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 комплекса программных средств 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 .

---

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

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

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

1. Введение

Одной из задач решаемых в ходе реализации транспортной стратегии России на 2035 год [**?**] является увеличение пропускной способности и увеличение скоростных параметров дорожной инфраструктуры до уровня лучших мировых достижений. Разработка и внедрение программных и математических инструментов для моделирования транспортных потоков и управления светофорными объектами для наиболее нагруженных участков дорожной сети учитывает общесоциальные целевые ориентиры транспортной стратегии: подвижность населения, снижение аварийности, рисков и угроз безопасности по видам транспорта, снижение доли транспорта в загрязнении окружающей среды. Одним из подходов к решению поставленных задач является использование продвинутых систем, управляющих светофорными объектами. В таблице **??** представлены наиболее известные АСУД (адаптивные системы управления дорожным движением).

Таблица 1. Модели адаптивных систем светофоров

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

Приведённые адаптивных систем светофоров были разработаны в различные временные периоды и для различных условий движения. Например, система SCOOT (Split Cycle Offset Optimization Technique) анализирует данные о дорожной обстановке и корректирует светофорные сигналы, чтобы предотвратить образование заторов до их появления. Для работы данной си-

стемы требуется установка плотной сети индукционных петель, камер и других датчиков движения на расстоянии около сорока метров от 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) — 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» [**?**] — наиболее широко используемая в России система, успешно применяемая в следующих регионах: Иркутск, Красноярский край, Иркутская область, Красноярск, Белгород, Ангарск, Воронеж, Хабаровск, Московская область. АСУДД «Микро» является децентрализованной системой и поддерживает до шести GPRS-серверов, которые позволяют подключить 250 перекрёстков. Отечественные оптические детекторы серии «Инфопро», используемые в АСУДД «Микро», предназначены для сбора статистических данных о транспортном потоке и данных реального времени для актуального управления; они работают на расстоянии до 100 метров. По дальности распознавания данные датчики не уступают их бельгийскому аналогу TrafiCam. Основным недостатком системы является тот факт, что реализа-

ция адаптивных алгоритмов находится в стадии разработки.

В данной работе предложен программный комплекс MARLIN24, реализующий часть методов новейших АСУД, использующих метод мультиагентного обучения с подкреплением (Multiagent Reinforcement Learning for Integrated Network, MARLIN) [**?**]. 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. К достоинствам подхода на основе MARLIN можно отнести следующие положения. 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. Существенным ограничением является рост вычислительной сложности при увеличении обзора агента, известный как «проклятие размерности». Структура комплекса программных средств MARLIN24 приведена на рисунке **??**.

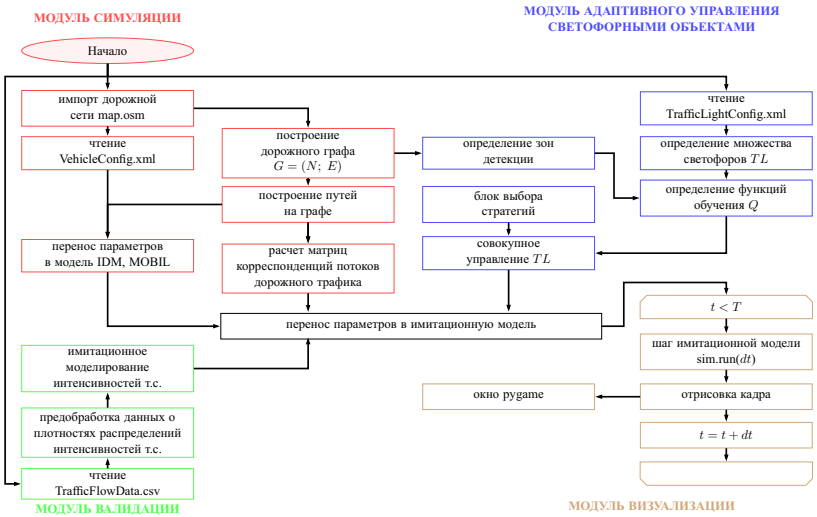


Рис. 1. Структура комплекса программных средств MARLIN24

В параграфе **??** приведено описание задачи машинного обучения с подкреплением для одного агента. Агент (светофорный объект) не располагает ресурсами и решает задачу целесообразности активации той или иной светофорной фазы. Обозначим множество всех действий агента символом $\mathcal{A}$. Среда — детектируемый перекресток с оптическими датчиками, которые распознают машины на отрезках дорог за сто метров до стоп-линий. Состояние среды отражает активность фазы светофорных объектов и время, которое машины находятся в детектируемой зоне. Обозначим множество всех состояний среды символом $\mathcal{S}$. В качестве математической модели светофорного объекта в работе рассматривается управляемый марковский процесс с конечным числом действий и состояний $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, r \rangle$. Проблема управления светофорными объектами сводится к задаче мультиагентного обучения с подкреплением (Multiagent Reinforcement Learning).

В параграфе **??** приведено описание задачи машинного обучения с подкреплением для нескольких агентов. Введены основные определения и обозначения для задачи обучения нескольких агентов. Задача поиска управления одного агента сведена к задаче машинного обучения нескольких агентов.

В параграфе **??** приведено решение задачи машинного обучения с подкреплением для нескольких агентов. Приведены критерий оптимального управления Вальда-Беллмана в предложении **??**, итерационная запись для функции оценки эффективности управления V, а также приведено предложение **??** о существовании и единственности оптимального решения. В параграфе **??** выведены формулы для поиска оптимального решения (**??**) и (**??**).

В параграфе **??** были проведены вычислительные эксперименты для реального участка дорожной сети, построены кривые обучения функции оценки эффективности управления и дана интерпретация оптимальному управлению.

2. Модуль симуляции

Модуль симуляции предназначен для оценки эффективности выбранного управления. Данный модуль позволяет имитировать

показания оптических датчиков. При работе он должен учитывать фазы, циклы и программы управления светофорных объектов.

2.1. Модели движения трансопртных средств

В симуляционном модуле при движении транспортных средств должны быть учтены минимальная безопасная дистанция, максимальная разрешенная скорость и коэффициент торможения транспортных средств.

Модель интеллектуального водителя IDM (Intelligent Driver Model)[**?**] позволяет описывать движение с учетом выбранных параметров. В модели IDM все транспортные средства рассматриваются как индивидуальные сущности, обладающие характеристиками и поведением. Модель IDM относится к классу моделей движения за лидером, она основана на взаимодействии между автомобилями, где каждый водитель регулирует скорость своего автомобиля в зависимости от расстояния до впереди идущего транспорта, его скорости и собственной скорости.

На рисунке **??** представлено взаимное расположение и характеристики текущего автомобиля, расположенного в i-ой позиции, и $(i-1)$-го автомобиля, находящегося перед ним.

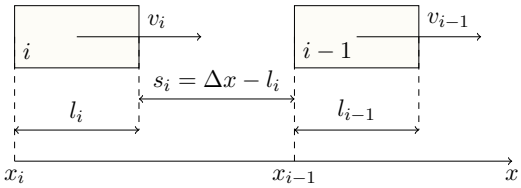


Рис. 2. Иллюстрация модели IDM

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

в классических обозначениях имеет вид:

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

При имитационном моделировании для нахождения значений скорости и ускорения будем пользоваться формулами, вытекающими из численного метода «пристрелки»:

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

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

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 (Microscopic Optimally Balanced Intersection Lanes)[**?**].

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 рисунок **??**, зависит, как правило, от двух следующих транспортных средств на текущей полосе движения и соответственно на целевой полосе движения.

Стимул для перестроения есть, если после первого фиктивного перестроения с правой полосы R на левую полосу L сумма

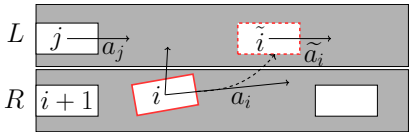


Рис. 3. Модель смены полосы MOBIL

собственного ускорения согласно модели IDM и ускорения соседних транспортных средств выше на порог изменения δ:

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

где $p \in \left[-\infty; \frac{1}{2}\right] \cup [1; +\infty]$ — вручную задаваемый коэффициент вежливости, символ $\widetilde{}$ обозначает измененные характеристики.

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_{safe}. Поскольку в модели IDM скорости, а, следовательно, и ускорения связаны формулой (**??**) и изменяются последовательно от лидирующего транспортного средства к последующему, то описать такое замедление можно формулой: $\widetilde{a_j} \geqslant -b_{\text{safe}}$.

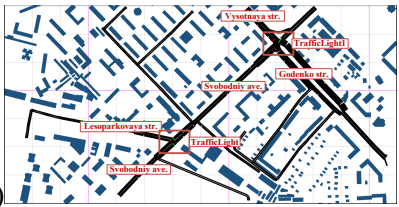
2.2. Описание работы модуля

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[**?**]. Для построения маршрутов транспортных средств в дорожной сети используется библиотека networkx[**?**].

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. Использование графических средств и модуля симуляции значительно ускоряет процесс отладки и исследования моделей управления светофорами.

Скриншот графического интерфейса MARLIN24 изображен

на рисунке **??**а, зоны работы оптических датчиков — на рисунке **??**б.

а) 

б) 

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

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, а под среднесуточной интенсивностью — усредненное количество машин по рабочим дням в течение года. Ранее в работе [**?**] при описании интенсивности транспортного потока использовалась статистка, описывающая количество машин, в работе [**?**] использовалась величина временного интервала между проездом двух автомобилей через сечение участка дорожной сети.

3.1. Модель транспортного потока

Для введения в модель зависимых случайных величин, описывающих интенсивность движения транспортных средств, рассмотрим временной интервал между проездом двух автомобилей через сечение участка дорожной сети. В основе подхода лежит использование копул из семейства Маршалла-Олкина [**?**] для описания совместного распределения временных интервалов появления автомобилей.

Опишем совместное распределение интенсивностей движения для полос 7 и 8 рисунка **??**б. 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 7 и 8 соответственно. По теореме Скляра [**?**] совместную функцию распределения можно представить копулой C

$$H_{XY}(x,y) = C(F(x), G(y)), \qquad \forall x, y \in \mathbb{R}. \tag{2}$$

Далее будем использовать двупараметрическую $(n = 2)$ копулу Маршалла-Олкина [**?**, **?**] с коэффициентами $0 \leqslant \alpha, \beta \leqslant 1$, $\theta_1 = \alpha, \theta_2 = \beta$

$$C(u_1, u_2, \ldots, u_n; \overline{\theta}) = \prod_{i=1}^{n} u_i^{1-\theta_i} \cdot \min(u_1^{\theta_1}, u_2^{\theta_2}, \ldots, u_n^{\theta_n}). \tag{3}$$

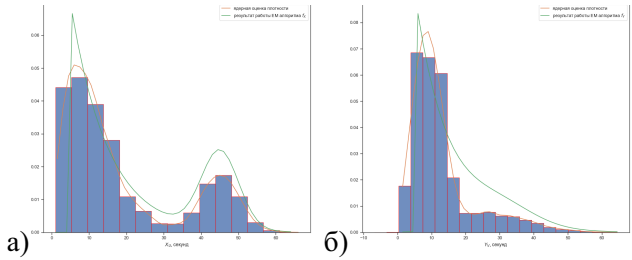
3.2. Описание модуля валидации

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

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 [**?**]. Далее формулируется упрощающее предположение о том, что каждая из рассматриваемых случайных величин представима в виде смеси нормальных распределений. С использованием EM-алгоритма [**?**], на вход которого продавались значения ядерной оценки плотности, определяются параметры смесей.

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

Результаты этапа предобработки на 10-й итерации EM алгоритма и вид маргинальных плотностей распределений случайных величин X и Y приведены на рисунке **??**а) и рисунке **??**б).



а) б)

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

Гистограмма выборки, полученной моделированием копулой Маршалла-Олкина с параметрами $\alpha = 0.9, \beta = 0.25$ представлена на рисунке **??**.

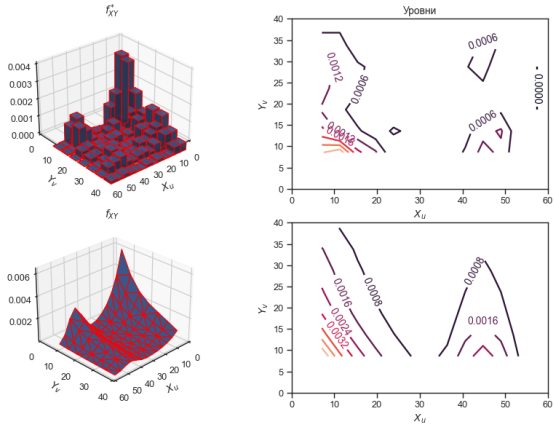


Рис. 6. Сравнение эмпирической f^*_{XY} и теоретической f_{XY} функции плотности распределения (X, Y) при моделировании значений (U, V) для копулы Маршалла-Олкина с параметрами $\alpha = 0.9, \beta = 0.25$

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 $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, r \rangle$ [**?**]. 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:

$$V = \sum_{t=0}^{\infty} \gamma^t r\left(s_{t+1} \mid s_t, \delta_t\right) = \lim_{T \to \infty} \mathbb{E}_{\mathcal{T}} \sum_{t=0}^{T} \gamma^t R_t, \tag{4}$$

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 (**??**).

Решение задачи поиска оптимального совокупного управления светофорными объектами дорожной сети ищется методом динамического программирования согласно принципу оптимальности Вальда—Беллмана. Справедливо слудующее утверждение.

[**?**] В задаче управления фазами светофорного объекта уравнение Вальда—Беллмана имеет вид

$$V^* = \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} p(s' \mid s, a)(r(s, a) + \gamma V^*(s')). \tag{5}$$

Формула (**??**) может быть переписана в итерационной записи, называемой Q-обучение. 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

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

Решение задачи поиска оптимального совокупного управления светофорными объектами дорожной сети ищется, чтобы увеличить максимальное совокупное вознаграждение, определяемое функцией Q. Справедливо слудующее утверждение. [**?**] Для задачи поиска оптимального управления светофорным объектом с любым количеством фаз

- существует единственное точное решение;
- оценка точности приближенного решения на n-ом шаге итерации
$$\rho(Q_n, Q_0) \leqslant \frac{\gamma^n \rho(Q_1, Q_0)}{1-\gamma},$$
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
$$V^*(s) = \max_{a \in \mathcal{A}} \lim_{l \to +\infty} Q_l(s,a), \tag{6}$$
$$a_t(s) = \arg\max_{a' \in \mathcal{A}} Q_l(s,a'). \tag{7}$$

4.2. Описание модуля

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

В имитационной среде (SIMULATION) моделируется транспортные потоки с интенсивностями, полученными в модуле

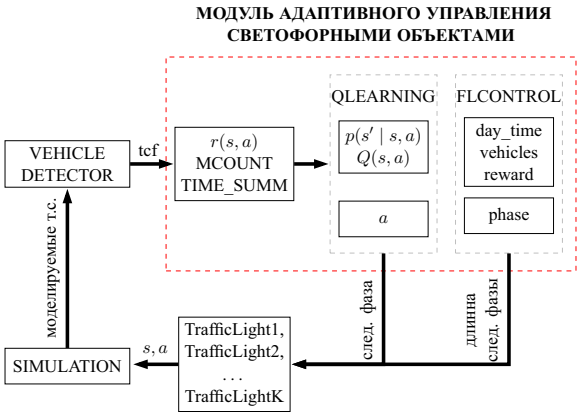


Рис. 7. Схема алгоритма MARLIN24

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 модуль QLEARNING, реализующий управление согласно выбранной стратегии совокупного управления. На основе выходных данных модуля принимается решение о переключении фазы светофоров.

5. Структура программного комплекса

Опишем подробнее процесс подсчета $\hat{\pi}(a|s)$, опираясь на структурную схему. Модуль адаптивного управления светофорными объектами загружает управляющий конфигурационный файл trafficLightConfig.xml. В конфигурационном файле содер-

жится 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.

Пусть 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 распределения $p(s,a)$ — 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 $s \in \mathcal{S}$

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

Наряду с политиками агента, при обработке интенсивностей записываются массивы $r_{a^{(k)}} = \left\{r(s_0,a^{(k)}), r(s_1,a^{(k)}), r(s_2,a^{(k)}), \dots\right\}$, $k \in K$. Элементы массивов $r\left(s_t,a^{(k)}\right)$ — время нахождения машин на активируемых агентом k и фазой $a^{(k)} \oplus s_t$ полосах.

5.1. Вычислительные эксперименты и обсуждение

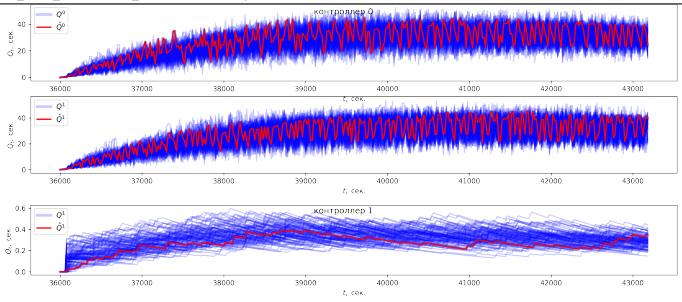
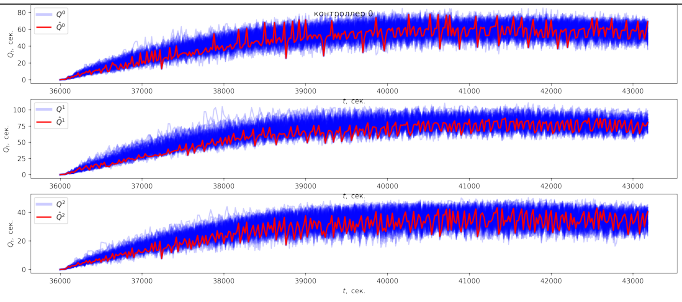
Для исследования представленных в работе моделей были проведены серии вычислительных экспериментов. Эксперименты проводились на ПК с процессором Intel Core i7-10510U CPU@1.80ГГц и оперативной памятью объемом 8ГБ.

В ходе серии из 1000 симуляций были построены усредненные кривые обучения функции оценки эффективности управления $\hat{Q}$ при равновесной по Нэшу стратегии для ограниченного и неограниченного пространств обзора состояний $\mathcal{O}$. Кривые обучения приведенны в таблице **??**.

Отметим, что при рассмотрении полного пространства состояний, суммарное время нахождения транспортных на участке, принадлежащем второму светофорному объекту, значительно меньше, чем на 0 и 1-ом, и поэтому в случае полного обзора оно не вносило вклада в изменение управления.

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

Таблица 2. Графики кривых обучения

частичный обзор $\mathcal{O} \subset \mathcal{S}^0 \times \mathcal{S}^1,$ $\mathcal{S} = \mathcal{S}^0 \times \mathcal{S}^1 \times \mathcal{S}^2$	
полный обзор $\mathcal{O} = \mathcal{S} = \mathcal{S}^0 \times \mathcal{S}^1 \times \mathcal{S}^2$	

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.

описать словами ограниченный обзор

Сравнение комплекса MARLIN24 и АСУДД24 в работе [**?**] показало сопоставимые результаты (таблица **??**).

Таблица 3. Сравнение показателей эффективности управления для различных моделей

Целевая функция	Ед. изм.	АСУДД24	MARLIN24	улучшение
Средняя задержка	сек. / маш.	10.63	9.4	11.6%
Пропускная способность	маш.	4 870	4 412	-9.4%
Суммарное время	сек.	51 792	41 286	20.3%

Литература

ARTICLE TITLE

Alexander Ivanov, Institute of Control Sciences of RAS, Moscow, Cand.Sc., assistant professor (aaivanov@mail.ru).
Boris Petrov, Institute of Control Sciences of RAS, Moscow, Doctor of Science, professor (Moscow, Profsoyuznaya st., 65, (495)000-00-00).
Vikor Sidorov, Moscow Institute of Physics and Technology, Moscow, student (viktor.sidorov@mipt.ru).

Abstract:

Keywords: .

УДК ...
ББК ...

Статья представлена к публикации членом редакционной коллегии ...

Поступила в редакцию ...
Дата опубликования ...