РАЗРАБОТКА АДАПТИВНОЙ СИСТЕМЫ УПРАВЛЕНИЯ СВЕТОФОРНЫМИ ОБЪЕКТАМИ С ИСПОЛЬЗОВАНИЕМ МАРКОВСКИХ ПРОЦЕССОВ ПРИНЯТИЯ РЕШЕНИЙ[1]

Тисленко Т.И.[2], **Семенова Д.В.**[3]

(Сибирский федеральный университет, г. Красноярск)

В статье представлены результаты разработки программного комплекса MARLIN24, предназначенного для адаптивного управления светофорными объектами. Структура комплекса включает модуль адаптивного управления светофорными объектами, модуль симуляции движения транспорта и модуль валидации. Математическая модель процесса управления светофорными объектами — управляемый марковский процесс с конечным числом действий и состояний. Задача поиска эффективного управления в целях уменьшения суммарного времени нахождения транспортных средств на детектируемых участках дорожной сети сводится к задаче мультиагентного обучения с подкреплением (Multiagent Reinforcement Learning for Integrated Network, MARLIN). Для поиска решения задачи MARLIN был построен алгоритм Q-обучения. Структура программного комплекса также включает модуль микросимуляции транспортных потоков «разумный водитель» (Intelligent Driver Model, IDM). Для имитации информации о реальной дорожной обстановке, поступающей в виде показаний оптического датчика, в модуле валидации используются многомерные распределения, полученные с помощью применения копул Маршала-Олкина к маргинальным эмпирических распределениям для временных отметок срабатывания оптического датчика. Для построения маргинальных распределений использовались данные об интенсивности движения через детектируемые участки дорожной сети в период с 2019 по 2020 год.

Ключевые слова: имитационное моделирование транспортных потоков, модель IDM, модель MOBIL, копулы Маршалла-Олкина, машинное обучение с подкреплением.

Одной из задач, решаемых в ходе реализации транспортной стратегии России на 2035 год [1], является увеличение про-

[1] *Работа поддержана Красноярским математическим центром, финансируемым Минобрнауки РФ (Соглашение 075-02-2024-1429).*

[2] *Тисленко Тимофей Иванович, аспирант(timtisko@mail.ru).*

[3] *Семенова Дарья Владиславовна, к.ф.-м.н., доцент (DVSemenova@sfu-kras.ru).*

пускной способности и скоростных параметров дорожной инфраструктуры до уровня лучших мировых достижений. При этом необходимо учитывать общесоциальные целевые ориентиры транспортной стратегии: подвижность населения, снижение аварийности, рисков и угроз безопасности по видам транспорта, снижение доли транспорта в загрязнении окружающей среды. Таким образом, актуальна проблема моделирования транспортных потоков и оценки эффективности управления светофорными объектами для наиболее нагруженных участков дорожной сети.

Системы, управляющие светофорными объектами, подразделяют на те, которые корректируют сигналы светофоров в реальном времени и реагируют на текущую дорожную обстановку — АСУДД (адаптивные системы управления дорожным движением) и неадаптивные — те, которые работают согласно фиксированному плану управления. В таблице 1 представлены наиболее известные АСУДД, которые были разработаны в различные временные периоды и для различных условий движения.

Таблица 1. Распространенные модели АСУДД

Критерий	**UTCS-1**	**SCOOT**	**OPAC**	**АСУДД «Микро»**	**MARLIN**
город	Вашингтон	Лондон	Арлингтон, Тускон	Красноярск	Торонто
временной период	1970е	1995	1983,1989	1993	2010
длительность фаз	фиксированная		переменная		
оптимизация	офлайн	онлайн			
предсказание	нет	есть			
устройство	централизованная		децентрализованная		
основные ограничения	постоянный сбор данных	сенсоры далеко	только для 8 фаз	находится в разработке	«ПРОКЛЯТИЕ РАЗМЕРНОСТИ»

Одна из ранних систем управления городским дорожным движением была UTCS-1 (Urban Traffic Control System) [2]. Система UTCS-1 использовала фиксированные сигнальные планы, которые менялись в зависимости от времени суток (утренний, дневной, вечерний планы). Для стабильной работы системы UTCS-1 необходимо регулярно вручную корректировать сигнальные пла-

ны, что является существенным недостатком. На данный момент считается устаревшей, так как более современные системы UTCS-2, UTCS-3, SCOOT позволяют гибко управлять дорожным движением в реальном времени.

Система SCOOT (Split Cycle Offset Optimization Technique) [3] анализирует данные о дорожной обстановке и корректирует светофорные сигналы, чтобы предотвратить образование заторов до их появления. Для работы данной системы требуется установка плотной сети индукционных петель, камер и других датчиков движения на расстоянии не менее сорока метров и не более двухста метров до регулируемых перекрёстков. Централизованное управление SCOOT направлено на устранение «эффекта волны». Однако продолжительность каждого сигнала светофора (Split) не указывает на изменение времени активной фазы в реальном времени.

Система OPAC (Optimized Policies for Adaptive Control) [4] — это адаптивная система управления светофорами, схожая по назначению с системой SCOOT, но использующая другой подход для оптимизации транспортных потоков. OPAC, разработанная в США, предназначена для улучшения дорожной ситуации в реальном времени путем адаптации фаз светофоров в зависимости от условий трафика. В системе OPAC существует ограничение в восемь фаз для каждого светофора. Это ограничение связано с практическими соображениями, поскольку каждая фаза представляет собой отдельное направление движения или определённую комбинацию разрешённых манёвров на перекрёстке (например, движение прямо, поворот налево или направо). Эти ограничения заложены аппаратно, то есть на уровне контроллеров, и не могут быть изменены конечным пользователем.

Автоматизированная система управления дорожным движением (АСУДД) «Микро» [5] — наиболее широко используемая в России система, успешно применяемая в следующих регионах: Красноярский край, Иркутская область, Белгородская область, Воронежская область, Хабаровский край, Московская область. АСУДД «Микро» является децентрализованной системой и под-

держивает до шести GPRS-серверов, которые позволяют подключить до 250 перекрёстков. Детекторы АСУДД «Микро» работают на расстоянии до ста метров. Основным недостатком системы является тот факт, что реализация адаптивных алгоритмов находится в стадии разработки.

В работах [6, 20, 21, 22, 23] для адаптивного управления светофорными объектами было предложено использовать метод мультиагентного обучения с подкреплением. Данный подход получил название MARLIN (Multiagent Reinforcement Learning for Integrated Network) и был успешно применен в современной АСУДД в Канаде. Целью обучения с подкреплением является сокращение времени проезда транспортных средств через выбранные участки дорожной сети. Управление светофорными объектами считается эффективным, если транспортные средства находятся на детектируемых участках менее двух циклов. Для работы требуется установка камер на расстоянии менее ста метров от стоп-линий. Агенты (светофоры) могут работать без информации о полной дорожной обстановке и управлять движением децентрализованно. Перечисленные особенности можно отнести к достоинствам подхода на основе MARLIN. Существенным ограничением является рост вычислительной сложности при увеличении обзора агента, известный как «проклятие размерности».

Структура статьи следующая. В параграфах **??** и **??** проблема управления светофорными объектами сводится к задаче мультиагентного обучения с подкреплением для одного и нескольких агентов соответственно. дописать текст из файлика в оверлифе

1. Общее описание программного комплекса MARLIN24

Фокус наших исследований сосредоточен на разработке программных и математических инструментов для адаптивного управления сетью светофорных объектов участков дорожной сети города Красноярска. В настоящей работе представлен новый программный комплекс MARLIN24, реализующий часть методов новейших АСУДД. Промежуточные результаты по разработке ком-

плекса и эксперименты представлены в [20, 21, 22, 23]. В статье в качестве примера приведены две модели участков дорожной сети города Красноярска: модель двух перекрестков (пр. Свободный–ул. Лесопарковая, пр. Свободный–ул. Годенко) и модель перекрестков микрорайона Покровский. Оценка эффективности применяемого управления и валидация полученных результатов осуществлялись с помощью симуляционных экспериментов на основе статистических данных с оптических детекторов за 2018 и 2019 года.

Общая структура комплекса MARLIN24 приведена на рисунке 1. Комплекс состоит из четырех модулей: симуляции, валидации, адаптивного управления светофорными объектами, визуализации. Модуль визуализации является опциональным. Его функционал и используемые компоненты будут описаны в остальных модулях.

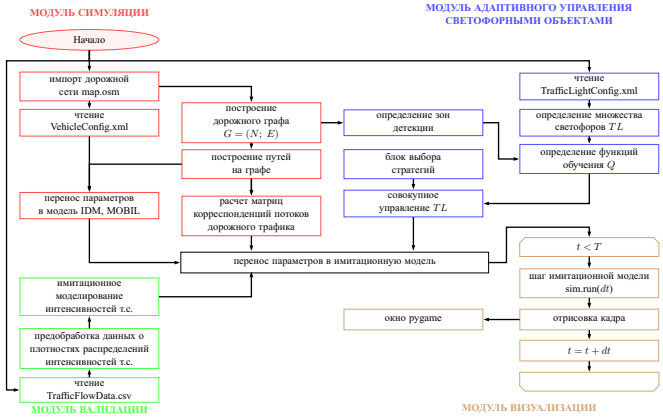


Рис. 1. Структура программного комплекса MARLIN24

2. Модуль симуляции

Модуль симуляции предназначен для оценки эффективности выбранного управления светофорными объектами. В модуле реализованы две микросимуляционные модели движения, описы-

вающая перемещения транспортных средств по дорожной сети. Первая модель IDM (Intelligent Driver Model) реализует движение транспортных средств по однополосной прямой дороге [7]. Вторая модель — MOBIL (Microscopic Optimally Balanced Intersection Lanes), описывает перестроение на многополосной дороге [8]. В процессе симуляции движения транспортных средств учитывается управление фазами и циклами светофорных объектов, а также предусмотрено имитирование показаний оптических датчиков и работы светофоров.

2.1. Модели движения транспортных средств

Для описания движения отдельных транспортных средств используется модель IDM [7]. Данная модель позволяет учитывать следующие параметры: минимальная безопасная дистанция, максимальная разрешенная скорость и коэффициент торможения транспортных средств. В таблице 2 представлены значения параметров для модели IDM, используемые далее в статье.

Таблица 2. Основные обозначения для модели IDM

Символ	Значение
i	номер транспортного средства
$s_{0,i}$	минимальная безопасная дистанция для т.с. i
$v_{0,i}$	максимальная желательная скорость i
δ	компонента «гладкости» ускорения
T_i	время реакции i-го водителя
a_i	ускорение i
b_i	коэффициент торможения i
s^*	желаемое расстояние между i и $i-1$

Модель IDM относится к классу моделей движения за лидером. Все транспортные средства рассматриваются как индивидуальные сущности, обладающие характеристиками и поведением. Водители транспортных средств регулируют скорость в зависимости от расстояния до впереди идущих и их скорости. На рисунке 2 представлено взаимное расположение и характеристики текущего автомобиля, расположенного в i-ой позиции, и $(i-1)$-го автомобиля, находящегося перед ним.

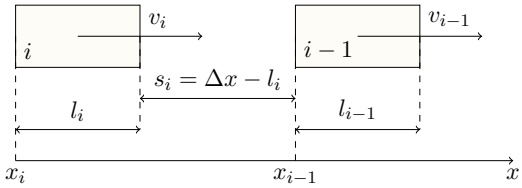


Рис. 2. Иллюстрация модели IDM

Система уравнений, описывающих текущую скорость i-го автомобиля и расстояние до $(i-1)$-го автомобиля в модели IDM в классических обозначениях имеет вид:

$$(1)\quad \begin{cases} \dfrac{dv_i}{dt} = \underbrace{a_i\left(1 - \left(\dfrac{v_i}{v_{0,i}}\right)^{\delta}\right)}_{a_{\text{free road}}} - \underbrace{a_i\left(\dfrac{s^*(v_i, \Delta v_i)}{s_i}\right)^2}_{a_{\text{deceleration}}}, \\ s^*(v_i, \Delta v_i) = s_{0,i} + v_i T_i + \dfrac{v_i \Delta v_i}{2\sqrt{a_i b_i}}, \end{cases}$$

где $a_{\text{free road}}$ обозначает ускорение, возникающее на свободной дороге, $a_{\text{deceleration}}$ — замедление, возникающее при сближении транспортных средств.

При имитационном моделировании для нахождения значений скорости и ускорения использовались формулы, вытекающие из численного метода «пристрелки» [9]:

$$\begin{cases} \dfrac{dv}{dt}(t) = a_{\text{free road}}(t) + a_{\text{deceleration}}(t), \\ v(t + \Delta t) = v(t) + \dfrac{dv}{dt}(t)\Delta t, \\ x(t + \Delta t) = x(t) + v(t)\Delta t + \dfrac{1}{2}\dfrac{dv}{dt}(t)(\Delta t)^2, \\ s(t + \Delta t) = x_i(t + \Delta t) - x(t + \Delta t) - l_i. \end{cases}$$

Шаг симуляции dt выбирается как шаг по времени при численном решении системы (1).

Существенным ограничением модели IDM является ее применимость только к однополосному движению. Одним из способов расширить ее применимость к многополосным дорожным сетям является введение алгоритмов, описывающих перестроение транспортных средств. В работе используется модель MOBIL [8].

В основе модели MOBIL лежит идея о том, что водители принимают решения о перестроении и изменении скорости движения из соображений проходимости и безопасности. Конкретное изменение полосы движения, например с правой полосы движения на левую полосу, как показано на рисунке 3, зависит, как правило, от двух следующих транспортных средств на текущей полосе движения и соответственно на целевой полосе движения.

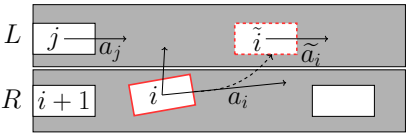


Рис. 3. Модель смены полосы MOBIL

Стимул для перестроения есть, если после первого фиктивного перестроения с правой полосы R на левую полосу L сумма собственного ускорения согласно модели IDM и ускорения соседних транспортных средств выше на порог изменения δ:

$$R \to L: \quad (\widetilde{a}_i - a_i) + m\left((\widetilde{a}_{i+1} - a_{i+1}) + (\widetilde{a}_j - a_j)\right) \geqslant \delta,$$

где $m \in \left[-\infty; \frac{1}{2}\right] \cup [1; +\infty]$ — вручную задаваемый коэффициент вежливости, символ «$\widetilde{\ }$» обозначает измененные характеристики. В [8] предложены следующие интерпретации коэффициента m:

$m > 1$, *альтруистичное поведение:* автомобиль не перестраивается, если он ухудшит общую дорожную ситуацию;

$0 \leqslant m \leqslant 0.5$, *реалистичное поведение:* движение остальных транспортных средств менее приоритетно;

$m < 0$, *вредительское поведение:* автомобиль перестраивается, если он замедлит остальные.

Также следует учитывать, что при перестроении i на соседнюю полосу транспортные средства j и следующие за ним должны двигаться с коэффициентом торможения больше, чем b. Поскольку в модели IDM скорости, а, следовательно, и ускорения связаны формулой (1) и изменяются последовательно от лидирующего транспортного средства к последующему, то описать такое замедление можно формулой: $\widetilde{a}_j \geqslant -b$.

2.2. Описание работы модуля

Модуль симуляции трафика на вход получает конфигурационный файл VehicleConfig.xml, в котором содержатся такие параметры как максимально разрешенная скорость, коэффициент торможения (покрытие дороги), количество полос. Дополнительно, в модуль симуляции поступает информация о дорожной сети в виде мультиграфа map.osm, предобработанная библиотекой osmnx [10]. Для построения маршрутов транспортных средств в дорожной сети используется библиотека networkx [11].

Каждый шаг по времени t в модуле симуляции может быть отображен в модуле визуализации с использованием библиотеки pygame. Модуль визуализации отрисовывает дорожную сеть, транспортные средства и отладочную информацию, а также позволяет изменять модельное время.

3. Модуль валидации

В модуле валидации решается задача имитационного моделирования интенсивности движения транспортного потока. В данной работе под интенсивностью транспортного потока будем понимать число автомобилей, проезжающих через поперечное сечение участка дорожной сети в единицу времени.В модуле предусмотрены два способа подсчета интенсивности: первый способ, учитывающий количество машин за фиксированный временной интервал dt [22], второй — величину временного интервала между проездом двух автомобилей [23].

3.1. Моделирование интенсивностей транспортных потоков

На практике оптические датчики закреплены за соответствующей отдельной полосой и движение по полосам, как правило, является зависимым. Обозначим множество детектируемых участков $Zones, |Zones| = Z$.

Пусть случайная величина X_z с функцией распределения $F_z(x)$ есть временной интервал между проездом двух автомобилей через сечение детектируемого участка в зоне $z, z \in Zones$. Тогда система зависимых случайных величин $(X_1, \ldots, X_Z)$ с сов-

местной функцией распределения $H(x_1 \dots x_Z)$ описывает интенсивность движения транспортных потоков через детектируемый участок дорожной сети $Zones$.

По теореме Скляра [12] совместная функция распределения представима копулой C

$$H_{X_1 \dots X_Z}(x_1, \dots, x_Z) = C(F_1(x_1), \dots, F_Z(x_Z). \tag{2}$$

Для описания совместного распределения временных интервалов проезда автомобилей будем использовать копулы Маршалла-Олкина [12, 13] при $0 \leqslant \theta_z \leqslant 1, \widetilde{z} = \arg\min(u_1^{-\theta_1}, \dots, u_Z^{-\theta_Z})$

$$C(u_1, \dots, u_Z; \overline{\theta}) = u_{\widetilde{z}}^{-\theta_{\widetilde{z}}} \prod_{z \in Zones} u_z. \tag{3}$$

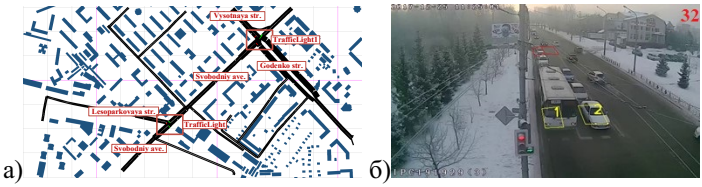
Тогда совместная плотность $h(x_1, \dots, x_Z)$ при $u_z = F_z(x_z)$ выражается через копулу (3) по формуле (2) следующим образом

$$h(x_1, \dots, x_Z) = (1 - \theta_{\widetilde{z}}) u_{\widetilde{z}}^{-\theta_{\widetilde{z}}} \prod_{z \in Zones} f_z(x_z). \tag{4}$$

3.2. Описание модуля валидации

Решение задачи моделирования зависимых распределений среднесуточных интенсивностей состоит из двух этапов: этапа предобработки и этапа имитационного моделирования.

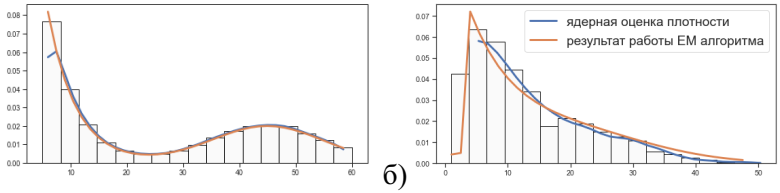
Опишем совместное распределение интенсивностей движения на примере полос 1 и 2 рисунка 4б. Пусть случайная величина X с функцией распределения $F(x)$ и случайная величина Y с функцией распределения $G(y)$ описывают временной интервал между проездом двух автомобилей через сечение детектируемого участка на полосах 1 и 2 соответственно. Скриншот графического интерфейса MARLIN24 изображен на рисунке 4а, зоны работы оптических датчиков — на рисунке 4б.



а) б)

Рис. 4. а)скриншот окна комплекса MARLIN24; б)зоны оптических датчиков (жёлтый цвет)

Этап предобработки состоит в оценивании плотности распределений случайных величин X и Y, описывающих число детектируемых транспортных средств на полосах 1 и 2 соответственно на основе данных, полученных с оптических детекторов города Красноярска с 2019 по 2020 год. На первом шаге строятся ядерные оценки плотности с ядром Епанечникова [15]. Далее формулируется упрощающее предположение о том, что каждая из рассматриваемых случайных величин представима в виде смеси нормальных распределений. С использованием EM-алгоритма [16], на вход которого продавались значения ядерной оценки плотности, определяются параметры смесей. Результаты этапа предобработки на 10-й итерации EM алгоритма и вид маргинальных плотностей распределений случайных величин X и Y приведены на рисунке 5а) и рисунке 5б).



а) б)

Рис. 5. Результаты этапа предобработки: оценка плотностей маргинальных плотностей: а) X; б) Y

Для этапа имитационного моделирования среднесуточных интенсивностей была разработана модификация метода дискретной суперпозиции Монте-Карло для генерации значений случайной величины (X, Y) [17, 23].

Гистограмма выборки, полученной моделированием копулой Маршалла-Олкина с параметрами $\overline{\theta} = (\frac{9}{10}, \frac{1}{4})$ и функция плотности распределения (4) представлены на рисунке 6.

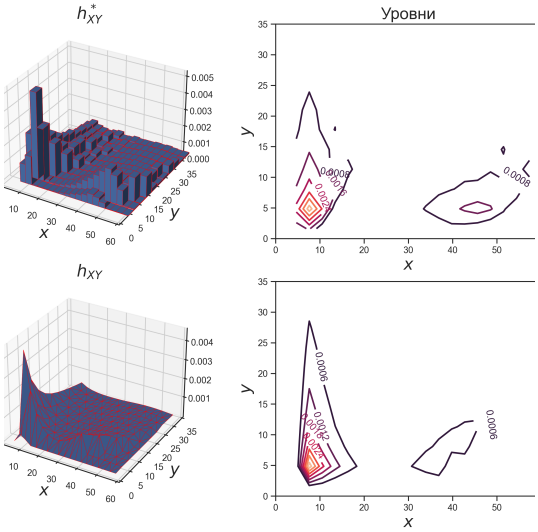


Рис. 6. Сравнение эмпирической h^*_{XY} с параметрами $\overline{\theta} = (\frac{9}{10}, \frac{1}{4})$ и теоретической h_{XY} функции плотности с.в. (X, Y)

Сравнение значений выборки с моделируемыми значениями (X, Y) приведено на рисунке 7.

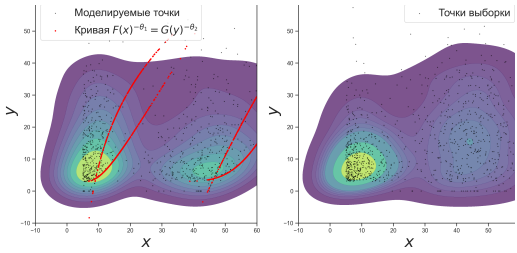


Рис. 7. Сравнение моделируемых значений и выборки (X, Y)

4. Модуль адаптивного управления светофорными объектами

Опишем задачу управления светофорным объектом как задачу управления агентом в стохастической среде. Агент (светофорный объект) не располагает ресурсами и решает задачу целесообразности активации той или иной фазы. Обозначим множество всех действий агента символом $\mathcal{A}$. Среда — детектируемые перекрестки с оптическими датчиками, которые распознают машины на отрезках дорог за сто метров до стоп-линий. Состояние среды отражает активность фаз светофорных объектов и время, которое машины находятся в детектируемой зоне. Обозначим множество всех состояний символом $\mathcal{S}$.

В качестве математической модели сети светофоров в работе рассматривается управляемый марковский процесс с конечным числом действий и состояний. Таким образом, проблема управления светофорными объектами сводится к задаче мультиагентного обучения с подкреплением (Multiagent Reinforcement Learning).

4.1. Задача мультиагентного обучения с подкреплением для светофорных объектов

Опишем поведение светофорных объектов (агентов) с помощью марковского процесса принятия решений $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, r \rangle$ [18]. Процесс принятия решений для агента будет выглядеть следующим образом. В момент времени t агент наблюдает состояние среды $s_t \in \mathcal{S}$ и выбирает действие $a_t \in \mathcal{A}$. Среда отвечает генерацией награды $R_t = r(s_t, a_t)$ и переходит в следующее состояние $s_{t+1} = s'$ с вероятностью $p(s' \mid s_t, a_t)$ по матрице переходов $\mathbb{P}$.

Функция оценки эффективности применяемого управления $\delta = \{a_t, t \in \mathbb{N}\}$ составляющая траекторию процесса $\mathcal{T} = \{s_0, a_0, s_1, a_1, \ldots, s_T, a_T\}$ получается как функция:

$$V = \sum_{t=0}^{\infty} \gamma^t r\left(s_{t+1} \mid s_t, \delta_t\right) = \lim_{T \to \infty} \mathbb{E}_{\mathcal{T}} \sum_{t=0}^{T} \gamma^t R_t, \tag{5}$$

где величина γ, $0 < \gamma < 1$, называется коэффициентом переоценки и показывает во сколько раз уменьшается отложенное вознаграждение за один временной шаг. Переоценка задает приоритет

получения награды в ближайшее время перед получением той же награды через некоторое время. Математический смысл условия $0 < \gamma < 1$ состоит в том, чтобы гарантировать ограниченность функционала V.

Формальная постановка задачи вычисления оценки эффективности управления светофорным объектом представлена ниже.

Дано: марковский процесс принятия решения $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, r \rangle$ для управления светофорным объектом, активная в начальный момент времени фаза светофорного объекта s_0.

Найти: управление светофорного объекта $\delta^* = \{a_t^*\}_{0 \leqslant t < \infty}$, которое доставит максимум функции оценки его эффективности (5).

Решение задачи поиска оптимального совокупного управления светофорными объектами дорожной сети ищется методом динамического программирования согласно принципу оптимальности Вальда—Беллмана. В задаче управления фазами светофорного объекта уравнение Вальда—Беллмана имеет вид [21]

$$V^* = \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} p(s' \mid s, a)(r(s, a) + \gamma V^*(s')). \tag{6}$$

Формула (6) может быть переписана в итерационной записи, называемой Q-обучение [18]. Функция суммарных вознаграждений при оптимальном управлении на шаге t имеет вид

$$V^* \left(\{s_{t'}, \delta\}_{t' \in \mathbb{N},\, t' \leqslant t} \right) = \max_{a \in \mathcal{A}} Q_t(s_t, a),$$

Считаем, что нам известно состояние среды s_{t+1} и оптимальное управление a_{t+1} на шаге $t+1$, соответствующий итерации l, и условимся, что итерация Q идет по индексу l, тогда функция Q для агента имеет рекурсивную запись

$$\begin{aligned} Q_{l+1}(s, a) &= \underbrace{p(s_{t+1}|s, a)}_{\alpha_l} \Big(r_{t+1} + \gamma V^*(s_{t+1}) \Big) + \\ &+ \underbrace{\sum_{s' \in S / s_{t+1}} p(s'|s, a)}_{1-\alpha_l} \Big(r(s'|s, a) + \gamma V(s') \Big) = \\ &= \alpha_l \Big(r_l + \gamma \max_{s'} Q_l(s_{t+1}, s') \Big) + \Big(1 - \alpha_l \Big) Q_l(s, a). \end{aligned}$$

Решение задачи поиска оптимального совокупного управления светофорными объектами дорожной сети ищется, чтобы увеличить максимальное совокупное вознаграждение, определяемое функцией Q. Для задачи поиска оптимального управления светофорным объектом справедливы следующие утверждения [21]:

- существует единственное точное решение;
- оценка точности приближенного решения на n-ом шаге имеет вид $\rho(Q_n, Q_0) \leqslant \dfrac{\gamma^n \rho(Q_1, Q_0)}{1-\gamma}$,
  где $Q_l \in \mathbb{R}_{\infty}^{|\mathcal{A}|+|\mathcal{S}|}$ — вектора значений $Q(s,a)$ на шаге l,
  $\forall q, w \in \mathbb{R}_{\infty}^{|\mathcal{A}|+|\mathcal{S}|}$ расстояние $\rho(q,w) = \max\limits_{j \in \mathbb{N},\, j \leqslant |\mathcal{A}|+|\mathcal{S}|} |q_j - w_j|$;
- приближенное решение находится согласно формулам

$$V^*(s) = \max_{a \in \mathcal{A}} \lim_{l \to +\infty} Q_l(s,a), \tag{7}$$

$$a_t(s) = \arg\max_{a' \in \mathcal{A}} Q_l(s,a'). \tag{8}$$

4.2. Описание модуля

Схема подсчета функции оценки эффективности управления управления светофорными объектами представлена на рисунке 8.

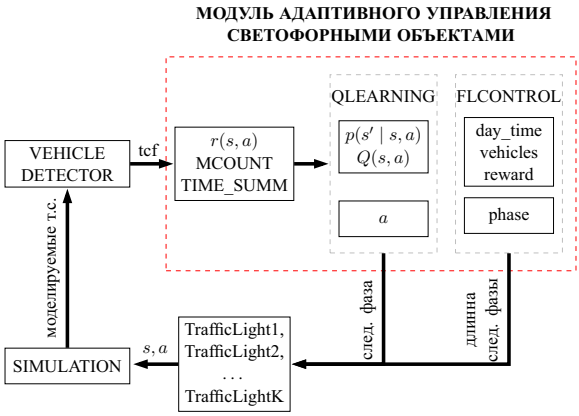


Рис. 8. Схема алгоритма MARLIN24

В имитационной среде (SIMULATION) моделируется транспортные потоки с интенсивностями, полученными в модуле валидации, автомобили перемещаются в имитационной среде (SIMULATION) пока не выйдут из ее зоны покрытия. При попадании машины на детектируемый участок дорожной сети z, во вспомогательном модуле, имитирующем поступление информации с оптических датчиков (VEHICLE DETECTOR), пары, состоящие из указателей на объект машины и текущего времени модели, добавляются в одну из коллекций tcf (time collection forward) для выбранного вручную множества светофорных объектов TrafficLight1, . . . , TrafficLightK. На следующем шаге симуляции машины удаляются из коллекции tcf, при проезде через зону. В течении периода времени dt во вспомогательном модуле выбора управления вызывается процесс переключения фаз (QLEARNING), реализующий управление согласно выбранной стратегии совокупного управления.

В качестве альтернативы процессу переключения фаз светофора (QLEARNING), использующемуся по умолчанию в модуле адаптивного управления светофорными объектами, можно выбрать изменение длительности фазы в следующем цикле. Данный подход был успешно реализован с помощью контроллера нечеткой логики (FLCONTROL) в работе [21]. Для управления светофором на основе метода Мамдани разработана система типа MISO (Multiple Input Single Output) с тремя входами [19].

Следующие лингвистические переменные рассматриваются в качестве входных данных: время суток с множеством термов $\mathcal{DT} = \{\text{morning}, \text{day}, \text{evening}, \text{night}\}$, плотность движения с множеством термов $\mathcal{V} = \{\text{run}, \text{wait}, \text{jam}\}$, время нахождения в зоне детекции с множеством термов $\mathcal{R} = \{\text{small}, \text{medium}, \text{large}\}$. Лингвистическая переменная выхода контроллера — длительность фазы, определенная множеством термов $P = \{$ long, medium, short$\}$. База правил представлена двенадцатью правилами (таблица 3) следующего вида:

R :IF $(day\_time = A)$ AND $(vehicles = B)$ AND $(reward = C)$ THEN $(phase = D)$, WHERE $\text{A} \in \mathcal{DT}, \text{B} \in \mathcal{V}, \text{C} \in \mathcal{R}, \text{D} \in \mathcal{P}$.

Таблица 3. Набор правил нечеткого вывода

	R_1	R_2	R_3	R_4	R_5	R_6
day_time	morning	night	-	-	day	day
$vehicles$	-	-	-	run	jam	-
$reward$	-	-	small	-	medium	medium
$phase$	short	short	short	short	short	medium

	R_7	R_8	R_9	R_{10}	R_{11}	R_{12}
day_time	evening	day	evening	-	-	-
$vehicles$	-	-	-	wait	jam	jam
$reward$	medium	large	large	medium	large	small
$phase$	medium	long	long	long	long	short

На основе выходных данных модуля принимается решение о переключении фазы светофоров или об изменении длительности следующей фазы.

5. Вычислительные эксперименты

Для исследования представленных в работе моделей были проведены серии вычислительных экспериментов. Эксперименты проводились на ПК с процессором Intel Core i7-10510U CPU@1.80ГГц и оперативной памятью объемом 8ГБ.

Опишем подробнее процесс подсчета политики управления $\hat{\pi}(a|s)$, опираясь на структурную схему комплекса MARLIN24 на рисунке 1. Модуль адаптивного управления светофорными объектами загружает управляющий конфигурационный файл trafficLightConfig.xml. В конфигурационном файле содержится информация о возможных направлениях движения, количестве фаз и циклах светофорных объектов. Далее комплекс программных средств MARLIN24 связывает показания датчика в имитационном модуле и рассчитывает оптимальное управление для светофорных объектов.

Оптические датчики (VEHICLE DETECTOR) в имитационной среде (SIMULATION) записывают момент появления t_i пронумерованного транспортного средства $i \in I \subset \mathbb{N}$ в зоне $z \in Zones = \left\{z^{(0)}, z^{(1)}, \ldots, z^{(m)}\right\}, m \in \mathbb{N}$. Отметим, что при имитационном моделировании псевдослучайная интенсивность

движения транспортных средств будет задана алгоритмически. Это означает, что мы можем сконструировать множество пар (i, z), что автомобиль i находится в детектируемой зоне z в момент времени t. Определим данное множество как отношение $\psi_t \subset \mathbb{N} \times Zones$, для которого $i\psi_t z$. Введем также отношение $\phi \subset Zones \times \mathcal{S}$, описывающее зоны z, в которых состояние s' разрешает движение. Сгруппируем автомобили в зонах в соответствии с фазой светофорного объекта s', которая разрешает движение транспортных средств в этих зонах и обозначим $I(s', t) = \{i|\ \ t_i < t, i\psi_t z, z\phi s'\}$.

Приведем рассуждения, исходя из которых считается функция вознаграждения. Для каждой полосы определено число машин на отрезке дороги, начинающемся с детектора и заканчивающемся стоп-линией перекрестка. Пусть $r : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ — функция вознаграждения агента при изменении наблюдаемого состояния s_t при действии $a_t = \delta(s)$. В момент времени t значение функции $r(s_t, a_t) = R_t$ определяется для следующей активной полосы и пропорционально времени, затраченному всеми машинами на преодоление детектируемых участков дороги $R_t = \sum\limits_{i \in I(s',t)} (t - t_i)$.

Далее для построенного множества светофорных объектов TL и зон детекции z определяются функция наград $r(s, a)$ число проехавших машин (MCOUNT), суммарное время проезда через детектируемые участки дорожной сети (TIME_SUMM) и обучающие функции Q. При симуляции, формируется двумерная выборка $\mathcal{X} = \{(s_i, a_i)\}_{i=1}^{T}$ объемом T порядка 10^6. В результате управления δ^*, принятого из соображений увеличения значения функции оценки эффективности Q, рассчитывается несмещенная оценка распределения $\mathcal{P} = \{p(s, a)|\ s \in \mathcal{S}, a \in \mathcal{A}\}$ двумерной случайной величины (s, a), где функция $p(s, a)$ — вероятность того, что в состоянии s агент принял решение a. На основании выборочных вероятностей $\hat{p}(s, a)$ вычисляются оценки политики агента $\hat{\pi}(a|s)$ для каждого $s \in \mathcal{S}$

$$\hat{\pi}(a|s) = \frac{\hat{p}(s, a)}{\sum\limits_{a \in \mathbb{A}} \hat{p}(s, a)} = \frac{\hat{p}(s, a)}{\hat{p}(s)}.$$

В ходе серии из 1000 симуляций были построены усредненные кривые обучения функции оценки эффективности управления $\hat{Q}$ при равновесной по Нэшу стратегии [24]. Кривые обучения приведены на рисунке 9, где синим цветом отмечены графики принимаемых значений Q для каждой эпохи по модельному времени t, красным цветом — их усредненные значения $\hat{Q}$.

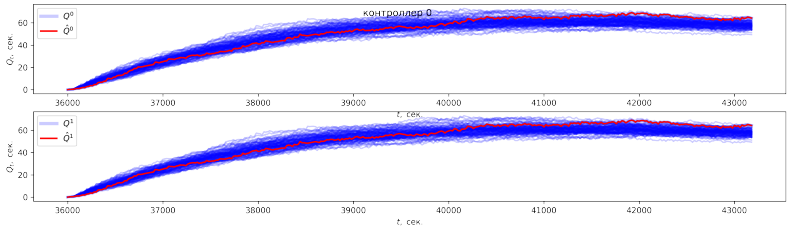


Рис. 9. Пример сходимости кривой обучения Q

В рамках вычислительных экспериментов было проведено сравнение кривых обучения агентов на протяжении 1000 эпох. В результате эффективного управления время ожидания транспортного средства в среднем не превышает длины цикла светофорного объекта. Также было продемонстрировано, что значительного улучшения управления при расширении покрытия дорожной сети может и не быть. Таким образом, координированное управление светофорными объектами в целях ускорения вычислений может быть рассмотрено только в тех участках, где его применение дает ощутимое улучшение в управлении. В остальных случаях может быть рассмотрен некоординированный подход, и, следовательно, «проклятие размерности», возникающее с ростом размерности матриц при вычислениях, не является серьезной проблемой.

В работе [20] было проведено сравнение показателей эффективности управления с комплекса MARLIN24 и АСУДД24. По результатам эксперимента [20], были получены сопоставимые результаты для среднего времени ожидания на участника движения для модели, управляемой марковским процессом принятия решения MARLIN. Данный результат демонстрирует насколько эф-

фективно адаптивные системы светофорных объектов могут быть применены для разгрузки сложных участков дорожной сети.

Также в работе [21] было проведено сравнение показателей эффективности управления, рассчитанных для координированного адаптивного управления светофорными объектами участка дорожной сети (MARLIN), некоординированного адаптивного управления (MARL), для светофорных объектов с фиксированным координационным планом (FIXED) и с переменной длительностью фаз (FUZZY) По результатам эксперимента, было получено качественное отличие для на 30% для среднего времени ожидания на участника движения для модели, управляемой марковским процессом принятия решения MARLIN.

6. Заключение

Литература

1. Транспортная стратегия Российской Федерации, утверждена распоряжением Правительства Российской Федерации от 22 ноября 2008 года №1734-р. – 2008. – Дата обращения: 25.06.2024. http://mintrans.gov.ru.
2. Carini, Raymond N. Application of the UTCS-1 Network Simulation Model to Select Optimal Signal Timings in a Multi-Linear Street System / Raymond N. Carini // Interim Report. Publication for Urban Traffic Control System. Joint Highway Research Project. – 1977. – P. 164.
3. Chandler, M.J.H. SCOOT and Bus Detection. OTRC Proc. Annual Summer Meeting / M.J.H. Chandler // Traffic Control Studies in London. – 1990. – Vol. P269. – Pp. 111–128.
4. Gartner, N.H. OPAC: Strategy for Demand-responsive Decentralized Traffic Signal Control / N.H. Gartner // IFAC Proceedings Volumes. – 1990. – Vol. 23. – Pp. 241–244.
5. АСУДД «МИКРО–М». – Дата обращения: 25.06.2024. http://asud55.ru/archives/1346.
6. El-Tantawy, S. Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto / S. El-Tantawy, B. Abdulhai, H. Abdelgawad // IEEE Transactions on Intelligent Transportation Systems. – 2013. – Vol. 14, no. 3. – Pp. 1140–1150.
7. Treiber, M. Congested traffic states in empirical observations and microscopic simulations / M. Treiber, A. Hennecke, D. Helbing // Transportation Physics Reviews E. – 2000. – Vol. 62. – Pp. 1805–1824.
8. Treiber, M. Realistische Mikrosimulation von StraBenverkehr mit einem einfachen Modell / M. Treiber, D. Helbing // 16. Symposium "Simulationstechnik ASIM 2002"Rostock. – 2002. – Pp. 514–520.

9. Зализняк, В.Е. Численные методы. Основы научных вычислений: учебное пособие для бакалавров / В.Е. Зализняк. – ЮРАЙТ, 2012. – С. 356.
10. Boeing, G., OSMnx: Python for Street Networks. – Дата обращения: 25.06.2024. https://osmnx.readthedocs.io/.
11. NetworkX Developers, NetworkX. – Дата обращения: 25.06.2024. https://networkx.org/.
12. Nelsen, R. B., An Introduction to Copulas. Springer, 2006. — 270 с.
13. Marshall, A.W. Families of Multivariate Distributions / A.W. Marshall, I. Olkin // Journal of the American Statistical Association. – 1988. – Vol. 83, no. 403. – Pp. 834–841.
14. Quesada-Molina, J.J. Bivariate copulas with quadratic sections / J.J. Quesada-Molina, J.A. Rodriguez-Lallena // Journal of Nonparametric Statistics. – 1995. – Vol. 5, no. 4. – Pp. 323–337.
15. Epanechnikov, V. A., "Non-Parametric Estimation of a Multivariate Probability Density,"Theory of Probability and Its Applications, 1969, pp. 153–158.
16. C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2007.
17. А. В. Войтишек, Основы метода Монте-Карло: Учебное пособие, ред. В. Н. Задорожный. Новосибирск: НГУ, 2010. — 108 с.
18. Sutton, R.S. Reinforcement Learning: An Introduction / R.S. Sutton, A.G. Barto. – Cambridge, MA: The MIT Press, 2015. – Pp. 145–147.
19. Zadeh, L. A. Fuzzy Sets, Fuzzy Logic, Fuzzy Systems / L. A. Zadeh. – World Scientific Press, 1996.
20. Tislenko, T.I. Multiagent Reinforcement Learning for Integrated Network: Applying to a Part of the Road Network of Krasnoyarsk City / T.I. Tislenko, D.V. Semenova, N.A. Sergeeva et al. // IEEE 16th International Conference on Application of Information and Communication Technologies (AICT). – 2022. – Pp. 1–5.
21. Tislenko, T.I. Modeling and Comparison of Different

Management Approaches on the Intersections Network / T.I. Tislenko, D.V. Semenova, A.A. Soldatenko // 2023 IEEE 26th International Conference, Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN). – 2023. – Pp. 25–29.

22. Тисленко, Т.И. Моделирование интенсивностей транспортных потоков при помощи копул Маршала-Олкина, // ИТ. НАУКА. КРЕАТИВ. Т. 5. Системы управления, информационные технологии и математическое моделирование : материалы I Междунар. форума (Омск, 14–16 мая 2024 г.) : в 5 Ч. / науч. ред. П. С. Ложников, отв. ред. И. Г. Ольгина. – Омск : Издательство ОмГТУ, 2024. – 70 с.
23. Тисленко, Т.И. Моделирование интенсивностей транспортных потоков для модуля валидации комплекса MARLIN24 / Т.И. Тисленко // Информационные технологии и математическое моделирование (ИТММ-2024): Материалы XXIII Международной конференции имени А.Ф. Терпугова. – Карши: ***, 2024. – С. ***–***.
24. Nash, J., "Equilibrium Points in N-Person Games,"Proceedings of the National Academy of Sciences, vol. 36, no. 1, 1950, pp. 48–49.

DEVELOPMENT OF AN ADAPTIVE TRAFFIC LIGHT CONTROL SYSTEM USING MARKOV DECISION PROCESSES

Timofey I. Tislenko, Siberian Federal University, Krasnoyarsk, postgraduate student (timtisko@mail.ru).
Darya V. Semenova, Siberian Federal University, Krasnoyarsk, Cand.Sc. (Physics and Mathematics), assistant professor (DVSemenova@sfu-kras.ru).

Abstract: The paper presents the results of developing the MARLIN24 software for adaptive traffic light control. The structure of the package includes a module for adaptive traffic light control, a traffic simulation module, and a validation module. The mathematical model of the traffic light control process is described using a Markov decision process with a finite number of actions and states. The task is to find the effective control that reduces the total time vehicles spend at the observed sections of the road network. We formulate the task as a multi-agent reinforcement learning problem (Multiagent Reinforcement Learning for Integrated Network, MARLIN). A Q-learning algorithm was developed to solve the MARLIN problem. The structure of the software package also includes a micro-simulation module for traffic flows. This module uses the "Intelligent Driver Model" (IDM). The validation module uses multivariate distributions to simulate real-time traffic data obtained from an optical sensors. The multivariate distributions are constructed by applying Marshall–Olkin copulas to marginal empirical distributions of timestamps for optical sensor triggers. Marginal distributions were constructed based on traffic intensity data for sections of the road network from 2019 to 2020.

Keywords: traffic flow simulation, IDM model, MOBIL model, Marshall-Olkin copulas, reinforcement learning .

УДК 519.1,519.2,519.6,519.8(075), 004.42, 519.85
ББК 221.7

Статья представлена к публикации членом редакционной коллегии ...

Поступила в редакцию ...
Дата опубликования ...