роботами, так и с внешними препятствиями, включая движущиеся объекты. Разработанная математическая модель учитывает три основных сценария возможных столкновений: пересечение траекторий роботов внутри группы, взаимодействие со стационарными препятствиями и вероятность столкновения с подвижными объектами. Каждый из этих сценариев детально анализируется для обеспечения максимальной безопасности движения, а их учет позволяет эффективно адаптировать маршруты роботов к изменяющимся условиям среды. Траектория движения каждого робота представляется в виде ломаной линии с промежуточными точками, которые оптимизируются для обеспечения безопасности движения. Особое внимание уделяется адаптации скорости на различных участках траектории: робот может изменять скорость в зависимости от текущих условий, чтобы минимизировать риск столкновений. Для оценки расстояний между объектами используется евклидова норма, позволяющая рассчитывать минимальные расстояния между центрами сферических представлений роботов и препятствий. Задача решается в два этапа. На первом этапе строится траектория для первого робота с учетом начальных условий и расположения препятствий. На втором этапе формируются траектории для остальных роботов с учетом уже спланированных маршрутов. Для оптимизации координат промежуточных точек и скоростей применяется генетический алгоритм, который минимизирует время перемещения и обеспечивает безопасность движения. Генетический алгоритм использует операторы скрещивания и мутации для создания разнообразных решений, а также выполняет проверку на соответствие условиям безопасности. Численное моделирование проведено на языке Python с использованием библиотеки Matplotlib для визуализации результатов. В ходе экспериментов было выполнено 50 тестов с различным количеством препятствий (от 5 до 10). Анализ результатов показал, что с увеличением числа препятствий возрастает как время расчета, так и качество сформированных траекторий. Это подтверждает эффективность предложенного метода для управления группами мобильных роботов в динамически меняющейся среде.

Группа роботов; мобильный робот; планирование траектории; препятствие; генетический алгоритм; модель.

L.A. Rybak, D.I. Malyshev, D.A. Dyakonov, A.A. Mamchenkova

A GENETIC ALGORITHM FOR PLANNING THE TRAJECTORY OF A GROUP OF MOBILE ROBOTS IN THE PRESENCE OF STATIONARY AND MOBILE OBSTACLES

The article discusses a trajectory planning method for a group of mobile robots that ensures safe movement and eliminates the possibility of collisions both between the robots themselves and with external obstacles, including moving objects. The developed mathematical model considers three main collision scenarios: intersection of robot trajectories within the group, interaction with stationary obstacles, and