УДК 004.89

АДАПТИВНОЕ УПРАВЛЕНИЕ ДОРОЖНЫМИ СИГНАЛАМИ НА ОСНОВЕ НЕЙРОСЕТЕВОГО ПРОГНОЗА МАКСИМАЛЬНОГО ВЗВЕШЕННОГО ПОТОКА

© **А. А. Агафонов, А. С. Юмаганов, В. В. Мясников**

Самарский национальный исследовательский университет им. академика С. П. Королёва, 443086, г. Самара, Московское шоссе, 34
E-mail: ant.agafonov@gmail.ru

Предлагается двухэтапный метод адаптивного управления сигналами светофоров, основанный на оценке прогнозируемого взвешенного потока транспортных средств, проходящих через перекрёсток. На первом этапе оценивается время прохождения перекрёстка каждым транспортным средством с использованием модели искусственной нейронной сети и формируется оценка прогнозируемого потока транспортных средств через перекрёсток для заданной фазы светофорного цикла. На втором этапе формируется оценка взвешенного потока, которая учитывает время ожидания транспортных средств. Предлагаемый метод выбора фазы светофора основывается на максимизации взвешенного транспортного потока. Результаты экспериментальных исследований позволяют сделать вывод о преимуществе предложенного подхода в сравнении с классическими подходами и современными методами управления сигналами светофоров на основе обучения с подкреплением.

Ключевые слова: управление сигналами светофоров, искусственная нейронная сеть, обучение с подкреплением, подключённые транспортные средства.

DOI: 10.15372/AUT20220510

Введение. Идеи цифровой экономики оказывают влияние на все аспекты современной жизни. Существенное влияние они оказывают на транспорт. В частности, де-факто стандартом в настоящее время являются транспортные средства (ТС) с электронными помощниками, которые контролируют полосу движения, учитывают дорожные знаки и правила дорожного движения, предупреждают столкновения и т. п. Указанные нововведения являются лишь одной из составных частей создаваемых интеллектуальных транспортных систем (ИТС) в [1]. Не меньшее, а, скорее, большее значение при построении ИТС играют решения, позволяющие обеспечивать комплексное управление транспортным движением на территории целого населённого пункта и/или региона.

Существующие тенденции развития транспортных систем показывают постоянный рост дорожных заторов, что приводит к значительному увеличению транспортных расходов (времени в пути и топлива) и выбросов в окружающую среду [2]. Хотя проектирование с нуля городской/областной транспортной инфраструктуры с ИТС позволяет решать комплексные задачи управления максимально эффективно (например, проектирование мест проживания и работы с обеспечением управления масштабами транспортного спроса и предложения на территориях), дороговизна, а порой и принципиальная невозможность изменения существующей транспортной топологии и жилой инфраструктуры делает значительно более важным решение прагматичной задачи — оптимизации движения ТС в рамках существующей инфраструктуры. Как следствие, становится актуальной задача управления транспортными потоками путём адаптивного светофорного регулирования, поскольку её решение допускает быстрое и относительно недорогое внедрение в существующую транспортную инфраструктуру при значительном (как правило, в разы) росте её эффективности (увеличению пропускной способности, снижению затрат топлива и

т. п.). Развитие информационно-коммуникационных технологий, «Интернета вещей», подключённых и автономных транспортных средств, сетей VANET приводит к увеличению объёма данных, которые могут использоваться для решения задачи управления, а также делает актуальным разработку новых методов решения этой задачи.

В работе рассматривается метод адаптивного управления фазами светофоров, основанный на выборе фазы, максимизирующей «взвешенный» поток транспортных средств, проходящих через перекрёсток. Предлагается модификация детерминированного метода управления, представленного в [3]:

— приведён метод оценки времени прохождения перекрёстка отдельным транспортным средством с использованием модели искусственной нейронной сети;

— представлена модификация метода управления для учёта времени ожидания транспортных средств на перекрёстках и оценки взвешенного транспортного потока;

— проведены экспериментальные исследования предложенного и базовых алгоритмов управления на разработанном крупномасштабном сценарии моделирования движения транспортных средств в системе моделирования SUMO (Simulation of Urban Mobility).

Таким образом, в данной работе рассматривается задача адаптивного управления дорожными сигналами светофоров на основании данных движения транспортных средств. Задача управления заключается в выборе фазы светофорного цикла, максимизирующей пропускную способность транспортной сети. Научная новизна заключается в разработке метода адаптивного управления, основанного на нейросетевой оценке взвешенного прогнозируемого потока транспортных средств, проходящих через перекрёсток за выбранную фазу светофорного цикла. Предлагаемый метод выбора фазы светофора основывается на максимизации оценки взвешенного транспортного потока, учитывающей время ожидания транспортных средств на перекрёстке.

Современное состояние исследований. Ранние работы [3–5], посвящённые решению задачи управления сигналами светофоров, рассматривают детерминированные подходы к выбору светофорного цикла и фазы на перекрёстке. Современные исследования сосредоточены на применении методов анализа данных и методов машинного обучения к решению транспортных задач. В частности, широкое распространение получило использование метода обучения с подкреплением (RL — reinforcement learning) для решения задачи управления сигналами светофоров [3, 4–10]. Классические оптимизационные методы управления в основном рассматриваются как базовые для сравнения с современными методами.

Классификация систем управления сигналами светофора представлена в [11], в которой рассматривались статистические алгоритмы, алгоритмы нечёткой логики и обучения с подкреплением, генетические и гибридные алгоритмы, используемые в системах управления светофорами. В [10] представлен обзор методов и практик, которые могут быть применены к решению проблемы управления сигналами светофора на перекрёстках, и описано влияние развития информационных и коммуникационных технологий на проблему управления. В [8] рассматривалось применение методов глубокого обучения с подкреплением в интеллектуальных транспортных системах. Представлены различные формулировки проблемы, параметры RL-моделей, среды моделирования, а также рассмотрены открытые вопросы. Комплексный обзор RL-моделей и алгоритмов, включающий описание состояния среды, действия и вознаграждения как части постановки RL-задачи дан в [6].

Рассмотрим подробнее детерминированные и основанные на RL-методах подходы к решению задачи управления светофорами.

Детерминированные подходы к управлению сигналами светофоров. Алгоритмы управления светофорами разрабатываются с конца 50-х гг. XX в., с тех пор как в [12] был предложен метод расчёта длины цикла светофора и фазового разделения для одиночного (изолированного) перекрёстка. Стратегии координированного управления с

фиксированным временем, направленные на оптимизацию смещений фаз сигналов светофора для смежных перекрёстков, разработаны в [13, 14]. Однако эти стратегии не реагируют на динамические изменения трафика и применимы только в условиях ненасыщенного трафика [5].

Адаптированные системы управления способны регулировать фазы светофорного цикла в зависимости от текущей дорожной ситуации на перекрёстке (например, переключаться на следующую фазу, оптимизируя длину очереди ТС на каждой полосе, или сохранять текущую фазу при обнаружении непрерывного транспортного потока) [15]. Самоорганизующееся управление светофором [16] представляет собой тип полнофункционального управления с дополнительными правилами адаптивного управления.

В [17] было введено понятие «давление», обозначающее разницу между количеством ТС на входящей и соответствующей исходящей полосах движения. Предлагаемый метод направлен на минимизацию нагрузки фаз, которые уравновешивают длину очереди между перекрёстками и максимизируют пропускную способность сети. В [18] на основе экспериментального анализа было показано, что система управления на базе алгоритма MaxPressure продемонстрировала наилучшую производительность в небольшой синтетической грид-сети по сравнению с RL-алгоритмами.

В [3] предложен алгоритм управления светофором, разработанный на модели детерминированного прогнозирования, который оценивает прогнозируемый транспортный поток для каждой фазы и выбирает фазу с максимальным потоком.

Подходы, основанные на обучении с подкреплением. Методы обучения с подкреплением — это класс методов машинного обучения, в которых агент настраивает свою функцию политики в ходе взаимодействия с окружающей средой, чтобы максимизировать численно определённое вознаграждение. Процесс обучения RL-агента на каждом шаге может быть представлен следующим образом:

— агент определяет состояние среды;

— на основе наблюдаемого состояния агент выбирает действие в соответствии с настроенной к данному моменту функцией политики (стратегии), что приводит к переходу в новое состояние среды;

— агент получает немедленное вознаграждение от среды и обновляет свою функцию политики так, чтобы максимизировать общее вознаграждение. Результатом такого обучения является оптимальная политика, т. е. оптимальное правило выбора действия по текущему состоянию.

Ключевой вопрос в формулировке задачи управления сигналами светофоров как проблемы обучения с подкреплением заключается в том, как определить понятия состояния, действия и вознаграждения. Для описания среды и определения состояния исследователи использовали различные характеристики транспортного потока, в том числе: длину очереди ТС, время ожидания, положение транспортных средств, среднюю скорость, продолжительность фазы [4, 19, 20]. Функции вознаграждения обычно определяются как взвешенная сумма различных факторов, которые можно измерить после совершения действия: длина очереди ТС, пропускная способность, время ожидания, давление [21–23].

RL-методы также можно разделить на методы, основанные на полезности и политике. Первые изучают функцию полезности действия-состояния (Q-функцию), т. е. «насколько хорошо действие» при текущем состоянии с точки зрения ожидаемого вознаграждения. Основными RL-алгоритмами этого класса являются Q-обучение [9, 24], глубокое Q-обучение [20], двойное Q-обучение [19, 25]. Вторые напрямую изучают функцию политики, которая определяет вероятность выполнения определённого действия в определённом состоянии [26–28]. Методы Actor-Critic сочетают в себе оба подхода: Actor изучает политику, которая определяет действие, выполняемое агентом, а Critic оценивает успешность этого действия, используя функцию полезности [29–32].

Несмотря на популярность RL-методов, они могут быть очень чувствительными к гиперпараметрам и нестабильно работать в сложных сценариях [18].

Постановка задачи управления и базовый алгоритм. Основные определения. Введём основные определения для постановки задачи управления.

Под перекрёстком понимается место пересечения, примыкания или разветвления дорог на одном уровне, ограниченное воображаемыми линиями, которые соединяют соответственно противоположные наиболее удалённые от центра перекрёстка начала закруглений проезжих частей. Каждая из проезжих частей делится разметкой на полосы движения, часть из которых отвечает за въезд на перекрёсток, а другая — за выезд. При этом разметка определяет допустимые направления движения, а сигналы светофора(ов) — возможность перемещения по соответствующей полосе в текущий момент времени. На рис. 1 показан типовой перекрёсток как пересечение двух дорог с разделёнными проезжими частями, каждая из которых имеет по три полосы движения.

Далее, фаза светофора — это множество непротиворечивых сигналов светофора. Фаза может выбираться в рамках некоторой предопределённой последовательности фаз, так называемого светофорного цикла, или может выбираться произвольно. При этом, если выбор фазы светофорного регулирования происходит на основании данных движения транспортных средств, то речь идёт о системах адаптивного управления дорожными сигналами [3].

Задача адаптивного управления сигналами/фазами светофоров заключается в выборе фазы светофорного цикла, максимизирующей пропускную способность транспортной сети.

Метод на основе детерминированной прогнозной модели. В работе [3] представлен адаптивный подход к решению задачи управления, заключающийся в выборе фазы светофорного цикла, которая максимизирует прогнозируемый поток транспортных средств, проходящих через перекрёсток за рассматриваемый временной интервал.

Метод максимального потока MaxFlow.
Входные данные: $\tau_{\min}$, t_p, P
Выходные данные: $phase$
if $t_p < \tau_{\min}$ **then**
    $t_p = t_p + 1$
else
    $t_p = 0$
    $phase = \arg\max\left(\{PredFlow\,(p) \text{ for } p \text{ in } P\}\right)$
end if

В алгоритме $\tau_{\min}$ — допустимый минимальный интервал переключения фаз, t_p — длительность текущей активной фазы светофора, P — множество фаз, $phase$ — следующая фаза.

Ключевым шагом алгоритма является расчёт прогнозируемого транспортного потока $PredFlow(phase)$ через перекрёсток для выбранной фазы светофорного цикла $phase$. Для оценки транспортного потока используется информация о параметрах движения ТС (в том числе подключённых и автономных ТС):

$$PredFlow(phase) = \sum_{l \in L_{phase}^{income}} \sum_{c \in C_l} I(t(c) < \tau_{\min}),$$

где $t(c)$ — оценка времени, необходимого для достижения перекрёстка транспортным средством c из множества ТС C_l, движущихся по определённой полосе l из множества полос для выезда L_{phase}^{income} для фазы $phase$, $I(val)$ — индикатор, который возвращает значение 1, если $val = True$, и значение 0 в противном случае.

Для оценки времени движения $t(c)$ в [3] использовалась детерминированная прогнозная модель, использующая набор аналитических закономерностей движения ТС, известных

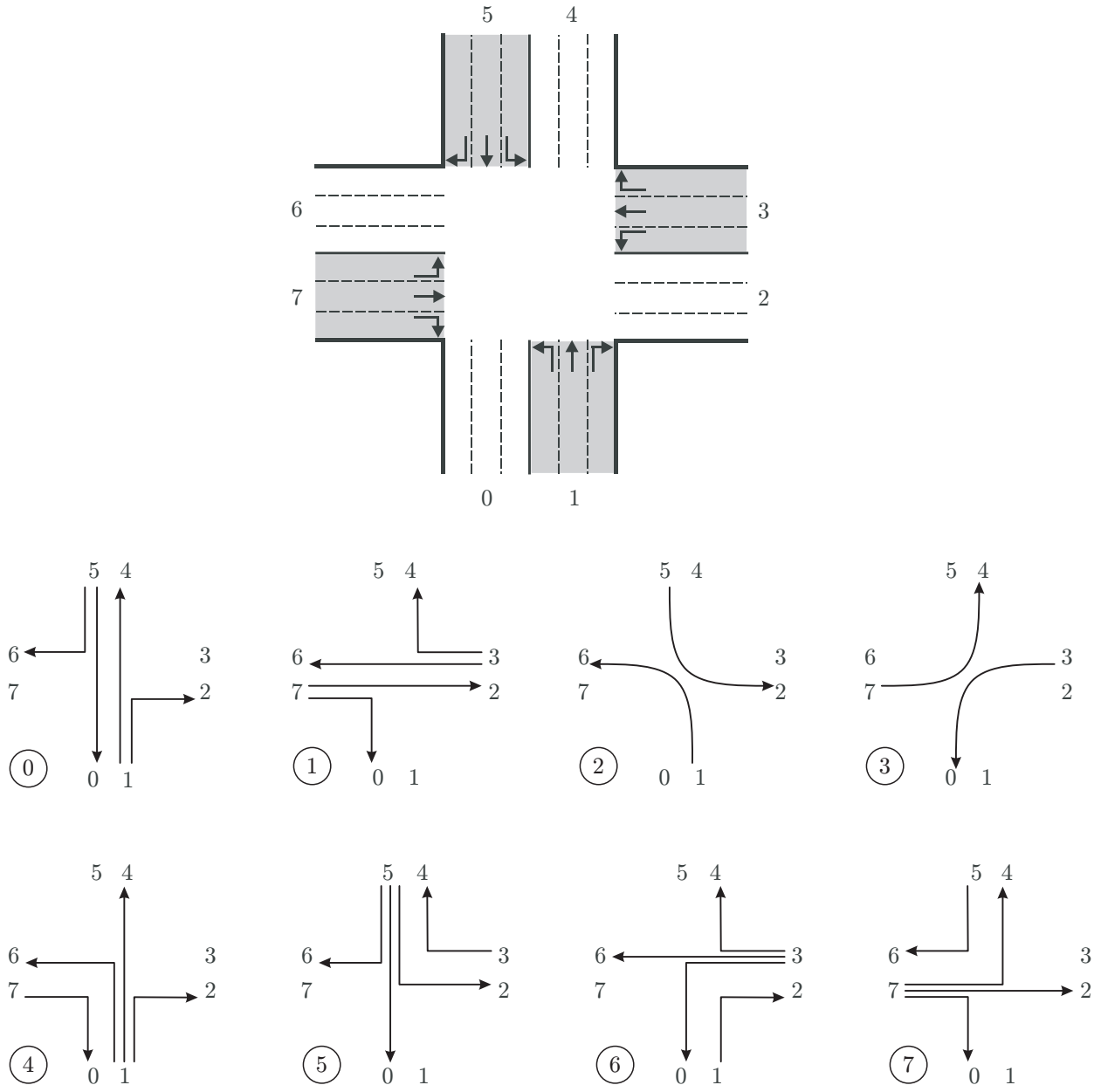

Рис. 1. Пример типового перекрёстка: восемь проезжих частей, каждая с тремя полосами движения

из классического курса физики/механики. Входными данными функции $t(c)$ являлись два параметра: текущая скорость движения ТС $v = v(c)$ и расстояние до перекрёстка $S = S(c)$.

Метод максимального взвешенного потока. Предлагаемый метод максимального потока обладает рядом недостатков:

— использует простую модель оценки времени прохождения перекрёстка;

— не учитывает время ожидания (простоя), которое ТС проводит на перекрёстке.

Учёт времени ожидания ТС необходим для предотвращения ситуаций, при которых время ожидания транспортных средств на второстепенной дороге (с малым транспортным потоком) будет увеличиваться до бесконечности.

Для исправления указанных недостатков предлагается использовать так называемый

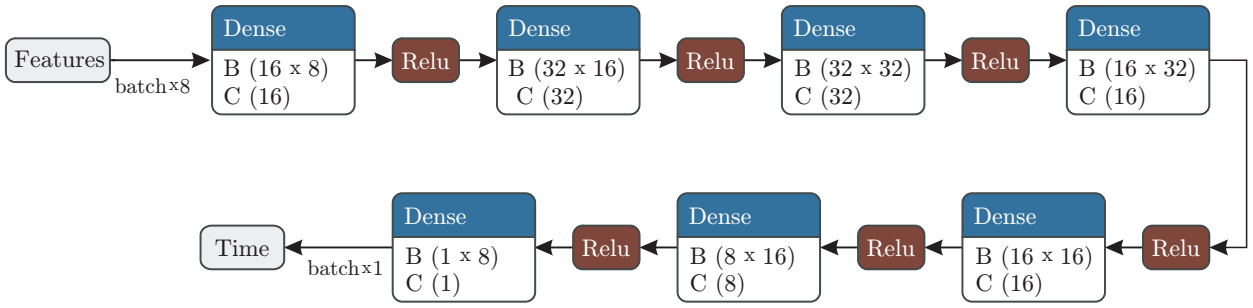

Рис. 2. Архитектура нейронной сети для оценки времени прохождения перекрёстка

взвешенный транспортный поток, который учитывает время ожидания ТС следующим образом:

$$PWFlow(phase) = \sum_{l \in L^{income}_{phase}} \sum_{c \in C_l} \eta(c,l) I(t^{dnn}(c) < \tau_{\min}).$$

Коэффициент $\eta(c,l)$ необходим для корректировки «веса» ТС c в транспортном потоке и зависит от времени ожидания $delay\,(c,l)$ (в секундах) этого ТС на полосе l на перекрёстке:

$$\eta(c,l) = 1 + \alpha\, delay\,(c,l),$$

где α — эмпирически выбираемый коэффициент. В экспериментах (на основе предварительно проведённого анализа) мы полагаем $\alpha = 0{,}01$.

Для оценки времени прохождения перекрёстка ТС c вместо детерминированной прогнозной модели $t(c)$ предлагается использовать модель на основе искусственной нейронной сети $t^{dnn}(c)$. В качестве входных параметров модели берутся следующие характеристики, прямо или косвенно описывающие транспортную ситуацию на текущем и смежном дорожных сегментах, а также движение рассматриваемого ТС c:

— положение ТС $S = S(c)$ (расстояние от текущей позиции ТС до перекрёстка);

— скорость движения ТС $v = v(c)$;

— ускорение ТС $a = a(c)$;

— максимально разрешённая скорость движения $v_{\max}$;

— число ТС перед рассматриваемым ТС c до перекрёстка $n = n(c)$;

— тип k ожидаемого перестроения на перекрёстке (0 — прямо, 1 — направо, 2 — налево, 3 — разворот);

— скорость v_0 и положение S_0 ближайшего ТС на полосе выезда.

Нейронная сеть состоит из 7 полносвязных слоёв. Архитектура используемой нейронной сети с указанием количества нейронов каждого слоя показана на рис. 2.

Следующая фаза светофорного цикла выбирается аналогично базовому методу: выбирается фаза, для которой взвешенный транспортный поток максимален, т. е. для выбора фазы используется выражение

$$phase = \arg\max\left(\{PWFlow(p) \;\; \text{for } p \;\; \text{in } P\}\right).$$

В экспериментальных исследованиях будем ссылаться на предложенный метод как на метод максимального взвешенного потока MaxPWFlow.

Таблица 1

Параметры сценариев движения

Сценарий	Светофоры	Перекрёстки	Сегменты	Поездки в сети
Cologne-8	8	78	149	1740
Cologne-316	316	2928	5808	13570

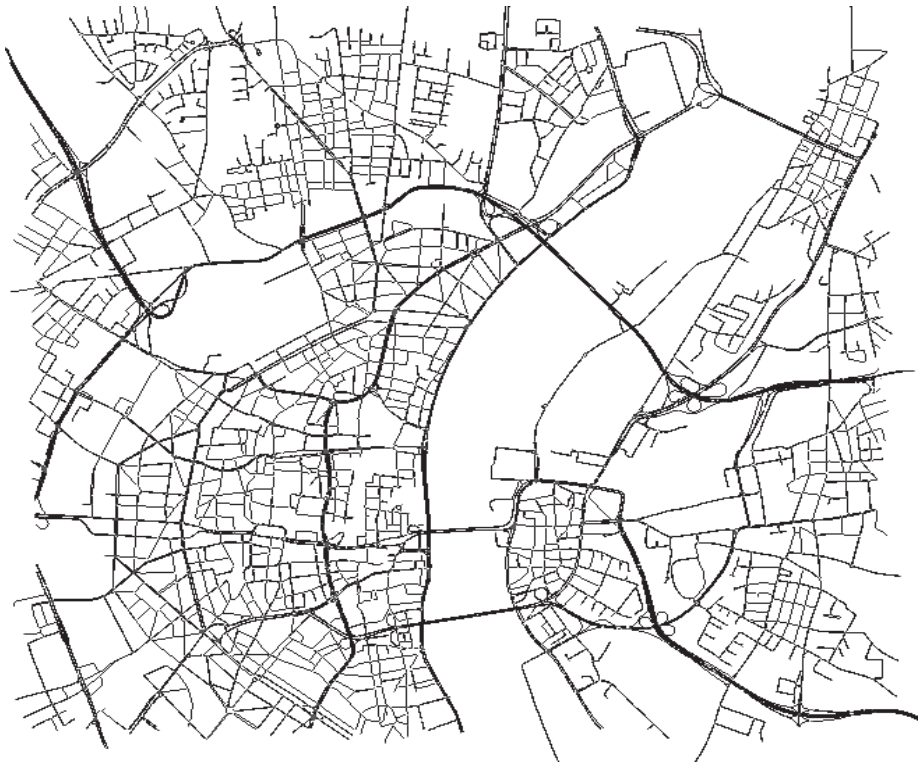

Рис. 3. Дорожная сеть сценария моделирования движения Cologne-316

Экспериментальные исследования. Экспериментальное исследование нацелено на сравнение эффективности предложенного метода MaxPWFlow с базовым методом MaxFlow [3], классическим алгоритмом MaxPressure [17] и современными методами на основе обучения с подкреплением. Экспериментальные исследования разработанных алгоритмов проводились в системе моделирования движения транспортных средств SUMO [33].

Сценарии движения. Для оценки эффективности алгоритмов использовались сценарии имитационного моделирования, основанные на сценарии движения SUMO «TAPAS Cologne» [34]:

1. Cologne-8 — сценарий, выполняющий моделирование движения транспортных средств в области транспортной сети малого размера [35].

2. Cologne-316 — созданный в данной работе крупномасштабный сценарий движения транспортных средств.

Параметры сценариев представлены в табл. 1.

Дорожная сеть сценария Cologne-316 показана на рис. 3.

Каждый сценарий определяется дорожной сетью и поездками, совершаемыми в сети (созданы на основе исходных данных о загруженности дорог). RL-модели обучались для каждого сценария на данных из нескольких эпизодов. Эпизоды различаются начальным расположением транспортных средств на сегментах сети, временем начала движения и

маршрутами движения. В течение одного эпизода выполняется моделирование всех поездок в сценарии.

Сравнение всех моделей проводилось на одинаковых данных на выборке из десяти эпизодов (начальные положения транспортных средств на сегментах сети, время начала движения и маршруты были одинаковыми для всех моделей в одном конкретном эпизоде).

Алгоритмы для сравнения. В качестве базовых алгоритмов для оценки эффективности предложенного метода использовались классические алгоритмы и методы на основе обучения с подкреплением:

— метод максимального потока MaxFlow [3];

— MaxPressure [17];

— IDQN — независимый алгоритм глубокого Q-обучения, основанный на свёрточной нейронной сети [35]. Состояние каждого агента характеризуют следующие величины: текущая фаза светофора, длина очереди для каждой входящей полосы, количество приближающихся транспортных средств и сумма скоростей приближающихся транспортных средств для каждой входящей полосы. В качестве функции вознаграждения использовалось общее время ожидания с отрицательным знаком.

— IPPO — это алгоритм обучения с подкреплением для оптимизации политики [27]. Пространство наблюдения и функция вознаграждения использовались те же, как и в методе IDQN.

— A2C [31].

Для сравнения на основе анализа современного состояния исследований были выбраны методы различных классов, показавшие хорошие результаты в решении задачи адап-

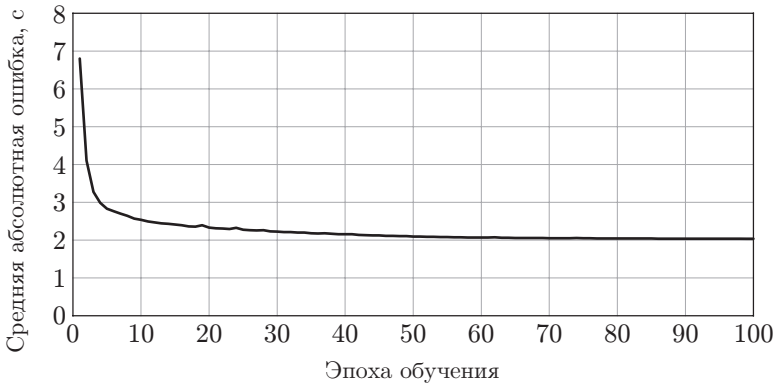

Рис. 4. Сходимость модели ИНС для прогнозирования времени движения

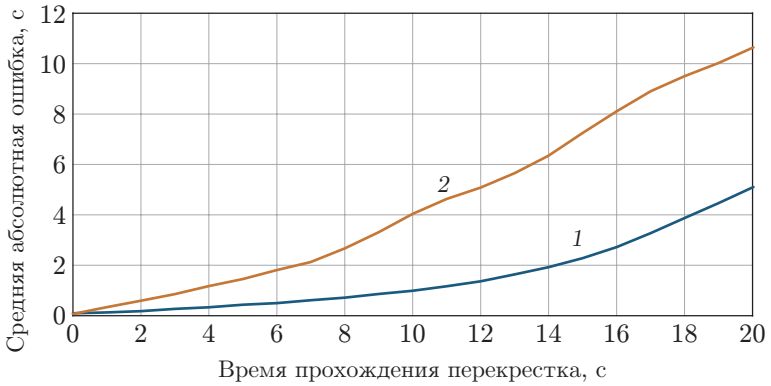

Рис. 5. Средняя абсолютная ошибка прогнозирования времени прохождения перекрёстка: *1* — ИНС, *2* — детерминированная модель

тивного управления сигналами светофоров. RL-алгоритмы были реализованы на языке программирования Python с использованием библиотеки PyTorch и библиотеки глубокого обучения с подкреплением PFRL [36].

Результаты экспериментов. На первом этапе исследований оценивалось качество прогнозирования времени прохождения перекрёстков с использованием детерминированной прогнозной модели $t(c)$ и применением модели искусственной нейронной сети (ИНС) $t^{dnn}(c)$. Для обучения модели ИНС реализовывались данные о движении транспортных средств на выборке из 70 эпизодов, для валидации и тестирования модели брались данные, полученные при моделировании 15 эпизодов движения. График зависимости средней абсолютной ошибки от эпохи обучения показан на рис. 4. Полученная кривая обучения подтверждает сходимость модели.

График зависимости средней абсолютной ошибки прогноза от времени прохождения перекрёстка (так называемого горизонта прогноза) для детерминированной прогнозной модели и модели ИНС показан на рис. 5. Из представленного графика видно, что предложенный алгоритм прогнозирования позволяет значительно (в 2–3 раза) снизить ошибку прогноза.

На втором этапе экспериментального исследования оценивалась сходимость RL-алгоритмов. Кривые обучения RL-алгоритмов на сценариях Cologne-8 (рис. 6, *a*) и Cologne-316 (рис. 6, *b*) подтверждают сходимость алгоритмов. Алгоритмы MaxPWFlow, MaxFlow и MaxPressure — необучаемые алгоритмы с точки зрения правила принятия решения о выборе фазы, поэтому для них подобные графики не представлены.

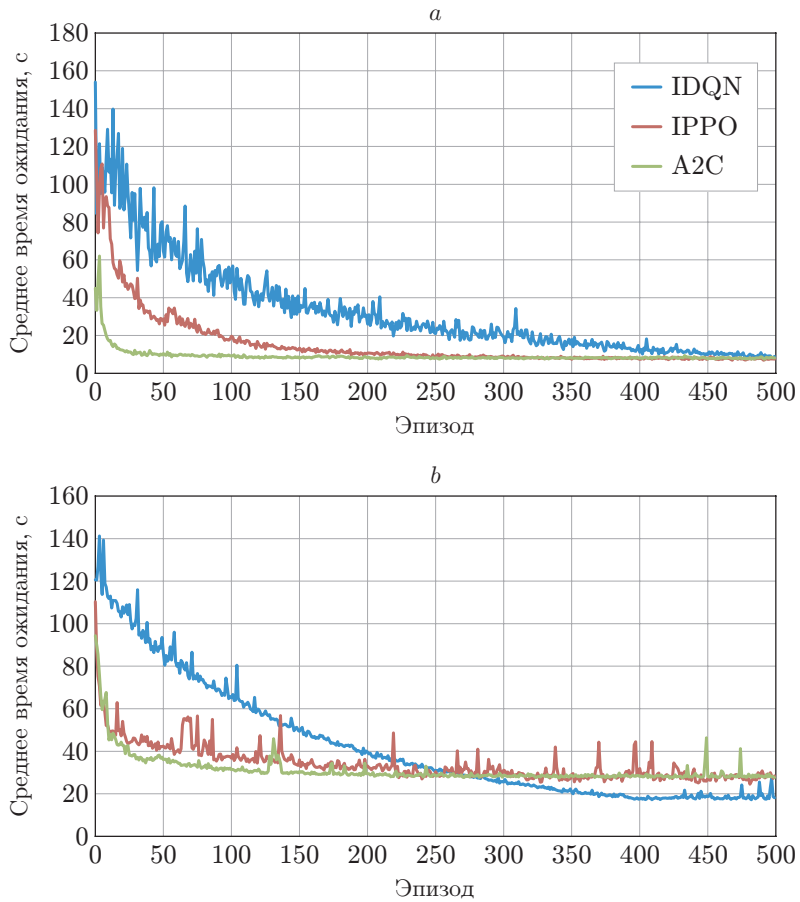

Рис. 6. Сходимость алгоритмов на сценариях: *a* — Cologne-8, *b* — Cologne-316

Таблица 2

Сравнение эффективности управления транспортными потоками

Модель	Среднее время ожидания		Среднее время движения	
	Cologne-8	Cologne-316	Cologne-8	Cologne-316
IDQN	4,14	15,3	89,62	329,6
IPPO	4,69	31,93	90,83	347,72
A2C	7,24	26,32	94,91	347,81
MaxFlow	3,46	17,58	88,86	334,99
MaxPressure	9,16	20,63	93,82	337,91
MaxPWFlow	**3,2**	**14,43**	**88,03**	**328,09**

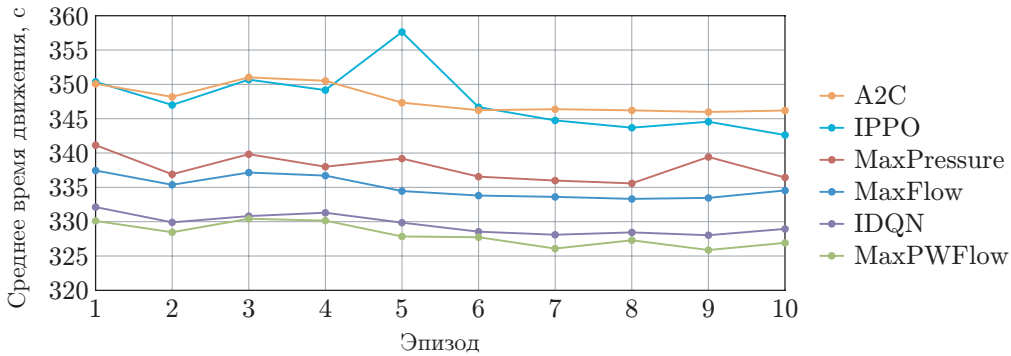

Рис. 7. Среднее время движения в каждом эпизоде сценария Cologne-316

Далее проводилось сравнение эффективности алгоритмов на тестовой выборке из десяти эпизодов по двум метрикам: среднее время ожидания и среднее время движения. Метрика среднего времени ожидания (в секундах) показывает среднее время, которое транспортное средство провело без движения на перекрёстке при движении по маршруту, среднее время движения (в секундах) — это среднее время, затрачиваемое транспортными средствами на совершение поездки от начального пункта их маршрута к месту назначения. Результаты сравнения эффективности алгоритмов представлены в табл. 2.

Экспериментальные исследования показывают, что предложенный алгоритм MaxPWFlow превосходит детерминированный метод MaxPressure и современные RL-алгоритмы по критериям среднего времени ожидания и среднего времени движения на рассматриваемых сценариях. Кроме того, правило принятия решения не требует обучения, что также является преимуществом представленного алгоритма.

В заключительной части исследований оценивалось время ожидания и время движения отдельно по каждому тестовому эпизоду. Результат сравнения по времени ожидания на сценарии Cologne-316 показан на рис. 7.

Графики показывают, что представленный метод превосходит остальные алгоритмы в каждом из эпизодов.

Заключение. В данной работе представлен метод адаптивного управления сигналами светофоров, основанный на оценке взвешенного прогнозируемого потока транспортных средств, проходящих через перекрёсток за выбранную фазу светофорного цикла. Для оценки времени прохождения транспортным средством перекрёстка предложено использовать модель искусственной нейронной сети. Предлагаемый метод выбора фазы светофора основывается на максимизации оценки взвешенного транспортного потока, учитывающей время ожидания транспортных средств на перекрёстке.

Экспериментальные исследования, проведённые на двух сценариях моделирования движения транспортных средств, подтверждают эффективность предложенного метода. Показано его преимущество по критериям среднего времени движения и среднего времени ожидания в сравнении с классическим методом MaxPressure и современными алгоритмами на основе машинного обучения с подкреплением.

В качестве перспективного направления можно отметить разработку более сложных алгоритмов оценки взвешенного транспортного потока, учитывающих время движения и время ожидания транспортных средств на перекрёстке.

Финансирование. Работа выполнена при поддержке Российского научного фонда (проект № 21-11-00321, https://rscf.ru/en/project/21-11-00321/).

СПИСОК ЛИТЕРАТУРЫ

1. **Указ** Президента Российской Федерации от 01.12.2016 г. № 642 // Президент России. URL: http://kremlin.ru/acts/bank/41449 (дата обращения: 19.09.2022).
2. **Schrank D., Albert L., Eisele B., Lomax T.** 2021 Urban Mobility Report. URL: https://trid.trb.org/view/1862637 (дата обращения: 20.05.2022).
3. **Мясников В. В., Агафонов А. А., Юмаганов А. С.** Детерминированная прогнозная модель управления сигналами светофоров в интеллектуальных транспортных и геоинформационных системах// Компьютерная оптика. 2021. **45**, № 6. С. 917–925.
4. **Wei H., Zheng G., Gayah V., Li Z.** A Survey on Traffic Signal Control Methods. arXiv:1904.08117 [cs, stat], Jan. 2020. URL: http://arxiv.org/abs/1904.08117 (дата обращения: 08.06.2022).
5. **Papageorgiou M., Diakaki C., Dinopoulou V. et al.** Review of road traffic control strategies // Proceeding of the IEEE. 2003. **91**, N 12. P. 2043–2065.
6. **Yau K.-L., Qadir J., Khoo H. et al.** A survey on Reinforcement learning models and algorithms for traffic signal control // ACM Comput. Surv. 2017. **50**, N 3. P. 1–38.
7. **Greguric M., Vujic M., Alexopoulos C., Miletic M.** Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data // Appl. Sci. 2020. **10**, N 11. P. 4011.
8. **Haydari A., Yilmaz Y.** Deep reinforcement learning for intelligent transportation systems: A Survey // IEEE Trans. Intell. Transport. Syst. 2022. **23**, N 1. P. 11–32.
9. **Abdulhai B., Pringle R., Karakoulas G.** Reinforcement learning for true adaptive traffic signal control // Journ. Transport. Eng. 2003. **129**. P. 278–285.
10. **Eom M., Kim B.-I.** The traffic signal control problem for intersections: A review. // Europ. Transport. Res. Rev. 2022. **12**, N 1. P. 50.
11. **Savithramma R., Sumathi R.** Road traffic signal control and management system: A survey // Proc. of the 3rd Int. Conf. on Intelligent Sustainable Systems (ICISS 2020). Palladam, India, 3–5 Dec., 2020. P. 104–110.
12. **Webster F. V.** Traffic Signal Settings. H.M. Stationery Office. London, 1958. 56 p.
13. **Little J., Kelson M., Gartner N.** MAXBAND: A program for setting signals on arteries and triangular networks // Transport. Res. Record Journ. Transportation Res. Board. 1981. **795**. P. 40–46.
14. **Li M.-T., Gan A.** Signal timing optimization for oversaturated networks using TRANSYT-7F // Transport. Res. Record. 1999. N 1683. P. 118–126.
15. **El-Tantawy S., Abdulhai B.** An agent-based learning towards decentralized and coordinated traffic signal control // Proc. of the IEEE Conf. on Intelligent Transportation Systems (ITSC). Madeira Island, Portugal, 19–22 Sept., 2010. P. 665–670.

16. **Cools S.-B., Gershenson C., D'Hooghe B.** Self-organizing traffic lights: A realistic simulation // Adv. Inform. and Knowledge Process. 2013. N 9781447151128. P. 45–55.
17. **Varaiya P.** The Max-Pressure Controller for Arbitrary Networks of Signalized Intersections // Advances in Dynamic Network Modeling in Complex Transportation Systems. Ser. Complex Networks and Dynamic Systems /Eds. by S. V. Ukkusuri and K. Ozbay. New York: Springer, 2013. P. 27–66.
18. **Genders W., Razavi S.** An Open-Source Framework for Adaptive Traffic Signal Control, arXiv:1909.00395 [cs, eess], Sep. 2019. URL: http://arxiv.org/abs/1909.00395 (дата обращения: 08.06.2022).
19. **Agafonov A., Myasnikov V.** Traffic signal control: A double q-learning approach // Proc. of the 16th Conf. on Computer Science and Information Systems (FedCSIS 2021). Sofia, Bulgaria, 2–5 Sept., 2021. P. 365–369.
20. **Wei H., Zheng G., Yao H., Li Z.** IntelliLight: A reinforcement learning approach for intelligent traffic light control // Proc. of the 24th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining. London, United Kingdom, 19–23 Aug., 2018. P. 2496–2505.
21. **Zhang Z., Yang J., Zha H.** Integrating independent and centralized multi-agent reinforcement learning for traffic signal network optimization. Sep. 2019. URL: https://arxiv.org/abs/1909.10651v1 (дата обращения 08.06.2022).
22. **Zheng G., Xiong Y., Zang X. et al.** Learning Phase Competition for Traffic Signal Control, arXiv:1905.04722 [cs, stat], May 2019. URL: http://arxiv.org/abs/1905.04722 (дата обращения 08.06.2022).
23. **Wei H., Chen C., Zheng G. et al.** PressLight: Learning max pressure control to coordinate traffic signals in arterial network // Proc. of the 25th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2019). Alaska, USA, 4–8 Aug., 2019. P. 1290–1298.
24. **Chin Y. K., Lee L. K., Bolong N. et al.** Exploring Q-learning optimization in traffic signal timing plan management // Proc. of the 3rd Int. Conf. on Computational Intelligence, Communication Systems and Networks. Bali, Indonesia, 26–28 Jul., 2011. P. 269–274.
25. **Gu J., Fang Y., Sheng Z., Wen P.** Double deep Q-network with a dual-agent for traffic signal control // Appl. Sci. 2020. **10**, N 5. P. 1622.
26. **Schulman J., Wolski F., Dhariwal P. et al.** Proximal Policy Optimization Algorithms, arXiv:1707.06347 [cs], Aug. 2017. URL: http://arxiv.org/abs/1707.06347 (дата обращения 08.06.2022).
27. **Ault J., Hanna J. P., Sharon G.** Learning an Interpretable Traffic Signal Control Policy, arXiv:1912.11023 [cs, stat], Feb. 2020. URL: http://arxiv.org/abs/1912.11023 (дата обращения: 08.06.2022).
28. **Li Y., He J., Gao Y.** Intelligent traffic signal control with deep reinforcement learning at single intersection // Proc. of the 7th Int. Conf. on Computing and Artificial Intelligence. Tianjin, China, 23–26 April, 2021. P. 399–406.
29. **Aslani M., Mesgari M. S., Wiering M.** Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events // Transport. Res. Pt C: Emerging Technologies. 2017. **85**. P. 732–752.
30. **Yang S., Yang B., Kang Z., Deng L.** IHG-MA: Inductive heterogeneous graph multi-agent reinforcement learning for multi-intersection traffic signal control // Neural Networks. 2021. **139**. P. 265–277.
31. **Wu Y., Mansimov E., Liao S. et al.** Scalable trust-region method for deep reinforcement learning using Kronecker-factored approximation, arXiv:1708.05144 [cs], Aug. 2017, arXiv: 1708.05144. URL: http://arxiv.org/abs/1708.05144 (дата обращения: 08.06.2022).

32. **Haarnoja T., Zhou A., Hartikainen K. et al.** Soft Actor-Critic Algorithms and Applications, arXiv:1812.05905 [cs, stat], Jan. 2019, arXiv: 1812.05905. URL: http://arxiv.org/abs/1812.05905 (дата обращения: 08.06.2022).
33. **Traffic Lights** - SUMO Documentation. URL: https://sumo.dlr.de/docs/Simulation/Traffic_Lights.html (дата обращения: 08.06.2022).
34. **TAPAS Cologne** - SUMO Documentation. URL: https://sumo.dlr.de/docs/Data/Scenarios/TAPASCologne.html (дата обращения: 08.06.2022).
35. **RESCO**, Oct. 2021, original-date: 2021-06-07T17:31:48Z. URL: https://github.com/Pi-Star-Lab/RESCO (дата обращения: 08.06.2022).
36. **PFRL**: a PyTorch-based deep reinforcement learning library. URL: https://github.com/pfnet/pfrl (дата обращения: 08.06.2022).

Поступила в редакцию 13.07.2022
После доработки 01.09.2022
Принята к публикации 02.09.2022