УДК 004.89
DOI: 10.26102/2310-6018/2026.55.4.015

Управление транспортным потоком на основе обучения с подкреплением

Е.И. Минаков✉, Н.И. Хазов

Тульский государственный университет, Тула, Российская Федерация

Резюме. Пробки на дорогах часто возникают из-за неэффективного управления светофорами на перекрестках, то есть из-за того, что их настройки недостаточно адаптированы к конкретным условиям. В настоящее время активно ведутся зарубежные исследования в области применения методов машинного обучения с подкреплением для оптимизации транспортных потоков на перекрестках, что еще раз показывает актуальность проблемы. Перспектива применения обучения с подкреплением заключается в способности управлять динамикой сложных процессов без вмешательства человека. Для поддержания эффективности и безопасности перемещения автомобилей в городских условиях существуют системы, управляющие потоками транспорта с помощью светофорных объектов. В работе рассмотрены существующие типы систем управления транспортными потоками. В ходе проведенного анализа выявлены их положительные и негативные качества. В статье предложена система интеллектуального управления, основанная на принципах обучения с подкреплением, дополненная аппроксиматором, в качестве которого используется нейронная сеть. Архитектура сети представляет собой многослойный перцептрон с двумя скрытыми слоями с ReLU функциями активации. Представлен процесс обучения агента и результаты моделирования системы управления в среде микроскопического моделирования SUMO. Результаты представлены в виде графика динамики обучения агента, тепловых карт перекрестков при имитации движения в час пик и при ДТП до воздействия и после. Предложенная система позволяет увеличить интенсивность движения в сети перекрестков на 40 % и 25 % при движении в час-пик и ДТП соответственно. Помимо этого, отражены дальнейшие перспективы ее развития.

Ключевые слова: транспортный поток, управление дорожным движением, обучение с подкреплением, нейронная сеть, машинное обучение, адаптивное управление.

Для цитирования: Минаков Е.И., Хазов Н.И. Управление транспортным потоком на основе обучения с подкреплением. *Моделирование, оптимизация и информационные технологии.* 2026;14(4). URL: https://moitvivt.ru/ru/journal/article?id=2223 DOI: 10.26102/2310-6018/2026.55.4.015

Traffic flow management based on reinforcement learning

E.I. Minakov✉, N.I. Khazov

Tula State University, Tula, the Russian Federation

Abstract. Traffic jams often occur due to inefficient control of traffic lights at intersections, that is, due to the fact that their settings are not sufficiently adapted to specific conditions. Currently, foreign research is actively underway in the field of applying machine learning methods with reinforcement to optimize traffic flows at intersections, which once again shows the urgency of the problem. The prospect of using reinforcement learning lies in the ability to control the dynamics of complex processes without human intervention. To maintain the efficiency and safety of moving cars in urban environments, there are systems that control traffic flows using traffic lights. The paper considers the existing types of traffic flow management systems. The analysis revealed their positive and negative qualities. The article proposes an intelligent control system based on the principles of reinforcement learning, supplemented by an approximator using a neural network. The network architecture is a multi-layered perceptron, with two hidden layers with ReLU activation functions. The process of agent training and the results of control system modeling in the SUMO microscopic modeling environment are presented. The results

are presented in the form of a graph of the dynamics of agent training, heat maps of intersections when simulating rush hour traffic and in case of an accident before and after exposure. The proposed system makes it possible to increase the traffic intensity in the intersection network by 40% and 25% during rush hour and traffic accidents, respectively. In addition, the future prospects of its development are reflected.

Keywords: traffic flow, traffic management, reinforcement learning, neural networks, machine learning, adaptive management.

For citation: Minakov E.I., Khazov N.I. Traffic flow management based on reinforcement learning. *Modeling, Optimization and Information Technology*. 2026;14(4). (In Russ.). URL: https://moitvivt.ru/ru/journal/article?id=2223 DOI: 10.26102/2310-6018/2026.55.4.015

Введение

Транспортная система относится к числу важнейших элементов, без которых невозможно полноценное развитие городской инфраструктуры и рост социального благосостояния государства. В связи с ростом интенсивности дорожного движения неизбежно увеличивается время поездок и эксплуатационных издержек, а также уровень вредных выбросов. Исходя из этого, обеспечение эффективного и безопасного передвижения по городским улицам является одним из важнейших вопросов [1, 2].

Для минимизации негативных эффектов повсеместно внедряются интеллектуальные системы управления дорожным движением, центральным элементом которых выступают алгоритмы управления светофорными объектами. Несмотря на существенный прогресс, действующие подходы остаются далеки от оптимальности в условиях реальной улично-дорожной сети, характеризующейся стохастичностью спроса, неоднородностью участников движения. Существующие решения охватывают широкий спектр: от фиксированных циклов с заранее заданными планами до адаптивных систем, а также моделей, использующих методы машинного обучения. Алгоритмы на основе обучения с учителем демонстрируют высокие результаты при наличии репрезентативных размеченных данных, позволяя строить краткосрочные прогнозы и поддерживать принятие решений. Однако в задачах управления движением полноценные обучающие выборки часто недоступны или быстро устаревают из-за нестационарности транспортного спроса, изменения дорожной инфраструктуры и внешних дестабилизирующих воздействий погодного, аварийного или иных характеров. Это приводит к снижению устойчивости и эффективности моделей, обученных на исторических данных.

В этих условиях перспективным направлением является интеллектуальное управление дорожным движением, которое основано на машинном обучении [3, 4]. Такой подход позволяет учитывать причинно-следственные связи управляющих воздействий, адаптироваться к изменяющимся условиям и работать при неполной информации.

Вместе с тем сохраняются открытые вопросы, среди которых можно выделить координацию множества перекрестков, разработку более устойчивых систем, ограниченные вычислительные ресурсы. Следовательно, несмотря на развитие методов, задача разработки более эффективных алгоритмов управления светофорными сигналами остается актуальной и требует дальнейших исследований и совершенствования.

Материалы и методы

Фиксированное управление светофорами представляет собой самый простой и традиционный способ регулирования дорожного движения, при котором фазы светофора переключаются по заранее заданному расписанию или фиксированным временным

интервалам. Этот подход часто используется в районах с предсказуемым и стабильным движением, например, в спальных районах или на перекрестках с низкой интенсивностью потока.

Основной принцип фиксированного управления заключается в том, что цикл работы светофора задается на постоянной основе, не меняясь в зависимости от текущего состояния трафика. Такой подход обеспечивает последовательный и предсказуемый ритм движения, но не учитывает изменения транспортного потока. Пример фаз интервалов приведен на Рисунке 1 [5].

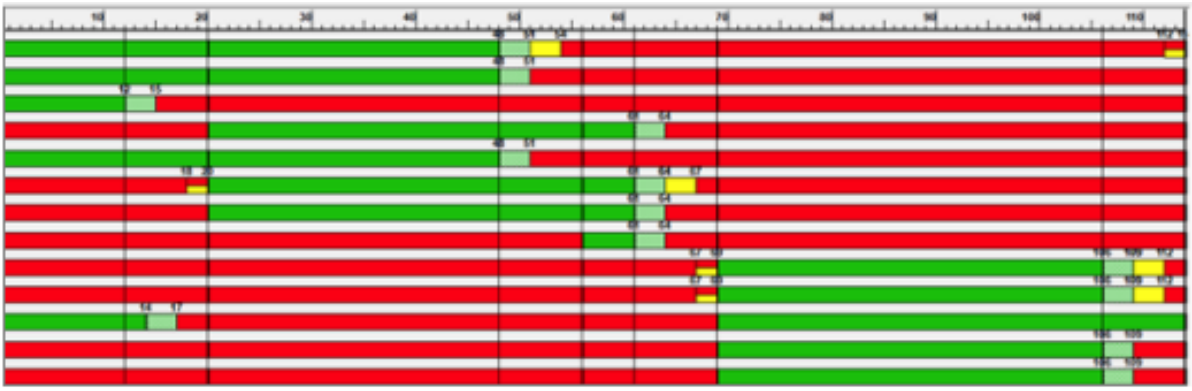

Рисунок 1 – Фазы светофорного регулирования
Figure 1 – Phases of traffic light regulation

Достоинством этих методов является их простая реализация. Основным недостатком является невозможность адаптации светофорного управления к изменяющимся параметрам транспортного потока.

Сценарное управление представляет собой более гибкий подход по сравнению с фиксированным управлением, основанный на заранее определенных сценариях, которые регулируют время фаз светофора в зависимости от времени суток, дня недели или сезонных изменений. В этом случае светофор работает по заранее заданным фиксированным фазам, которые меняются в зависимости от сценария. Эти сценарии могут включать, например, более длинные фазы зеленого сигнала в часы пик или изменение продолжительности фаз в зависимости от сезона. Сценарии заранее запрограммированы, и система не может учесть все изменения в транспортном потоке [6].

Адаптивное управление представляет собой более сложную систему, которая динамически регулирует продолжительность фаз светофора в зависимости от реального состояния транспортного потока. В таких системах используются датчики и камеры для мониторинга интенсивности движения, плотности автомобилей и других факторов, которые могут влиять на движение транспорта. Адаптивное управление может изменять режим работы светофоров в реальном времени, обеспечивая оптимизацию потока и минимизацию заторов [7].

Одним из популярных методов адаптивного управления является использование алгоритмов, которые учитывают данные о текущем уровне загруженности перекрестков и соответствующим образом корректируют время работы светофора [8]. Такие системы значительно эффективнее фиксированных и сценарных, так как они способны реагировать на изменения. Схема адаптивной системы приведена на Рисунке 2.

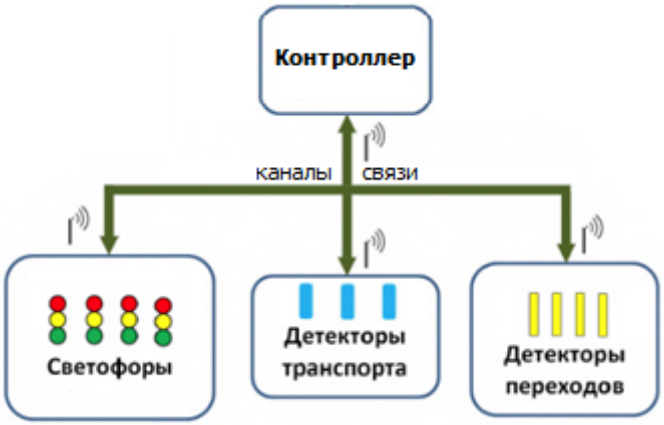

Рисунок 2 – Система адаптивного управления
Figure 2 – Adaptive control system

Интеллектуальное управление. Интеллектуальное управление светофорами, подразумевает использование методов искусственного интеллекта и представляет собой высокотехнологичный подход к регулированию трафика. В отличие от адаптивных систем, интеллектуальное управление основывается на алгоритмах машинного обучения, которые могут обучаться на исторических и реальных данных, чтобы принимать более сложные и обоснованные решения.

Одним из перспективных методов интеллектуального управления, исходя из зарубежного опыта и анализа ряда работ [9, 10], является использование обучения с подкреплением (Reinforcement learning, RL). Это более сложный и современный подход, при котором система самостоятельно учится на основе опыта [11]. В таких системах светофорный агент принимает решения, используя алгоритмы машинного обучения, которые оптимизируют поведение с учетом полученной информации о предыдущих действиях и результатах. Каждый этап работы системы оценивается через механизм вознаграждения, который позволяет корректировать стратегию управления для достижения наилучшего результата. Этот метод особенно полезен в условиях высокой неопределенности и динамичных изменений на дорогах. Преимущество заключается в способности системы оптимизировать свое поведение без явной настройки на каждом перекрестке, основываясь только на накопленных данных [12]. Однако обучение таких систем требует времени и больших вычислительных ресурсов, что может ограничить их внедрение в реальных условиях.

Обучение с подкреплением, обучением с учителем и обучение без учителя – три основные группы машинного обучения, каждая из которых отличается формулировкой задач и обучением алгоритмов по данным. RL – это подход, при котором агент обучается интерпретировать окружающую среду, выполняя действия и отслеживая результаты. За каждое выполненное действие агент может получить положительное, отрицательное или нейтральное вознаграждение в зависимости от предпринятого действия и реакции среды. Схема взаимодействия агента и среды представлена на Рисунке 3.

Рисунок 3 – Схема процесса обучения с подкреплением
Figure 3 – Reinforcement learning process diagram

Основные используемые элементы определяются следующим образом: s_t представляет собой текущее состояние окружающей среды в конкретный момент времени, a_t – действие, которое совершает агент для изменения состояния системы, r_t – положительное или отрицательное вознаграждение, полученное за изменение состояния системы в лучшую или отрицательную сторону соответственно. Величина вознаграждения определяется функцией вознаграждения $R(s_t, a_t, s_{t+1})$, которая не имеет конкретной записи и определяется разработчиками при создании системы. Математическая формализация функции вознаграждения приведена в формуле:

$$R : S \cdot A \cdot S \rightarrow \mathbb{R}. \quad (1)$$

Для любой тройки (текущее состояние s_t, действие a_t, следующее состояние s_{t+1}) функция возвращает действительное числовое значение награды и передает его агенту, в роли которого выступает управляющий элемент, например, светофор.

Пространство состояний *S*, пространство действий *A* и функция вознаграждения задаются средой. Вместе они составляют кортежи (*s*, *a*, *r*), являющиеся основными информационными единицами при описании систем RL.

Помимо перечисленных основных элементов существует функция перехода между состояниями, которая описывается следующими формулами. С помощью нее определяется переход между состояниями s_t и s_{t+1}:

$$s_{t+1} \sim P(s_{t+1} | s_t, a_t), \quad (2)$$

$$P : S \cdot A \rightarrow [0; 1]. \quad (3)$$

Глубокое RL (Deep RL, DRL) является под областью RL. В основе DRL используется аппроксимация функций, где в качестве аппроксимирующего семейства функций используются глубокие нейронные сети.

DRL предполагает обучение агента принятию решений путем взаимодействия со средой для максимизации совокупного вознаграждения с использованием глубоких нейронных сетей. Обучение заключается в нахождении оптимальной стратегии управления $\pi : S \rightarrow A$, которая максимизирует награду *r* [7].

Так как светофорное регулирование может длиться бесконечно, то оптимальное вознаграждение будет стремиться к бесконечности. Для сходимости вводится коэффициент дисконтирования γ, который уменьшает влияние вознаграждения на каждом шаге. Значение коэффициента γ варьируется в диапазоне [0; 1] в зависимости от конфигурации агента. Значения, близкие к 0, указывают на то, что агент предпочитает краткосрочные вознаграждения, тогда как значения, близкие к 1, позволяют агенту

максимизировать возможные более высокие будущие вознаграждения [13]. С учетом коэффициента дисконтирования выражается кумулятивная функция вознаграждения G – это целевая величина, которую агент стремится максимизировать [14].

$$G = \mathbb{E}_\tau \sum_{t=0}^{T} \gamma^t r_t \to max, \quad (4)$$

где $\tau = (s_0, a_0, s_1, a_1)$ – последовательность пар действий и состояний, γ – коэффициент дисконтирования.

Исходя из дискретности пространств состояний и действий, подходящим для решения задачи управления светофорами является Q-learning и его расширение в глубоком обучении DQN. Смысловая основа в Q-learning и DQN одинакова и заключается в решении уравнения Беллмана. Применительно к методам Q-learning и DQN правило обновления Q значений описывается формулами:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha[r_t + \gamma maxQ(s',a')], \quad (5)$$

$$Q(s,a,\Theta) = Q(s,a,\Theta) + \alpha[r_t + \gamma maxQ(s',a',\Theta) - Q(s,a,\Theta)], \quad (6)$$

где $Q(s,a,\Theta)$ – оценка для выполнения действия a в состоянии s, r_t – вознаграждение, полученное за выполнение действия a_t в момент времени t, γ – коэффициент дисконтирования, Θ – веса основной нейронной сети, α – скорость обучения.

Обновление весов Θ происходит по формуле:

$$\Theta \leftarrow \Theta - \alpha \nabla_\Theta L(\Theta), \quad (7)$$

$$L(\Theta) = [r + \gamma max_{a'} Q(s',a',\Theta^-) - Q_\Theta(s,a,\Theta)]^2, \quad (8)$$

где Θ^- – веса целевой нейронной сети.

Главным же из преимуществ интеллектуальных систем такого типа является возможность распространения на несколько управляющих агентов, что может позволить добиться еще более высокой эффективности [15]. На Рисунке 4 приведена обобщенная схема мультиагентной системы.

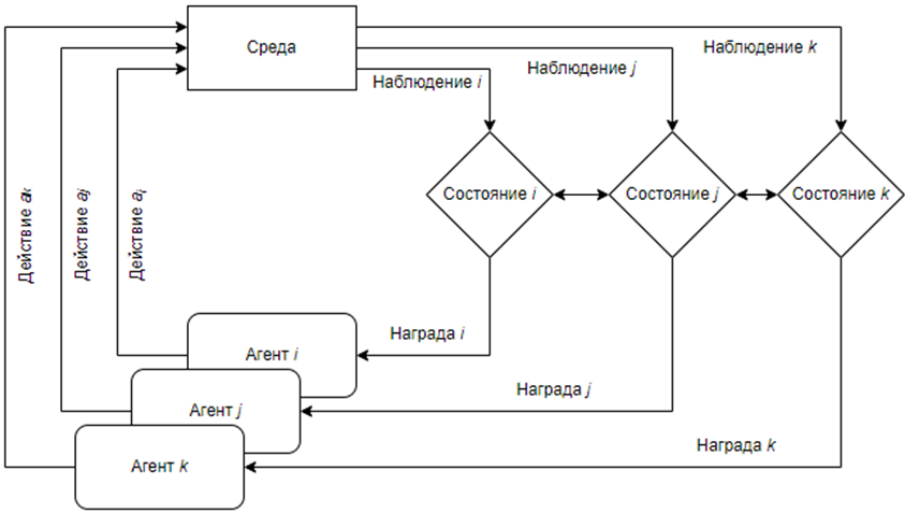

Рисунок 4 – Мультиагентная система
Figure 4 – Multi-agent system

Оценка $Q(s,a,\Theta)$ выполняется по формуле:

$$Q^i_{new}(s^i_t, a^i_t, \Theta) \leftarrow (1-\alpha)Q^i(s^i_t, a^i_t, \Theta) + \alpha\left[r^i_t + \gamma \sum_j n_{ij} max_{a'} Q^i(s^i_t, a^i_t, \Theta)\right], \quad (9)$$

где n_{ij} – вес агента j для агента i.

Результаты

Создана система, где в роли агента выступает пара нейронная сеть и светофор, в роли среды сеть перекрестков, созданных в Simulation Urban Mobility (SUMO), определены вектор состояний s, функция вознаграждений R и возможные действия. Структура полученной системы приведена на Рисунке 5.

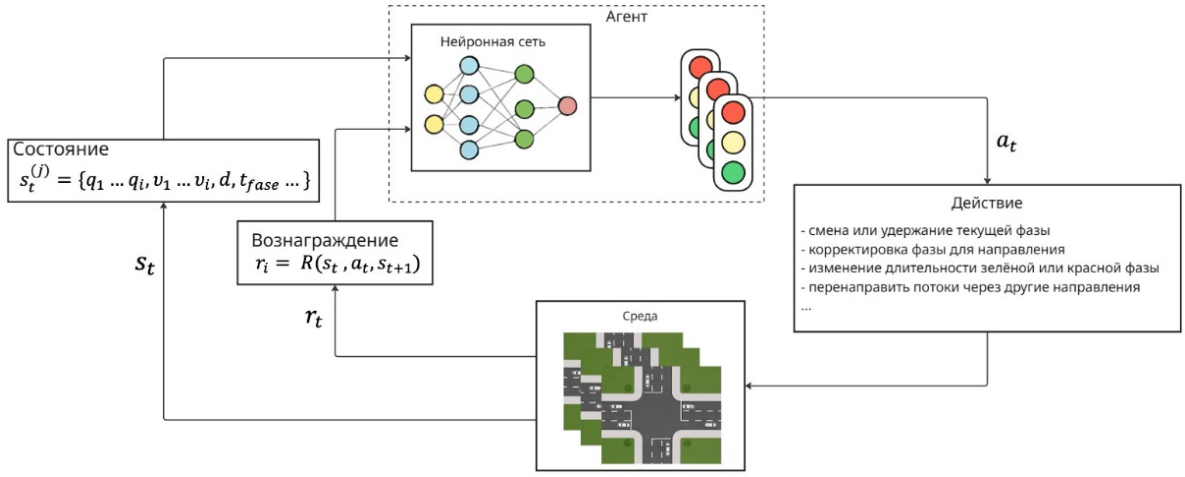

Рисунок 5 – Обучение с подкреплением, применяемое к сети перекрестков
Figure 5 – Reinforcement learning applied to a network of intersections

Возможные действия определены следующим набором: изменение фазы зеленого на −60, −55…60; изменение фазы красного на −60, −55…60; смена фазы; изменение дополнительной секции на −15, −10…15.

Вектор состояний определяется набором данных:

$$s = \{t_g, t_r, m, n, \tau, v, \lambda_n, \lambda_s, \lambda_e, \lambda_w\}, \quad (10)$$

где t_g – длительность зеленой фазы, t_r – длительность красной фазы, m – текущая фаза $\{0;1\}$, n – длина очереди, τ – время ожидания, v – скорость потока, $\lambda_n, \lambda_s, \lambda_e, \lambda_w$ – интенсивность движения от соседнего перекрестка к агенту.

Сеть состоит из входного слоя, который обрабатывает нормализованные состояния дорожной среды, преобразуя состояние окружающей среды в формат, подходящий для обучения. Двух скрытых полносвязных (fully conection) слоев, оснащенных функциями активации ReLU. Выходного слоя, генерирующего Q-значения, соответствующие каждому возможному действию, обеспечивая основу для процесса принятия решения агентом. Архитектура используемой DQN сети представлена на Рисунке 6.

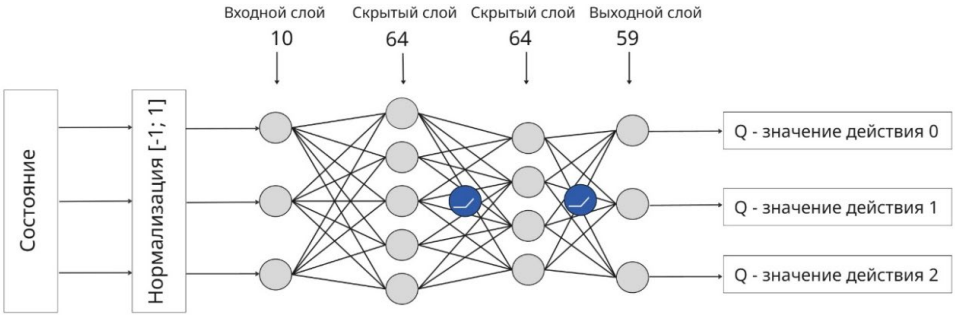

Рисунок 6 – Архитектура используемой нейронной сети
Figure 6 – Architecture of the neural network used

При моделировании была использована сеть перекрестков 3×3, созданная в среде SUMO. В модели предусмотрены режимы час-пик и ДТП. Час-пик имитируется резким увеличением количества автомобилей в сети, режим ДТП создается стоящим на одной из полос автомобилем.

В ходе обучения величина вознаграждений постепенно увеличивается, при этом наблюдаются колебания в некотором интервале значений. Сходимость обучения обусловлена полученной динамикой вознаграждений [14], что отражает успешность выполнения задач и степень адаптации агента к динамике среды [16]. Динамика наград в процессе обучения представлена на Рисунке 7.

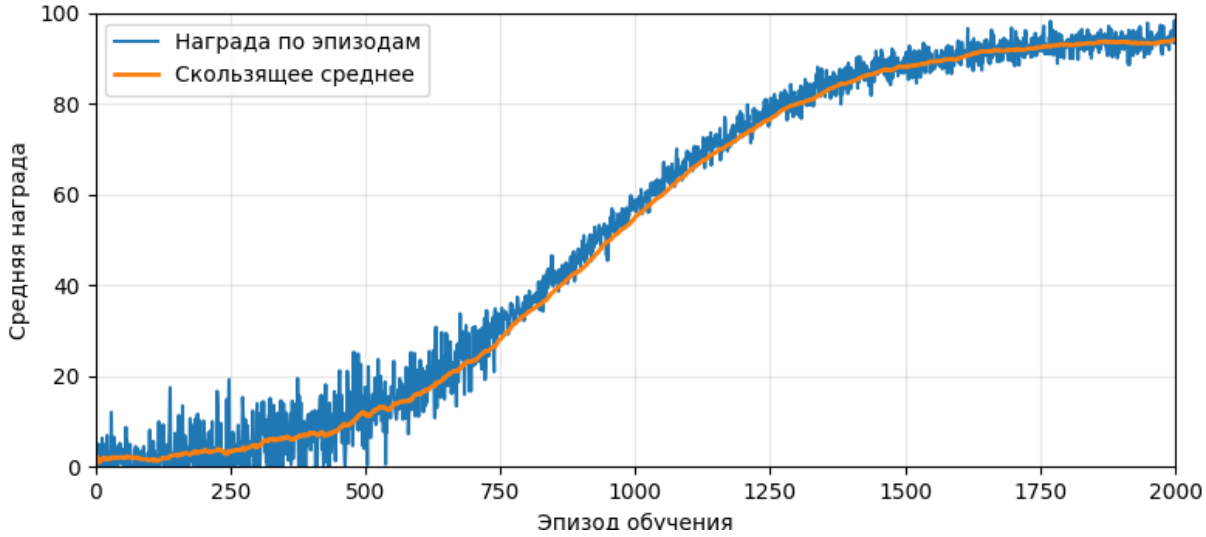

Рисунок 7 – Значение вознаграждения при обучении
Figure 7 – The value of reward in learning

Для обучения использовалась функция вознаграждения, учитывающая изменение интенсивности транспортного потока $\Delta\lambda$ на подъездах к агенту, средней скорости Δv, длины очереди Δn, времени ожидания проезда Δt, количества автомобилей, проезжающих перекресток N:

$$R = \Delta\lambda + \Delta v + \Delta n + 0{,}5 \cdot \Delta t + 0{,}1 \cdot N. \quad (11)$$

Также введены штрафы за частое мигание светофором и долгое не переключение.

Гиперпараметры обучения: эпизоды 2000, длина эпизода 800, скорость обучения 0,001, коэффициент дисконтирования 0,99, размер буфера воспроизведения опыта 10^6, скорость разведки в эпсилон стратегии 1→0,05.

На Рисунке 8 представлены сеть перекрестков с режимом час-пик (слева сверху), где на каждом перекрестке высокая плотность движения и режим ДТП (справа сверху), при котором высокая плотность только на одном перекрестке соответственно. В нижнем ряду на Рисунке 8 показаны сети перекрестков после воздействия на них системы управления.

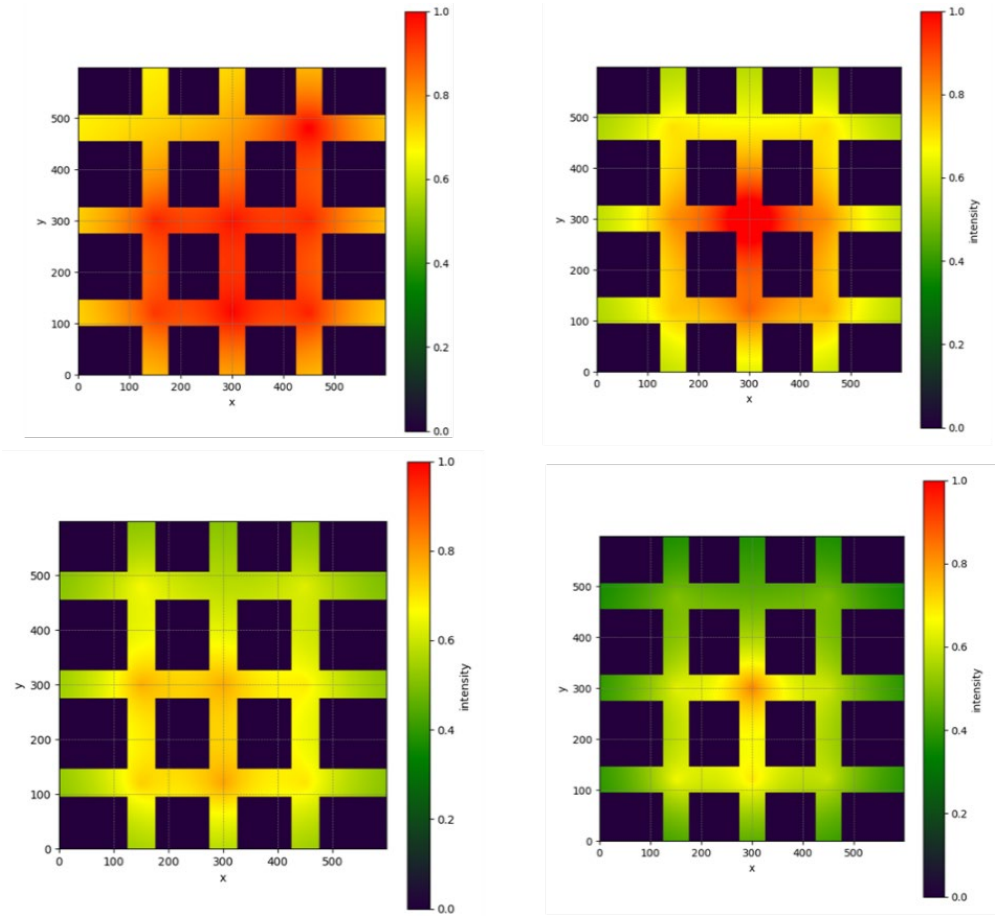

Рисунок 8 – Тепловые карты сети перекрестков
Figure 8 – Heat maps of the intersection network

Исходя из результатов моделирования видно снижение плотности движения на 40 % в режиме час-пик и на 25 % при ДТП. Время нормализации параметров дорожного движения составило 15 минут и 7 минут соответственно. Полученные результаты позволяют сделать вывод о том, что представленный подход является перспективным для управления группой перекрестков. Стоит отметить, что дальнейшее совершенствование системы может быть достигнуто за счет модернизаций обучающей модели, функции вознаграждения, гиперпараметров модели.

Заключение

Современные зарубежные исследования показали, что агенты DRL являются более эффективными в задачах управления транспортным потоком, тем самым их применение потенциально повышает эффективность светофорного регулирования.

В ходе работы проанализирован отечественный опыт внедрения систем управления транспортным потоком. Обозначены достоинства и недостатки рассмотренных методов фиксированной фазы, сценарного, адаптивного и интеллектуального управления.

Рассмотрены основные компоненты обучения с подкреплением S, A, R, P, γ и схема процесса обучения. Предложена система интеллектуального управления транспортным потоком на основе обучения с подкреплением, применяемая к группе перекрестков. Показаны результаты применения обучения, используемая архитектура

нейронной сети и тепловые карты перекрестков до и после применения на них системы управления.

Таким образом, предложенный подход позволяет решить проблему управления транспортным потоком не только при большой загруженности дорог, но и при заторе, образовавшемся из-за ДТП, что делает его перспективным решением для применения в интеллектуальных транспортных системах.

В дальнейших исследованиях необходимо изучить закономерности при изменении функции вознаграждения, возможных действиях и состояниях на различных структурах перекрестков.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Raeisi M., Mahboob A.S. Intelligent Control of Urban Intersection Traffic Light Based on Reinforcement Learning Algorithm. In: *2021 26th International Computer Conference, Computer Society of Iran (CSICC), 03–04 March 2021, Tehran, Iran*. IEEE; 2021. P. 1–5. https://doi.org/10.1109/CSICC52343.2021.9420622
2. Zhou M., Yu Y., Qu X. Development of an Efficient Driving Strategy for Connected and Automated Vehicles at Signalized Intersections: A Reinforcement Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*. 2020;21(1):433–443. https://doi.org/10.1109/TITS.2019.2942014
3. Ducrocq R., Farhi N. Deep Reinforcement Q-Learning for Intelligent Traffic Signal Control with Partial Detection. *International Journal of Intelligent Transportation Systems Research*. 2023;21(1):192–206. https://doi.org/10.1007/s13177-023-00346-4
4. Farazi N.P., Ahamed T., Barua L., Zou B. *Deep Reinforcement Learning and Transportation Research: A Comprehensive Review.* arXiv. URL: https://doi.org/10.48550/arXiv.2010.06187 [Accessed 31st October 2025].
5. Qadri S.Sh.S.M., Gökçe M.A., Öner E. State-of-art review of traffic signal control methods: challenges and opportunities. *European Transport Research Review*. 2020;12(1). https://doi.org/10.1186/s12544-020-00439-1
6. Рутковский В.Н., Капский Д.В. Анализ, разработка и реализация адаптивных алгоритмов (гибкого) светофорного регулирования. *Системный анализ и прикладная информатика*. 2023;(3):4–16. https://doi.org/10.21122/2309-4923-2023-3-4-16

   Rutkovsky V.N., Kapski D.V. Analysis, development and implementation of adaptive algorithms for (flexible) traffic light regulations. *System analysis and applied information science*. 2023;(3):4–16. (In Russ.). https://doi.org/10.21122/2309-4923-2023-3-4-16
7. Агафонов А.А., Ефименко Е.Ю. Сравнение алгоритмов управления сигналами светофоров в крупномасштабном сценарии моделирования движения транспортных средств. В сборнике: *Информационные технологии и нанотехнологии (ИТНТ-2022): Сборник трудов по материалам VIII Международной конференции и молодежной школы: Том 3, 23–27 мая 2022 года, Самара, Россия*. Самара: Самарский национальный исследовательский университет имени академика С.П. Королева; 2022. С. 031382.
8. Агафонов А.А., Юмаганов А.С., Мясников В.В. Адаптивное управление дорожными сигналами на основе нейросетевого прогноза максимального взвешенного потока. *Автометрия*. 2022;58(5):85–97. https://doi.org/10.15372/AUT20220510

   Agafonov A.A., Yumaganov A.S., Myasnikov V.V. Adaptive traffic signal control based on neural network prediction of weighted traffic flow. *Optoelectronics, Instrumentation and Data Processing.* 2022;58(5):503–513. https://doi.org/10.3103/s8756699022050016

9. Dake D.K., Gadze J.D., Klogo G.S., Nunoo-Mensah H. Traffic Engineering in Software-defined Networks using Reinforcement Learning: A Review. *International Journal of Advanced Computer Science and Applications*. 2021;12(5):330–345.
10. Fadila J.N., Wahab N.H.A., Alshammari A., et al. Comprehensive review of smart urban traffic management in the context of the fourth industrial revolution. *IEEE Access.* 2024;12:196866–196886. https://doi.org/10.1109/access.2024.3509572
11. Liang X., Du X., Wang G., Han Zh. A Deep Reinforcement Learning Network for Traffic Light Cycle Control. *IEEE Transactions on Vehicular Technology*. 2019;68(2):1243–1253. https://doi.org/10.1109/TVT.2018.2890726
12. Kunjir M., Chawla S. *Offline Reinforcement Learning for Road Traffic Control.* arXiv. URL: https://doi.org/10.48550/arXiv.2201.02381 [Accessed 20th October 2025].
13. Tan K.L., Sharma A., Sarkar S. Robust Deep Reinforcement Learning for Traffic Signal Control. *Journal of Big Data Analytics in Transportation*. 2020;2(3):263–274. https://doi.org/10.1007/s42421-020-00029-6
14. Орлова Е.В. Обучение с подкреплением как технология искусственного интеллекта для решения социально-экономических задач: оценка производительности алгоритмов. *π-Economy*. 2023;16(5):38–50. https://doi.org/10.18721/JE.16503
Orlova E.V. Reinforcement learning as an artificial intelligence technology to solve socio-economic problems: algorithms performance assessment. *π-Economy*. 2023;16(5):38–50. (In Russ.). https://doi.org/10.18721/JE.16503
15. Saadi A., Abghour N., Chiba Z., Moussaid Kh., Ali S. A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control. *Journal of Big Data*. 2025;12(1). https://doi.org/10.1186/s40537-025-01104-x
16. Корчагин А.П. Гибридная система обучения агентов с использованием A2C и эволюционных стратегий. *Моделирование, оптимизация и информационные технологии*. 2025;13(3). https://doi.org/10.26102/2310-6018/2025.50.3.029
Korchagin A.P. Hybrid agent training system using A2C and evolutionary strategies. *Modeling, Optimization and Information Technology*. 2025;13(3). (In Russ.). https://doi.org/10.26102/2310-6018/2025.50.3.029

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Минаков Евгений Иванович, доктор технических наук, профессор, профессор кафедры «Радиоэлектроника», Тульский государственный университет, Тула, Российская Федерация.
e-mail: eminakov@bk.ru
ORCID: 0000-0002-3234-9907

Evgeniy I. Minakov, Doctor of Engineering Sciences, Professor, Professor at the Department of Radio Electronics, Tula State University, Tula, the Russian Federation.

Хазов Никита Ильич, аспирант, Тульский государственный университет, Тула, Российская Федерация.
e-mail: nikita.hazov511@yandex.ru

Nikita I. Khazov, Postgraduate, Tula State University, Tula, the Russian Federation.

Статья поступила в редакцию 13.02.2026; одобрена после рецензирования 17.04.2026; принята к публикации 22.04.2026.

The article was submitted 13.02.2026; approved after reviewing 17.04.2026; accepted for publication 22.04.2026.