УДК 004.896 **DOI** 10.15622/ia.24.3.1

М.Ю. Медведев, В.Х. Пшихопов, И.Д. Евдокимов

АЛГОРИТМ РОБАСТНОГО УПРАВЛЕНИЯ ОДНОМЕРНЫМ ДИНАМИЧЕСКИМ ОБЪЕКТОМ НА ОСНОВЕ ТАБЛИЧНОГО Q-МЕТОДА ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Медведев М.Ю., Пшихопов В.Х., Евдокимов И.Д. **Алгоритм робастного управления одномерным динамическим объектом на основе табличного Q-метода обучения с подкреплением.**

Аннотация. В статье представлен обзор в области систем управления динамическими объектами на базе методов машинного обучения с подкреплением. На основе проведенного анализа сделан вывод о актуальности развития методов управления, базирующихся на обучении с подкреплением. В статье предлагается интеллектуальный алгоритм робастного управления устойчивыми динамическими объектами с одним входом и одним выходом, базирующийся на табличном Q-методе обучения нулевого порядка. Алгоритм осуществляет стабилизацию выходной величины объекта управления с заданной погрешностью, если параметры и внешние возмущения объекта являются кусочно-постоянными неизвестными величинами, а вектор состояния является измеряемым. Новизна предложенного алгоритма заключается в новом инкрементальном способе формирования управления, который позволяет, базируясь на множестве из трех возможных действий, стабилизировать объект управления. Предложенный способ формирования множества управляющих воздействий позволяет обеспечить требуемую точность стабилизации выхода объекта, изменяя амплитуду приращения управления. Также элементом новизны является предложенное выражения для расчета вознаграждения, которое позволяет ограничить изменения управления. Предлагаемый алгоритм обладает высокой вычислительной эффективностью. После обучения вычисление управления сводится к вычислению индексов по результатам измерения, чтению данных из памяти по вычисленным индексам и нахождению максимального значения в векторе небольшой размерности. В работе исследованы условия сходимости алгоритма обучения и ограниченности ошибки управления. Разработанный алгоритм демонстрируется на примере синтеза робастного управления двигателем постоянного тока с независимым возбуждением. В ходе численного моделирования исследуется качество замкнутой системы при изменении параметров и задающего воздействия. Анализ результатов позволяет сделать выводы об эффективности синтезированного алгоритма. В статье приводятся результаты экспериментов, которые демонстрируют техническую реализуемость полученного алгоритма. Данный вопрос является важным, так как анализ источников показывает практически полное отсутствие технической реализации систем управления динамическими объектами, синтезированными с использованием методов обучения с подкреплением.

Ключевые слова: робастное управление, обучение с подкреплением, Q-алгоритм обучения, динамические объекты, неопределенные параметры, сходимость алгоритма обучения.

1. Введение. В настоящее время становятся популярными методы интеллектуального управления, базирующиеся на машинном обучении. Обещающие результаты показывают методы обучения с подкреплением [1], направленные на решение проблемы адаптивного

управления динамическими объектами в условиях неопределенности. Задача синтеза прямого адаптивного управления на базе Q-метода обучения с подкреплением была описана в работе [2]. Здесь можно выделить несколько основных проблем.

Первая проблема заключается в обеспечении асимптотической устойчивости замкнутой системы управления, базирующейся на методах машинного обучения.

Вторая проблема состоит в том, что в процессе обучения имеется возможность выработки обучающимся алгоритмом недопустимых воздействий. В этой связи обучение осуществляется в виртуальной среде, на реальном объекте реализуется дообучение. Чтобы минимизировать риски потери качества и устойчивости управления требуется наличие виртуальной модели, максимально близкой к реальному объекту. Однако, наличие такой модели делает задачу синтеза алгоритма управления реализуемой классическими методами автоматического управления [3]. В этом случае возникает вопрос о преимуществах, которые дает интеллектуальный регулятор.

В работах [4 – 13] проблема сохранения качества и устойчивости системы управления при он-лайн обучении интерпретируется как обеспечение свойства равномерной абсолютной ограниченной устойчивости (Uniformly Ultimate Boundedness Stability – UUB).

При этом можно выделить подход, базирующийся на модели, и безмодельный подход. В работах [4, 6 – 8, 10 – 13] используется подход, основанный на моделях. Рассматриваются дискретные [4, 7] или непрерывные [6, 8, 10 – 13] объекты управления с неопределенными правыми частями и возмущениями. Обычно накладываются условия дифференцируемости [4, 11] или непрерывности в смысле выполнения условий Липшица [6 – 8, 10 – 13] правых частей, внешние возмущения считаются ограниченными, а объект – управляемый. В рассматриваемых работах машинное обучение используется для аппроксимации неопределенных правых частей с их последующей идентификацией, а также для адаптивного решения задачи оптимального управления. В подавляющем числе методов, базирующихся на моделях, для аппроксимации используется однослойная нейронная сеть [4, 6, 8, 10 – 13], в частности популярной является аппроксимация радиальными базисными функциями [4, 11, 12]. Для оценивания оптимального решения уравнения Гамильтона-Якоби-Беллмана и уменьшения ошибки Беллмана также используются однослойные нейронные сети и архитектура обучения с подкреплением, известная как Actor-Critic [1].

Можно выделить ряд ограничений предлагаемых подходов.

Практически во всех статьях, основанных на моделях, за исключением работы [7], используются однослойные нейронные сети. Такие сети способны аппроксимировать нелинейные функции многих переменных, при этом ошибка аппроксимации стремится к нулю, если число нейронов (и, соответственно, число параметров настройки), стремится к бесконечности. Такая аппроксимация не дает принципиальных преимуществ по сравнению с обычной параметризацией и разложением по заданному базису. Она удобна с точки зрения выбора закона настройки параметров на основе теории устойчивости Ляпунова. Однако такой подход не использует преимущества многослойных нейронных сетей, с помощью которых получены наиболее впечатляющие результаты в области искусственного интеллекта.

Также можно отметить, что обеспечение свойства равномерной абсолютной ограниченной устойчивости не гарантирует выполнение прямых показателей качества. На наш взгляд он-лайн адаптация и не может обеспечить выполнение таких требований. Для этого необходимо дополнительно применять офф-лайн методы обучения.

В рамках основанного на моделях подхода различные аспекты проблемы обеспечения безопасности при обучении изучаются в настоящее время [14, 15, 16]. Так в работе [14] обеспечивается построение ограниченной области притяжения методом, сходим с методом Ляпунова. В работе [15] вводятся ограничения на состояние системы управления, которые выполняются, если известны индикаторная функция, указывающая на выход состояния за заданные ограничения и субоптимальные удовлетворяющие заданным ограничениям действия. В статье [16] выполнение заданных ограничений на состояние системы управления осуществляется в рамках структуры, формирующий параметрическую аппроксимацию решения уравнения Беллмана и формирование ограничений методами обучения с подкреплением.

Проблема обеспечения устойчивости также актуальна в безмодельных подходах [5, 9, 17 – 19], в которых управление строится только на основе анализа входных и выходных данных.

В работах [5] и [9] также обеспечивается свойство UUB. В работах [17, 18] вводятся управляющие барьерные функции (Control Barrier Functions), а также используется предварительное обучение, после которого осуществляется функционирование замкнутой системы. В статье [19] в процессе обучения методом Ляпунова обеспечивается устойчивость. Однако, как в основанных на моделях

подходах не обеспечивается выполнения прямых показателей качества в процессе обучения.

Безмодельные подходы достаточно популярны в различных приложениях. Так в работе [20] рассматривается проблема управления объектом первого порядка с запаздыванием. Для ее решения успешно применяется метод обучения с подкреплением Deep Deterministic Policy Gradient (DDPG). В [21] с применением метода DDPG решена задача перехвата подвижного объекта, который движется по прямолинейной или круговой траектории.

Проведенный анализ позволяет сделать следующие выводы:

1. Преимущественно исследуются методы он-лайн обучения (адаптации), которые гарантируют устойчивость, но не гарантируют заданного качества управления в процессе обучения.

2. В системах управления динамическими объектами, в подавляющем большинстве статей при анализе устойчивости используются однослойные сети, что удобно для использования теории устойчивости Ляпунова.

3. Проводимый анализ устойчивости не позволяет выбрать структуру нейронной сети, функцию активации и гиперпараметры при обучении, хотя эти характеристики прямо влияют на устойчивость процесса обучения [22].

4. Несмотря на заявленные теоретические гарантии обеспечения устойчивости в статьях отсутствует практическая реализация теоретических результатов. Все исследования остаются в рамках симуляций. Это указывает на недостаточную развитость инженерных методик синтеза систем управления на базе обучения с подкреплением. При этом такая ситуация резко отличается от методов обучения с учителем, на основе которых успешно строятся системы распознавания [23, 24], обработки сигналов [25] и текстов [26], планирования движения [27, 28], реализации регуляторов [29].

В данной работе развивается инженерная методика синтеза регулятора с использованием оф-лайн Q-метода [30]. Преимуществом оф-лайн подхода является то, что он способен обеспечивать заданное качество в течение всего переходного процесса, так как во время функционирования регулятор не обучается. В данной работе рассматривается задача с дискретными значениями управления. Предлагается вариант, позволяющий обеспечить за счет малого интервала дискретизации по уровню и инкрементального способа формирования управления обеспечить слежение за задающим сигналом с малой ошибкой. Кроме того, имеется широкий класс управляемых объектов с релейным характером управления, для

которых такая постановка задачи является актуальной. В частности, релейное управление применяется при управлении мощными двигателями и тепловыми процессами.

Новизна результатов, представленных в настоящей статье, определяется следующими положениями:

1. Предложен инкрементальный способ формирования управляющего воздействия, который обеспечивает робастную стабилизацию управляемого выхода объекта с заданной точностью. Метод требует построения только трех таблиц ценности действий, что определяет его вычислительную эффективность.

2. Предложен способ формирования вознаграждения, предотвращающий резкие изменения управляющих воздействий, что позволяет добиться более плавных переходных процессов в замкнутой системе.

3. Проведенные натурные эксперименты показывают отсутствие необходимости дообучения алгоритма управления при переходе из виртуальной среды к реальному объекту для рассмотренного класса задач. Как правило, такой бесшовный переход считается проблематичным и представляет собой проблемы, известную под названием «разрыв реальности».

2. Постановка задачи. Рассматривается система управления, представленная на рисунке 1, включающая односвязный объект управления (ОУ) и устройство управления (УУ). УУ формирует на вход ОУ управления u_k. С выхода ОУ на вход УУ подаются состояние x_k и вознаграждение R_k. На вход объекта также воздействует возмущение f.

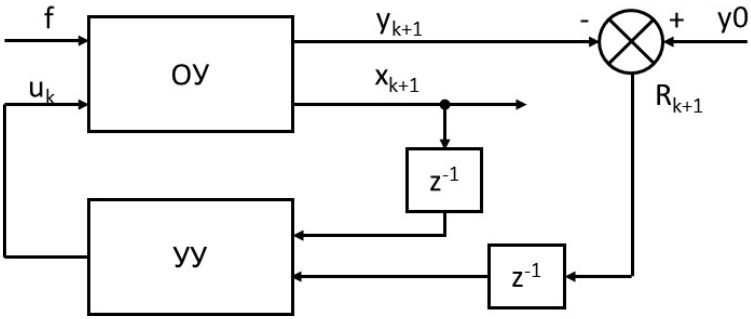

Рис. 1. Структура системы управления

Объектом управления является линейная динамическая стационарная система с неизвестными параметрами и неизмеряемым возмущением, которая описывается системой уравнений вида:

$$\frac{dx(t)}{dt} = Ax(t) + bu(t) + hf(t), \quad y(t) = c^T x(t), \tag{1}$$

где $x(t)$ – вектор состояния размерности n; $u(t)$ – скалярное управляющее воздействие; $f(t)$ – скалярное возмущение; $y(t)$ – регулируемая выходная величина; A – матрица постоянных неизвестных параметров размерностью n×n; b, h – векторы постоянных неизвестных параметров размерностью n×1; c – вектор известных параметров размерностью n×1.

Предположение 1. Предполагается, что матрица A является Гурвицевой. Управляющее воздействие принадлежит заданному интервалу $u(t) \in [u_{min}, u_{max}]$. Выходная величина объекта (1) ограничена и также принадлежит известному интервалу $y(t) \in [y_{min}, y_{max}]$. Внешнее воздействие также ограничено по величине $f_{min} \leq f(t) \leq f_{max}$.

Задача состоит в разработке и исследовании интеллектуального робастного алгоритма управления $u = u(x, y0)$, обеспечивающего стабилизацию выходной величины $y(t)$ в заданной окрестности желаемого значения $y0$. При разработке будем использовать методы обучения с подкреплением на основе временных различий [1].

3. Решение поставленной задачи. Для решения поставленной задачи модель объекта (1) представляется в дискретной форме

$$x_{k+1} = x_k + \Delta t(Ax_k + bu_k + hf_k), \quad y_k = c^T x_k, \tag{2}$$

где Δt – шаг дискретизации по времени; k – обозначение дискретного момента времени $t_k = k\Delta t$.

Управление u_k выбирается из множества допустимых управлений

$$u_k \in \{u_i\}, \qquad i = 1,2, \dots, N_u, u_{min} \leq u_i \leq u_{max}, \tag{3}$$

где N_u – положительное число.

Вознаграждение в момент времени k выбирается в виде квадрата ошибки, взятого с противоположным знаком:

$$R_k = -(y0 - y_k)^2, \quad (4)$$

где y_k – значение управляемой величины динамического объекта (1) в момент времени t.

В качестве алгоритма обучения используется метод Q-обучения нулевого порядка [1], который относится к методам обучения с разделенной стратегией и описывается выражением:

$$Q_{t+1}(\tilde{x}_k) = Q_t(\tilde{x}_k) + \alpha [R_{k+1} + \gamma \max_u Q_{t+1}(\tilde{x}_{k+1}, u) - Q_t(\tilde{x}_k)], \quad (5)$$

где $Q_{t+1}(\tilde{x}_k)$ – значение ценности состояния $\tilde{x}_k$ в момент времени t+1; $Q_t(\tilde{x}_k)$ – значение ценности состояния $\tilde{x}_k$ в момент времени t; R_{k+1} – вознаграждение за переход в состояние $\tilde{x}_k$; α – шаг обучения; γ – коэффициент обесценивания; $\tilde{x}_k = [x_k \;\; y0]^T$.

Выражение (5) формирует обновление ценности действий в каждом состоянии следующим образом.

1) В момент времени t в состоянии $\tilde{x}_k$ выбирается текущее управление u_k в соответствии с выражением

$$u_k = \max_u Q_t(\tilde{x}_k, u). \quad (6)$$

2) Управление u_k (6) подается на вход объекта (1).

3) На следующем шаге $k+1$ осуществляется вычисление вознаграждения согласно (4) и для нового состояния $\tilde{x}_{k+1}$ определяется оптимальное управление u_{k+1} согласно (6) и его ценность $Q_{t+1}(\tilde{x}_{k+1})$.

4) Согласно выражению (5) осуществляется обновление функции ценности $Q_k(\tilde{x}_k)$.

Таким образом, особенностью выражения (5) является то, что обновление ценности состояния $\tilde{x}_k$, в котором находилась система в момент времени t, осуществляется в момент времени t+1, после вычисления максимально полезного действия в состоянии $\tilde{x}_{k+1}$.

Как известно [1], в результате применения описанной процедуры осуществляется максимизация вознаграждения, определяемого выражением

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \quad (7)$$

Метод Q-обучения, как метод с разделенной с разделенной стратегией, допускает использование ε-жадного алгоритма [1] в процессе обучения и жадного алгоритма в процессе функционирования обученного агента. Основное преимущество метода Q-обучения заключается в том, что обновление функции ценности $Q_k(\tilde{\mathrm{x}}_\mathrm{k})$ осуществляется сразу после выполнения действия u_k и определения нового состояния $\tilde{x}_{\mathrm{k+1}}$ и вознаграждения R_{t+1}. При этом не требуется хранить массивы состояний и вознаграждений как в методе Монте-Карло, т.е. все операции по обновлению функции ценности (5) являются скалярными. Это позволяет уменьшить требования к вычислительным затратам.

Кроме дискретизации по времени, в процессе обучения Q-методом, используется дискретизация по уровню. Рассмотрим рабочую область объекта (1), в которой переменные состояния $x_i(t)$ ограничены

$$x_{imin} \le x_i(t) \le x_{imax}\,, \quad (8)$$

где x_{imin}, x_{imax} – известные константы; i=1,2,…,n.

Такие ограничения широко распространены на практике. Например, в электрических системах сигналы ограничены напряжением питания, в электрических двигателях рабочие токи ограничены и при выходе за заданный диапазон осуществляется отключение питания.

Тогда дискретизация диапазона $[x_{imin}, x_{imax}]$ на $n_d + 1$ значение осуществляется в соответствии с выражением

$$x_{ik} = round\big(n_d * x_i(k\Delta t)/(x_{imax} - x_{imin})\big), i = 1,2,\dots,n, \mathrm{k} = 0,1,2,\dots \quad (9)$$

Проведем оценку устойчивости системы управления (2) – (6). Для краткости перепишем выражение (5) в виде

$$Q_{k+1} = Q_k + \alpha\,[R_{k+1} + \gamma \max_u Q_{k+1} - Q_k]. \quad (10)$$

Введем в рассмотрение следующую квадратичную функцию

$$V_k = Q_k^2 + x_k^2. \quad (11)$$

Рассмотрим приращение функции (11)

$$\Delta V_{k+1} = V_{k+1} - V_k = Q_{k+1}^2 + x_{k+1}^2 - Q_{k-1}^2 - x_{k-1}^2. \quad (12)$$

Подстановка (10) и (2) в правую часть (12) приводит к выражению

$$\Delta V_{k+1} = (Q_k + \alpha [R_{k+1} + \gamma \max_u Q_{k+1} - Q_k])^2 + \left(x_k + \Delta t(Ax_k + bu_k + hf_k)\right)^2 - Q_k^2 - x_k^2. \quad (13)$$

Согласно Предположению 1 объект управления является устойчивым, а внешние воздействия ограничены. В этой связи вектор x_k также ограничен, поэтому при достаточно малом шаге дискретизации по времени Δt слагаемым $\Delta t(Ax_k + bu_k + hf_k)$ в выражении (13) можно пренебречь. Тогда, при $\Delta t \to 0$ получаем

$$\Delta V_{k+1} = (Q_k + \alpha [R_{k+1} + \gamma Q_{k+1} - Q_k])^2 - Q_k^2. \quad (14)$$

В выражении (14) раскроем скобки и сократим подобные

$$\Delta V_{k+1} = 2Q_k \alpha [R_{k+1} + \gamma \max_u Q_{k+1} - Q_k] + (\alpha [R_{k+1} + \gamma \max_u Q_{k+1} - Q_k])^2. \quad (15)$$

Выражение $R_{k+1} + \gamma \max_u Q_{k+1} - Q_k$ в (15) хорошо известно, как ошибка Q-метода обучения. Сходимость этой ошибки при $k \to \infty$ доказана [1, 11]. Поэтому, согласно (15) можно утверждать, что замкнутая система устойчива по Ляпунову, что гарантирует ограниченность всех сигналов.

Описанный алгоритм является стандартным алгоритмом Q-обучения. Он обладает тем недостатком, что при повышении требований к точности стабилизации необходимо увеличивать мощность множества управлений (3). Однако такое увеличение приводит к пропорциональному росту числа Q-таблиц и сложностям при обучении, так как известно, что табличные методы хорошо работают при небольшом числе N_u [1].

В этой связи в данной работе предлагается инкрементальный способ вычисления управления, в котором вычисляется не абсолютное значение управления, а его приращение в соответствии с выражением:

$$u_k = u_{k-1} + \Delta u_k. \quad \Delta u_k = \max_u Q_t(\tilde{x}_k, u), \quad u \in \{-\Delta u, 0, \Delta u\}, \quad (16)$$

где Δu – приращение управления на текущем шаге.

При этом, в выражении (16) $\tilde{x}_{\text{k}} = [x_k \;\; y0 \;\; u_{k-1}]^T$, т.е. состояние системы включает переменные состояния объекта, задающее воздействие и предыдущее значение управления.

Из выражения (16) следует, что регулятор $u_k = u_{k-1} + \max_u Q_t(\tilde{x}_{\text{k}}, u)$ является динамическим, так как значения управления u_k зависят не только от текущего состояния $\tilde{x}_{\text{k}}$, но и от предыдущего значения управляющего воздействия u_{k-1}.

Очевидно, что управление (16) не меняет условия сходимости по сравнению с управлением (6). При этом, выражение (16) обеспечивает более плавное изменение управляющего воздействия.

В системах обучения с подкреплением изменение управляющего воздействия обуславливается не только множеством допустимых управлений, но и вознаграждением. В этой связи в данной работе предлагается модифицированный способ формирования вознаграждения, при котором в выражение (4) добавляется штраф для резких изменений управления:

$$R_k = -(y0 - y_k)^2 - \beta * |\Delta u|, \tag{17}$$

где β – коэффициент штрафа; $\Delta u = u_k - u_{k-1}$.

На практике всегда присутствуют измерительные шумы, поэтому их необходимо фильтровать. Данный вопрос не является предметом настоящей статьи, поэтому в ней применяется простейший алгоритм сглаживания вида

$$u = \lambda * u_k + (1 - \lambda) * u_{k-1}, \tag{18}$$

где λ – коэффициент сглаживания.

При $\lambda = 0.5$ алгоритм (8) является алгоритмом скользящего усреднения по двум изменениям.

3. Результаты численного моделирования. На рисунках 2 – 4 представлены результаты моделирования замкнутой системы управления (2) – (6), обученной на 3 000 000 эпизодах для следующих параметрах: время моделирования эпизода 8.0 с; шаг моделирования 0.01; шаг обучения $\alpha = 0.05$; коэффициент обесценивания $\gamma = 1.0$.

В качестве объекта управления выбран двигатель постоянного тока, модель которого в дискретной форме имеет вид

$$\omega_{k+1} = \omega_k + \frac{\Delta t}{J}(k_m I_k - M_{lk}),$$
$$I_{k+1} = I_k + \frac{\Delta t}{L}(u_k - k_m \omega_k - r I_k), \tag{19}$$

где ω_k, I_k, u_k, M_{lk} – частота вращения, ток статора, управляющее напряжение и момент нагрузки двигателя.

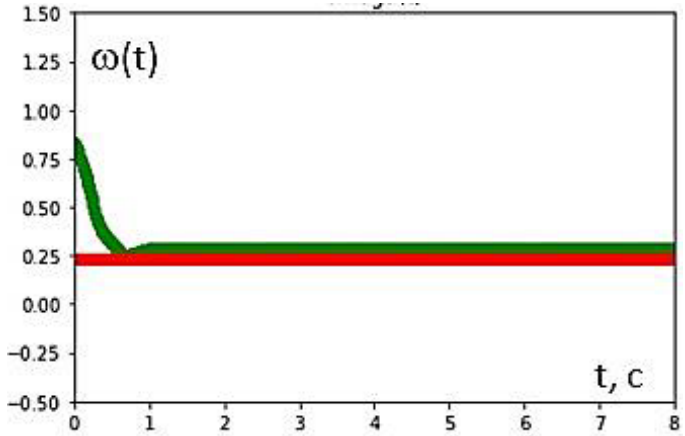

Рис. 2. Изменение частоты вращения в замкнутой системе управления (3) – (6), (19), (20)

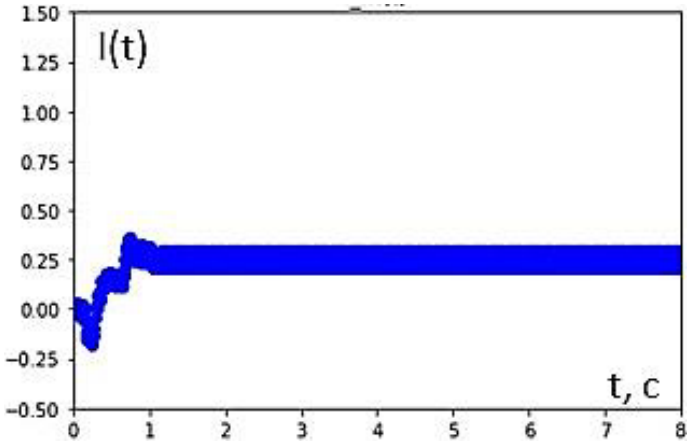

Рис. 3. Изменение тока в замкнутой системе (3) – (6), (19), (20)

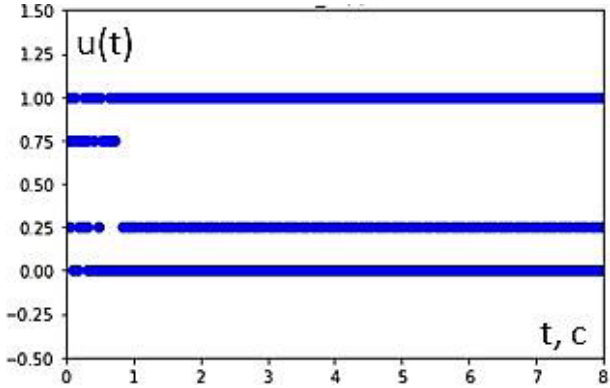

Рис. 4. Изменение управляющего воздействия в замкнутой системе (3) – (6), (19), (20)

Параметры двигателя (19) являются интервальными и заданы в следующих диапазонах: $J \in [0.1,\ 0.3]$; $k_m \in [0.7,\ 1.3]$; $M_l \in [0.1,\ 0.3]$; $L \in [0.07,\ 0.13]$; $r \in [0.7,\ 1.3]$.

Множество управляющих воздействий содержит 5 значений

$$u_k \in \{0, 0.25u_{max}, 0.5u_{max}, 0.75u_{max}, u_{max}\}. \quad (20)$$

При моделировании замкнутой системы значения параметров равны $J = 0.219$, $k_m = 0.944$, $M_l = 0.237$, $L = 0.113$, $r = 1.061$, $\omega_0 = 0.234$.

Полученное значение СКО равно 0.065. При этом СКО рассчитывается по всему переходному процессу. Для установившего режима СКО равно 0.015, что составляет примерно 45 % от размера ячейки дискретизации пространства состояния.

По сравнению с работой [30], в которой реализовано управление с обратной связью по выходной величине ω_k можно отметить следующее. В работе [30] реализован алгоритм, при котором необходимо обучение при неизвестных, но постоянных параметрах двигателя и заданной нагрузке. При изменении параметров и нагрузки требуется заново обучать алгоритм управления. В данном случае не требуется каждый раз обучать алгоритм управления. Для примера, на рисунках 5–8 представлены переходные процессы в замкнутой системе при скачкообразном изменении нагрузки и желаемой частоты двигателя.

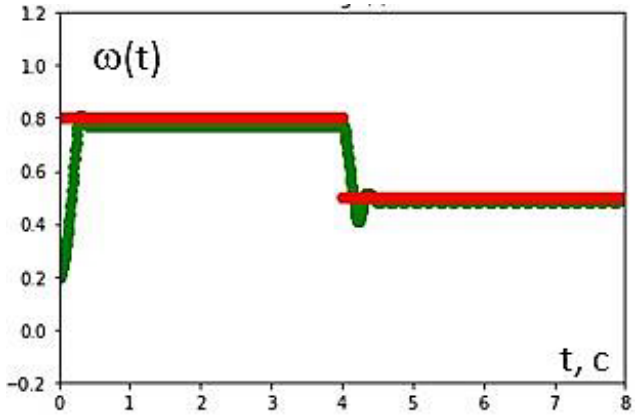

Рис. 5. Изменение частоты в системе (3) – (6), (19), (20) при скачке задающего воздействия ω_0

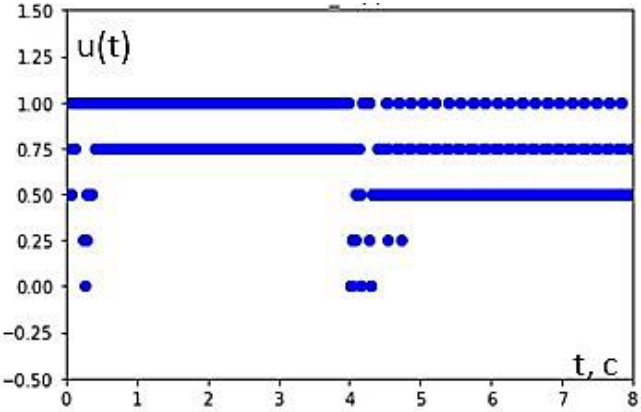

Рис. 6. Изменение управления в системе (3) – (6), (19), (20) при скачке задающего воздействия ω_0

На рисунке 5 и рисунке 6 в момент времени t = 4 с происходит скачкообразное изменение задающей частоты вращения вала двигателя с 0.8 до 0.5. Алгоритм управления (6) после переходного процесса стабилизирует частоту вращения в окрестности нового значения. При моделировании замкнутой системы ее параметры сгенерированы случайным образом и равны: $J = 0.119$, $k_m = 0.969$, $M_l = 0.115$, $L = 0.106$, $r = 0.922$. СКО ошибки управления в течение всего времени составило 0.034, а в установившемся режиме 0.017.

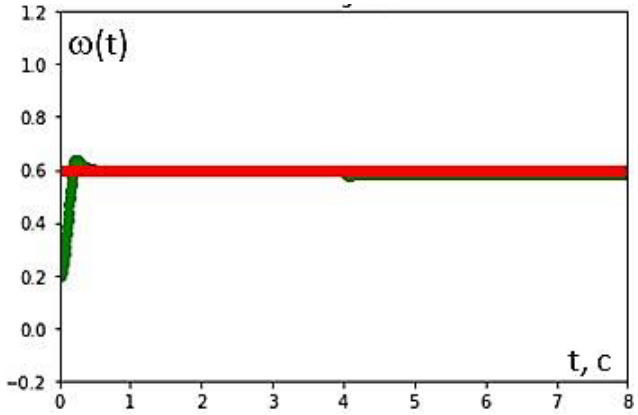

Рис. 7. Изменение частоты в системе (3) – (6), (19), (20) при скачке нагрузки M_l

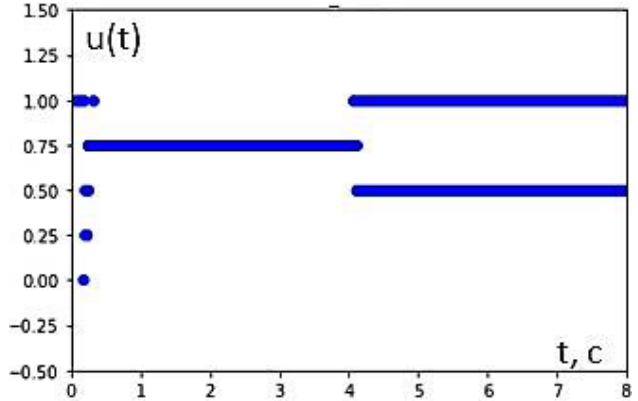

Рис. 8. Изменение управления в системе (3) – (6), (19), (20) при скачке нагрузки M_l

На рисунке 9 представлена проекция функции ценности для действия $u_k = 0$ при $\omega_0 = 0.1$. Функция построена в логарифмическом масштабе. Так как значения функции отрицательные, то построена функция логарифма от модуля функции ценности. Поэтому ее оптимальное значение находится в минимуме, который при малых токах примерно соответствует значению $\omega = \omega_0$. При увеличении тока оптимальное значение функции ценности смещается в сторону возрастания ω.

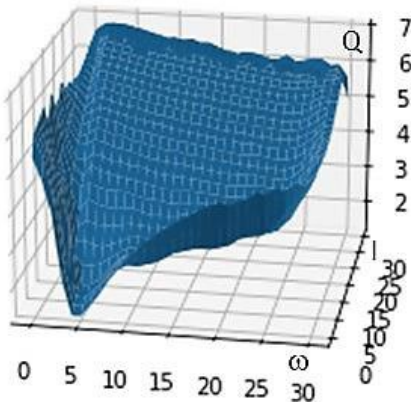

Рис. 9. Проекция функции ценности $\log(-Q(\omega, I))$ при $u_k = 0$ и $\omega_0 = 0.1$

На рисунках 10 и 11 представлены результаты моделирования обученной замкнутой системы для управления частотой вращения двигателя для инкрементального управления (16).

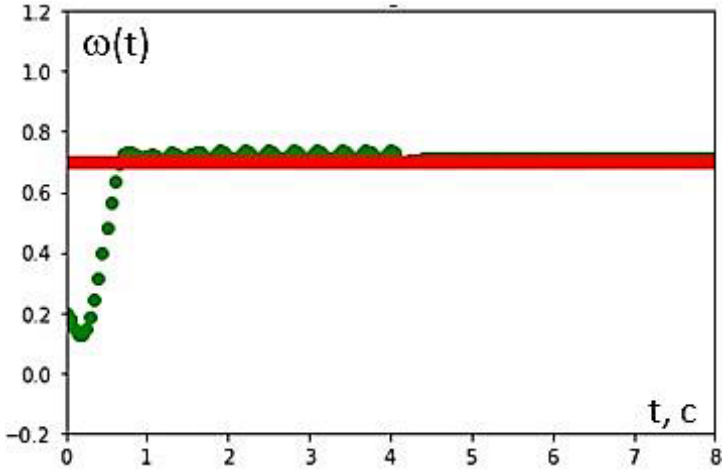

Рис. 10. Изменение частоты в замкнутой системе (16), (4), (5), (19)

Моделирование проводилось при следующих параметрах двигателя: $J = 0.29$, $k_m = 1.081$, $M_l = 0.15$, $L = 0.081$, $r = 0.975$, $\omega_0 = 0.7$. В момент времени t=4 с имеется скачок нагрузки до значения $M_l = 0.2$. При этом величина $\Delta u = 0.1$, шаг обновления управления равен 0.05 с, а шаг моделирования двигателя 0.01.

СКО замкнутой системы в установившемся режиме составило 0.017, что примерно в 2 раза меньше размера ячейки дискретизации. Отметим, что моделируется гибридная система, в которой управляющий алгоритм является дискретным, а объект непрерывным.

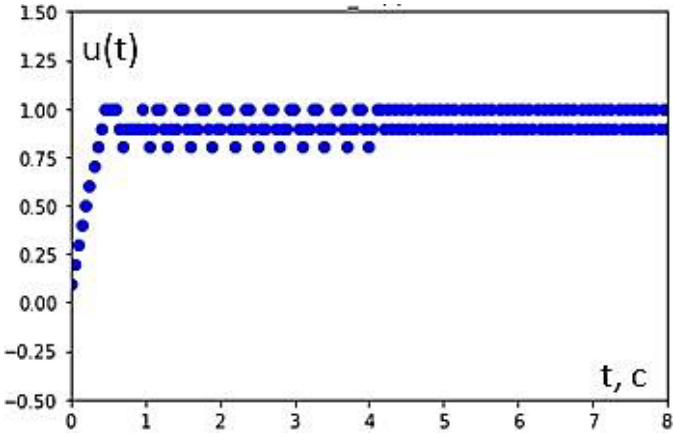

Рис. 11. Изменение управления в замкнутой системе (16), (4), (5), (19)

4. Результаты натурных экспериментов. Для тестирования метода управления приводом на основе Q-обучения был выбран малый привод постоянного тока RB300200-0B101REего характеристики были измерены эмпирически и приведены в таблице 1.

Таблица 1. Параметры привода постоянного тока

Параметры привода	**Величина**	**Погрешность**
Момент инерции, (кг ∗ м^2)	10^{-6}	$5{\cdot}10^{-7}$
Коэффициент вязкого трения, (Н*м*с/рад)	$2{\cdot}10^{-6}$	$5{\cdot}10^{-7}$
Противо-ЭДС, (В*с/рад)	0.0169	0.002
Сопротивление, (Ом)	10	1.5
Индуктивность, (Гн)	$2.83{\cdot}10^{-3}$	$3{\cdot}10^{-4}$
Максимальное напряжение, (В)	12	–
Максимальная угловая скорость, (рад/с)	710	–

В ходе экспериментов исследованы формирование управления на основе множества (3). Для обучения и формирования управления использованы средства Matlab/Simulink.

При управлении на основе множества (3) обучение длилось 500 тыс. эпизодов по 5 секунд каждый. После обучения в виртуальной среде алгоритм управления применяется на реальном приводе. Шаг дискретизации управляющего напряжения равен 0.5 В. Состояния по угловой скорости дискретизируются с шагом 10 рад/с.

Схема системы управления в Matlab/Simulink представлена на рисунке 12.

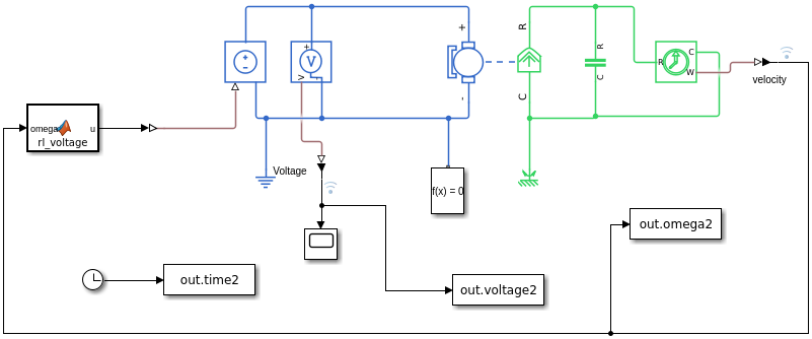

Рис. 12. Схема системы управления в Matlab/Simulink

Схема подключения привода представлена на рисунке 13, а внешний вид стенда для эксперимента – на рисунке 14.

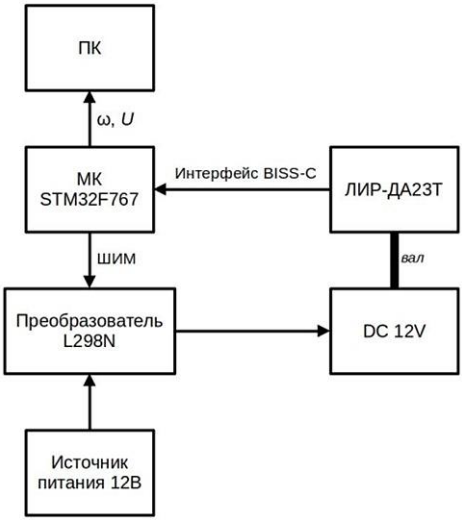

Рис. 13. Блок-схема подключения привода

Как видно из рисунка 13, подключение привода осуществляется через микроконтроллер STM32F767 и преобразователь L298N. Для измерения частоты вращения вала двигателя используется датчик ЛИР-ДА23Т.

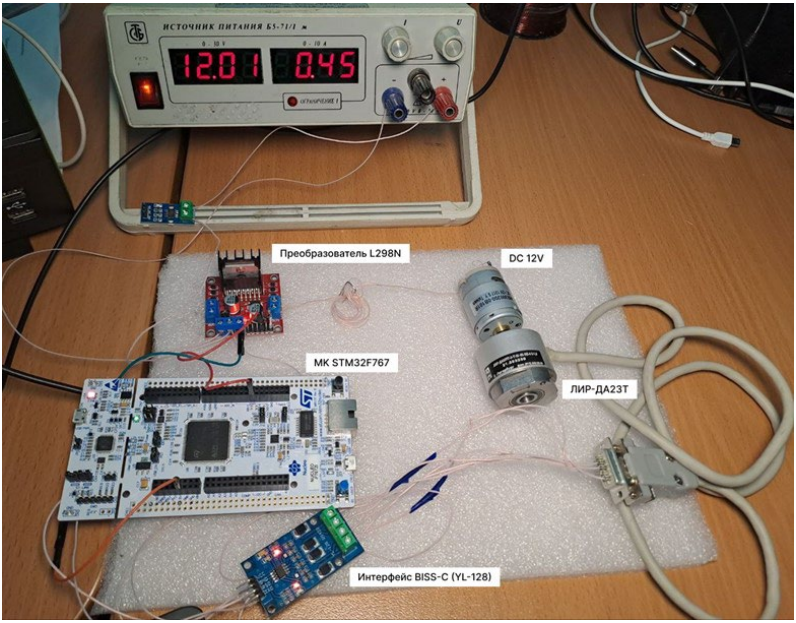

Рис. 14. Схема управления реальным приводом

Переходные процессы обученной системы управления представлены на рисунках 15 и 16.

Результаты эксперимента показывают, что ошибка поддержания частоты составляет около 25 рад/с, что составляет 2,5 шага дискретизации по частоте. Основной вклад в ошибку управления вносят измерительные шумы.

Результаты моделирования для системы управления с вознаграждением (17) и фильтром (18) представлены на рисунках 17 и 18. Как видно из рисунка 17 ошибка управления составила около 10 рад/с, что соответствует используемому шагу дискретизации частоты вращения по уровню.

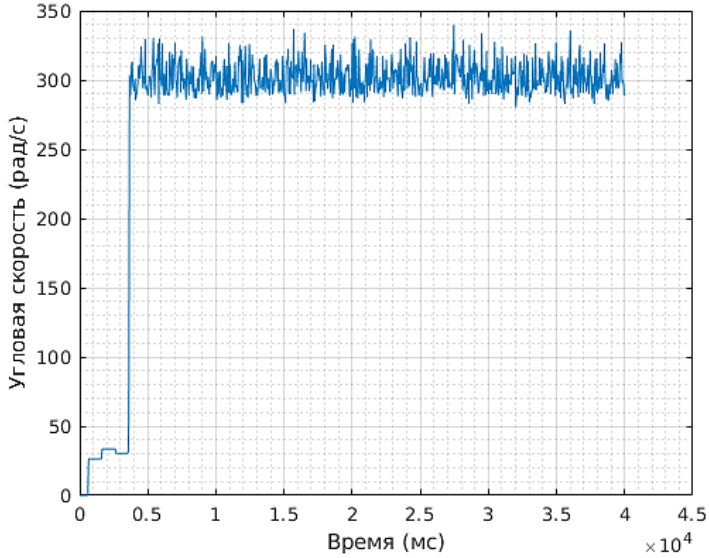

Рис. 15. Угловая скорость для управления из множества (3)

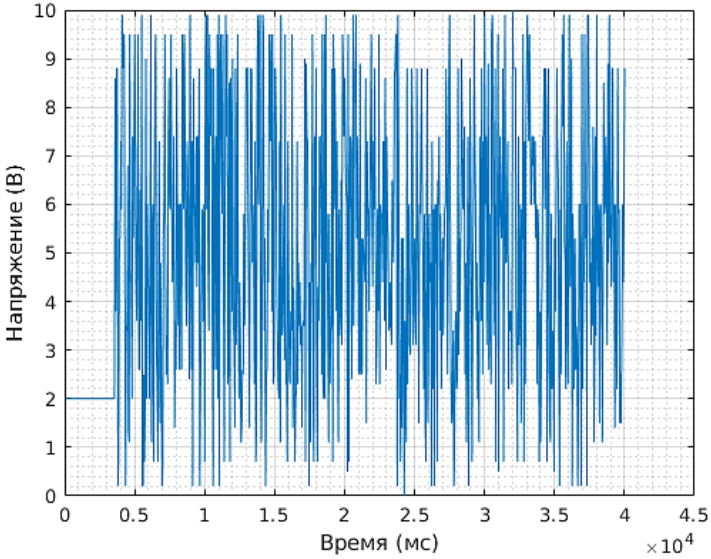

Рис. 16. Угловая скорость для управления из множества (3)

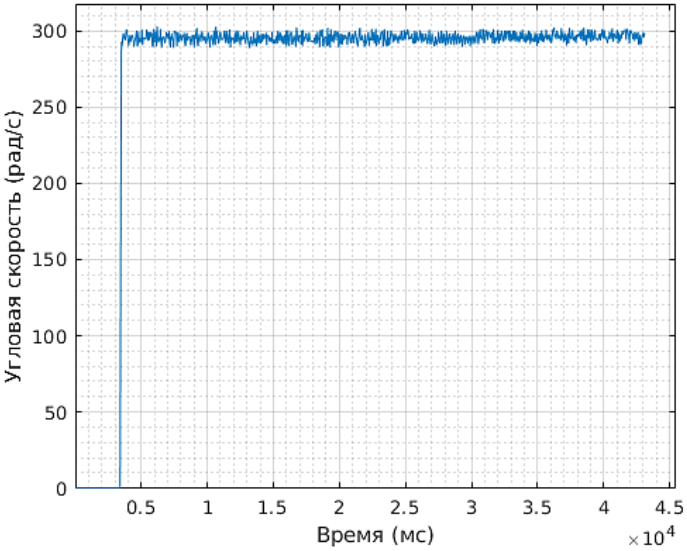

Рис. 17. Угловая скорость для управления из множества (3), вознаграждения (19) и фильтра (20)

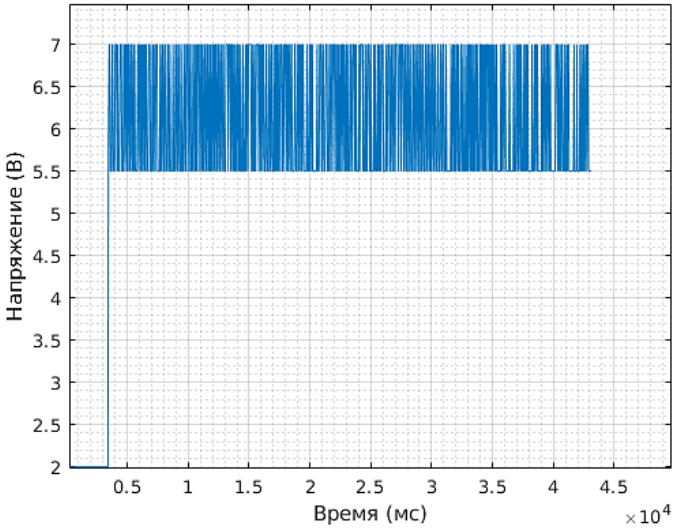

Рис. 18. Угловая скорость для управления из множества (3), вознаграждения (19) и фильтра (20)

Заключение. В статье представлены результаты исследования по применению Q-метода обучения с подкреплением для стабилизации выходной величины динамического объекта в заданном значении. Проведенный анализ литературы показал, что в большинстве работ развиваются он-лайн методы обучения, которые не позволяют удовлетворить требуемым прямым показателям качества в процессе адаптации. В этой связи офф-лайн методы обучения представляют интерес для технических систем.

Результаты, полученные в данной статье, позволяют сделать следующие выводы:

- офф-лайн Q-метод обучения с дискретным входом и выходом позволяет стабилизовать выход динамической системы с возмущением, параметры которой являются интервальными. При этом стабилизация осуществляется с точностью, определяемой шагом дискретизации;
- с использованием натурного эксперимента показано отсутствие необходимости дообучения алгоритма управления при переходе из виртуальной среды к реальному объекту. Этот факт позволяет создавать на основе алгоритмов обучения с подкреплением новый класс робастных систем управления для исполнительных механизмов;
- полученный алгоритм является вычислительно эффективным. Для нахождения управления достаточно реализовать доступ к элементам табличных функций ценности по имеющимся индексам и функцию поиска максимума в массиве размером $1 \times N_u$. По сравнению с традиционными ПИД-регуляторами, требующими реализации численного интегрирования и дифференцирования, такой регулятор обладает более высоким быстродействием;
- в работе предложен подход к уменьшению числа N_u, который заключается в инкрементальной форме вычисления управления (18). В этом случае управление u_k становится динамическим, так как оно зависит от предыдущего значения u_{k-1}. Изменяя величину Δu_k можно уменьшать величину дискретизации управления при $N_u = 3$.

Представленный в данной статье подход может быть эффективно применен для построения робастных систем управления объектами 2-3-го порядков. Очевидно, увеличение числа переменных приведет к необходимости использования различных, в том числе нейросетевых, аппроксимаций. Однако, использование аппроксимации функции ценности, вместе с разделенной стратегией обучения, при необходимости использования оценок вознаграждений вместо

фактических вознаграждений или полных доходов, приводит к проблеме расходимости алгоритмов обучения. Кроме того, сходимость процесса аппроксимации, строго говоря, доказана только для однослойных нейронных сетей (которыми может быть представлено, например, радиально-базисное или полиномиальное разложение). Наиболее точная многомерная аппроксимация достигается многослойными нейронными сетями, для которых не доказана асимптотическая сходимость в динамических системах.

Литература

1. Sutton R., Barto A. Reinforcement Learning. An Introduction. Second Edition. Cambridge: MIT Press, 2018. vol. 1. no. 1. pp. 9–11.
2. Sutton R.S., Barto A.G., Williams R.J. Reinforcement learning is direct adaptive optimal control. IEEE Control Systems Magazine. 2002. vol. 12(2). pp. 19–22.
3. Pshikhopov V., Medvedev M. Multi-Loop Adaptive Control of Mobile Objects in Solving Trajectory Tracking Tasks. Automation and Remote Control. 2020. vol. 81. pp. 2078–2093. DOI: 10.1134/S0005117920110090.
4. Shih P., Kaul B., Jagannathan S., Drallmeier J. Near Optimal Output-Feedback Control of Nonlinear Discrete-Time Systems in Nonstrict Feedback Form with Application to Engines. IEEE International Joint Conference on Neural Networks. 2007. pp. 396–401.
5. Xu B., Yang C., 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.
6. Mu C., Ni Z., Sun C., 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.
7. 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.
8. 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.
9. Luo B., Yang Y., Liu D., Wu H.-N. Event-Triggered Optimal Control with Performance Guarantees Using Adaptive Dynamic Programming. IEEE Transactions on Neural Networks and Learning Systems. 2019. vol. 31(1). pp. 76–88.
10. Yang X., Xu M., Wei Q. Dynamic Event-Sampled Control of Interconnected Nonlinear Systems Using Reinforcement Learning. IEEE Transactions on Neural Networks and Learning Systems. 2022. vol. 35(1). pp. 923–937. DOI: 10.1109/TNNLS.2022.3178017.
11. Zhang H., Zhao X., Wang H., Zong G., 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. 2022. vol. 35(2). pp. 1559–1571. DOI: 10.1109/TNNLS.2022.3183991.
12. Dong C., Chen L., Dai S.-L. Performance-Guaranteed Adaptive Optimized Control of Intelligent Surface Vehicle Using Reinforcement Learning. IEEE Transactions on Intelligent Vehicles. 2023. vol. 9. no. 2. pp. 3581–3592. DOI: 10.1109/TIV.2023.3338486.

13. Dao P.N., Phung M.H. 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. DOI: 10.1016/j.compeleceng.2024.109870.
14. 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.
15. Thananjeyan B., Balakrishna A., Rosolia U., Li F., McAllister R., Gonzalez J.E., Levine S., Borrelli F., Goldberg K. Safety Augmented Value Estimation From Demonstrations (SAVED): Safe Deep Model-Based RL for Sparse Cost Robotic Tasks. IEEE Robotics and Automation Letters. 2020. vol. 5(2). pp. 3612–3619.
16. Zanon M., Gros S. Safe Reinforcement Learning Using Robust MPC. IEEE Transactions on Automatic Control. 2020. vol. 66(8). pp. 3638–3652. DOI: 10.1109/TAC.2020.3024161.
17. 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. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-19). 2019. vol. 33. no. 01. pp. 3387–3395.
18. 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.
19. 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.
20. Боровик В.С., Шидловский С.В. Обучение с подкреплением в системах управления объектами с транспортным запаздыванием. Автометрия. 2021. Т. 57(3). С. 48–57.
21. Галяев 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.
22. 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. DOI: 10.17587/mau.25.345-353.
23. 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.
24. 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.
25. 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.
26. 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. DOI: 10.15622/ia.24.1.10.
27. 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. DOI: 10.17587/mau.25.559-567.
28. Gaiduk A.R., Martjanov O.V., Medvedev M.Yu., Pshikhopov V.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. DOI: 10.17587/mau.21.470-479.
29. 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.
30. Карапеев А.Н., Косенко Е.Ю., Медведев М.Ю., Пшихопов В.Х. Исследование интеллектуального адаптивного алгоритма управления на базе метода обучения с подкреплением. Известия ЮФУ. Технические науки. 2025. № 2. С. 162–175.

Медведев Михаил Юрьевич — д-р техн. наук, доцент, ведущий научный сотрудник, НИИ робототехники и процессов управления, Южный федеральный университет (ЮФУ). Область научных интересов: адаптивное и робастное управление мобильными роботами, оценивание возмущений, методы анализа и синтеза систем автоматического управления. Число научных публикаций — 260. medvmihal@sfedu.ru; улица Шевченко, 2, 347922, Таганрог, Россия; р.т.: +7(863)437-1694.

Пшихопов Вячеслав Хасанович — д-р техн. наук, профессор, директор, НИИ робототехники и процессов управления, Южный федеральный университет (ЮФУ). Область научных интересов: управление мобильными роботами в неопределенных средах, оптимальное управление роботами, анализ и синтез систем группового управления, интеллектуальное управление и планирование в робототехнике. Число научных публикаций — 300. pshichop@rambler.ru; улица Шевченко, 2, 347922, Таганрог, Россия; р.т.: +7(863)437-1694.

Евдокимов Игорь Дмитриевич — аспирант, инженер, НИИ робототехники и процессов управления, Южный федеральный университет (ЮФУ). Область научных интересов: адаптивное и робастное управление мобильными роботами. Число научных публикаций — 1. ievdokimov@sfedu.ru; улица Шевченко, 2, 347922, Таганрог, Россия; р.т.: +7(863)437-1694.

Поддержка исследований. 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».

DOI 10.15622/ia.24.3.1

M. MEDVEDEV, V. PSHIKHOPOV, I. EVDOKIMOV

A ROBUST CONTROL ALGORITHM FOR SINGLE INPUT SINGLE OUTPUT DYNAMIC OBJECT BASED ON TABLE-BASED Q-METHOD OF REINFORCEMENT LEARNING

Medvedev M., Pshikhopov V., Evdokimov I. **A Robust Control Algorithm for Single Input Single Output Dynamic Object Based on Table-Based Q-Method of Reinforcement Learning.**

Abstract. The article provides an overview in the field of dynamic object control systems based on reinforcement learning. Based on the analysis, it is concluded that the development of control methods based on reinforcement learning is relevant. The article proposes an intelligent algorithm for robust control of stable dynamic objects with one input and one output, based on the tabular Q-learning method of zero order. The algorithm stabilizes the output value of the control object with a given error if the parameters and external disturbances of the object are piecewise constant unknown quantities, and the state vector is measurable. The novelty of the proposed algorithm lies in a new incremental method of control formation, which allows, based on a set of three possible actions, to stabilize the control object. The proposed method of forming a set of control actions makes it possible to ensure the required accuracy of stabilizing the output of an object by changing the amplitude of the control increment. The proposed algorithm has high computational efficiency. After training, the control calculation is reduced to calculating indexes based on measurement results, reading data from memory based on calculated indexes, and finding the maximum value in a small vector. For a discrete description of the control object, the conditions of convergence of the learning algorithm and the limitation of the control error are investigated. The developed algorithm is demonstrated by the example of the synthesis of robust control of a DC motor with independent excitation. In the course of numerical simulation, the quality of a closed system is investigated when the parameters and the control action change. The analysis of the simulation results allows us to draw conclusions about the effectiveness of the synthesized algorithm. The article also provides the results of a real experiment that demonstrate the technical feasibility of the algorithm obtained. This issue is important, since the analysis of sources shows an almost complete lack of technical implementation of control systems for dynamic objects synthesized using reinforcement learning methods.

Keywords: robust control, reinforcement learning, Q-learning algorithm, dynamic objects, uncertain parameters, convergence of the learning algorithm.

References

1. Sutton R., Barto A. Reinforcement Learning. An Introduction. Second Edition. Cambridge: MIT Press, 2018. vol. 1. no. 1. pp. 9–11.
2. Sutton R.S., Barto A.G., Williams R.J. Reinforcement learning is direct adaptive optimal control. IEEE Control Systems Magazine. 2002. vol. 12(2). pp. 19–22.
3. Pshikhopov V., Medvedev M. Multi-Loop Adaptive Control of Mobile Objects in Solving Trajectory Tracking Tasks. Automation and Remote Control. 2020. vol. 81. pp. 2078–2093. DOI: 10.1134/S0005117920110090.
4. Shih P., Kaul B., Jagannathan S., Drallmeier J. Near Optimal Output-Feedback Control of Nonlinear Discrete-Time Systems in Nonstrict Feedback Form with Application to Engines. IEEE International Joint Conference on Neural Networks. 2007. pp. 396–401.

5. Xu B., Yang C., 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.
6. Mu C., Ni Z., Sun C., 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.
7. 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.
8. 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.
9. Luo B., Yang Y., Liu D., Wu H.-N. Event-Triggered Optimal Control with Performance Guarantees Using Adaptive Dynamic Programming. IEEE Transactions on Neural Networks and Learning Systems. 2019. vol. 31(1). pp. 76–88.
10. Yang X., Xu M., Wei Q. Dynamic Event-Sampled Control of Interconnected Nonlinear Systems Using Reinforcement Learning. IEEE Transactions on Neural Networks and Learning Systems. 2022. vol. 35(1). pp. 923–937. DOI: 10.1109/TNNLS.2022.3178017.
11. Zhang H., Zhao X., Wang H., Zong G., 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. 2022. vol. 35(2). pp. 1559–1571. DOI: 10.1109/TNNLS.2022.3183991.
12. Dong C., Chen L., Dai S.-L. Performance-Guaranteed Adaptive Optimized Control of Intelligent Surface Vehicle Using Reinforcement Learning. IEEE Transactions on Intelligent Vehicles. 2023. vol. 9. no. 2. pp. 3581–3592. DOI: 10.1109/TIV.2023.3338486.
13. Dao P.N., Phung M.H. 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. DOI: 10.1016/j.compeleceng.2024.109870.
14. 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.
15. Thananjeyan B., Balakrishna A., Rosolia U., Li F., McAllister R., Gonzalez J.E., Levine S., Borrelli F., Goldberg K. Safety Augmented Value Estimation From Demonstrations (SAVED): Safe Deep Model-Based RL for Sparse Cost Robotic Tasks. IEEE Robotics and Automation Letters. 2020. vol. 5(2). pp. 3612–3619.
16. Zanon M., Gros S. Safe Reinforcement Learning Using Robust MPC. IEEE Transactions on Automatic Control. 2020. vol. 66(8). pp. 3638–3652. DOI: 10.1109/TAC.2020.3024161.
17. 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. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-19). 2019. vol. 33. no. 01. pp. 3387–3395.
18. 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.
19. 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.

20. Borovik V.S., Shidlovskii S.V. [Reinforcement Learning in Plant Control Systems with Transport Lag]. Optoelectronics. Instrumentation and Data Processing. 2021. vol. 57. pp. 265–272.
21. Galyaev A.A., Medvedev A.I., Nasonov I.A. [Neural network algorithm for intercepting targets moving along known trajectories by a Dubins' car]. Avtomat. i Telemekh. – Automation and telemechanics. 2023. no. 3. pp. 3–21. (In Russ.).
22. Khapkin D.L., Feofilov S.V. [The Method of Synthesis of a Stable Closed-Loop Object Control System with Limiters]. Mekhatronika, Avtomatizatsiya, Upravlenie – Mechatronics, automation, control. 2024. vol. 25(7). pp. 345–353. DOI: 10.17587/mau.25.345-353. (In Russ.).
23. Favorskaya M., Pakhirka A. [Restoration of semantic-based super-resolution aerial images]. Informatics and Automation. 2024. vol. 23(4). pp. 1047–1076. DOI: 10.15622/ia.23.4.5. (In Russ.).
24. Chen K., Ignat'eva S.A., Bogush R.P, Ablameyko S.V. [Re-identification of people in video surveillance systems using deep learning: an analysis of existing methods]. Avtomat. i Telemekh. – Automation and telemechanics. 2023. no. 5. pp. 61–112. DOI: 10.31857/S0005231023050057. (In Russ.).
25. Ponimash Z.A., Potanin M.V. [Method and algorithm for extracting features from digital signals based on neural networks transformer]. Izvestija JuFU. Tehnicheskie nauki – Izvestiya SFEDU. Technical sciences. 2024. no. 6. pp. 52–64. DOI: 10.18522/2311-3103-2024-6-52-64. (In Russ.).
26. Golubinskiy A., Tolstykh A., Tolstykh M. [Automatic Generation of Scientific Articles Abstracts Based on Large Language Models]. Informatics and Automation. 2025. vol. 24(1). pp. 275–301. DOI: 10.15622/ia.24.1.10. (In Russ.).
27. 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. DOI: 10.17587/mau.25.559-567.
28. Gaiduk A.R., Martjanov O.V., Medvedev M.Yu., Pshikhopov V.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. DOI: 10.17587/mau.21.470-479.
29. Zhilov R.A. [Building a PID controller using neural networks]. Izvestija Kabardino-Balkarskogo nauchnogo centra RAN – News of the Kabardino-Balkarian Scientific Center of RAS. 2022. no. 5(109). pp. 38–47. DOI: 10.35330/1991-6639-2022-5-109-38-47. (In Russ.).
30. Karapeev A.N., Kosenko E.Y., Medvedev M.Yu., Pshikhopov V.Kh. [Research of an intelligent adaptive control algorithm based on the reinforcement learning method]. Izvestija JuFU. Tehnicheskie nauki – Izvestiya SFEDU. Engineering Sciences. 2025. no. 2. pp. 162–175. (In Russ.).

Medvedev Mikhail — Ph.D., Dr.Sci., Associate Professor, Leading researcher, R&D Institute of robotics and control systems, Southern Federal University (SFedU). Research interests: planning and control of autonomous robots with a focus on adaptive and robust control, path planning methods, and neural network planning. The number of publications — 260. medvmihal@sfedu.ru; 2, Shevchenko St., 347922, Taganrog, Russia; office phone: +7(863)437-1694.

Pshikhopov Viacheslav — Ph.D., Dr.Sci., Professor, Director, R&D Institute of robotics and control systems, Southern Federal University (SFedU). Research interests: planning and control of autonomous robots and manipulators with a focus on time-optimal control, control in obstructed unmapped environments, and intelligent and group control. The number of

publications — 300. pshichop@rambler.ru; 2, Shevchenko St., 347922, Taganrog, Russia; office phone: +7(863)437-1694.

Evdokimov Igor — Post-graduate student, engineer, R&D Institute of robotics and control systems, Southern Federal University (SFedU). Research interests: planning and control of autonomous robots with a focus on adaptive and robust control. The number of publications — 1. ievdokimov@sfedu.ru; 2, Shevchenko St., 347922, Taganrog, Russia; office phone: +7(863)437-1694.

Acknowledgements. The research was funded by the Russian Science Foundation project No. 25-61-00017, «Intelligent methods of trajectory control of robotic complexes under conditions of parametric and external perturbations», https://rscf.ru/project/25-61-00017/ implemented by the Southern Federal University.