

РАЗРАБОТКА АДАПТИВНОЙ СИСТЕМЫ УПРАВЛЕНИЯ СВЕТОФОРНЫМИ ОБЪЕКТАМИ С ИСПОЛЬЗОВАНИЕМ МАРКОВСКИХ ПРОЦЕССОВ ПРИНЯТИЯ РЕШЕНИЙ¹

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 Маршалла-Олкина к маргинальным эмпирическим 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 [?], является увеличение про-

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 (DVSeменова@sfu-kras.ru).

пусковой 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. Модели адаптивных систем управления дорожным движением

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»
работы, авторы					

Одн из ранних систем управления городским дорожным движением была UTCS-1 (Urban Traffic Control System)[?]. 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 планы, 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)[?] 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)[?] — 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: Красноярский край, Иркутская область, Белгородская область, Воронежская область, Хабаровский край, Московская область.

АСУДД «Микро» является децентрализованной системой и поддерживает до шести GPRS-серверов, которые позволяют подключить до 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. Эти детекторы работают на расстоянии до 100 метров. По дальности распознавания данные датчики не уступают их бельгийскому аналогу TraftCam. Основным недостатком системы является тот факт, что реализация адаптивных алгоритмов находится в стадии разработки [?].

В работах [?, ?] 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».

Фокус наших исследований сосредоточен на разработке программных и математических инструментов для адаптивного управления сетью светофорных объектов участков дорожной сети города Красноярск. В настоящей работе представили новый программный комплекс MARLIN24, реализующий часть методов новейших АСУДД. Оценка эффективности применяемого управления и валидация полученных результатов осуществляется с по-

мощью симуляционных экспериментов на основе статистических данных с оптических детекторов за 2018 и 2019 года.

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

1. **Общее описание программного комплекса MARLIN24**

Структура MARLIN24 приведена на рисунке 1 и состоит из четырех модулей: симуляции, валидации, адаптивного управления светофорными объектами, визуализации.

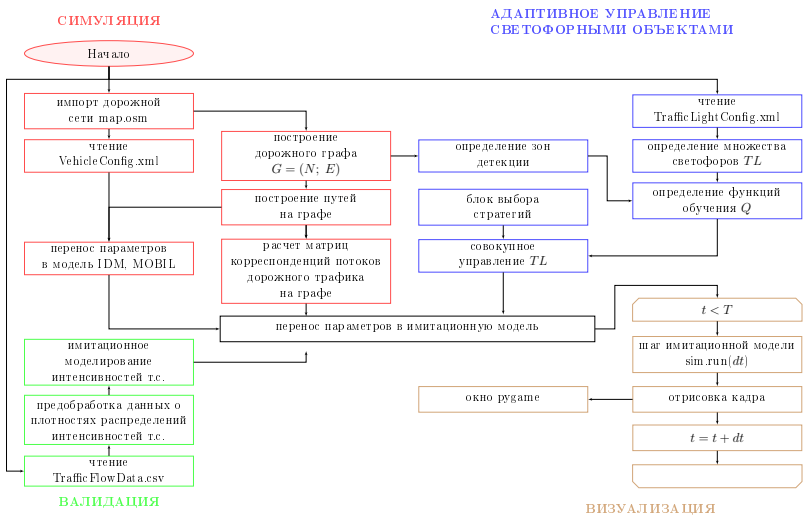


Рис. 1. Структура программного комплекса MARLIN24

Модуль симуляции предназначен для оценки эффективности выбранного управления. Он реализует микросимуляционные модели движения, предназначенные для описания перемещения транспортных средств по дорожной сети. Часть программного кода, отвечающая за валидацию, позволяет моделировать интен-

сивности движения транспортных потоков. Включение модулей является опциональным, так графический интерфейс в модуле визуализации позволяет производить визуальную оценку применяемого управления. Данный модуль отображает транспортные средства, дорожную сеть, дополнительную информацию о содержании названий дорог, их координатах. Также в модуле реализована функция ускорения и замедления модельного времени, выделения маршрутов, масштабирования и перемещения зоны обзора.

2. Модуль симуляции

2.1. модели движения транспортнх средств вся математика

2.2. описание работы модуля

вся математика Модуль симуляции предназначен для оценки эффективности выбранного управления. 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[?]`. Данный модуль позволяет имитировать показания оптических датчиков. 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 световыми объектами.

Скриншот графического интерфейса MARLIN24 изображен на рисунке 2а, зоны работы оптических датчиков — на рисунке 2б.



Рис. 2. а)скриншот окна комплекса MARLIN24;
б)зоны оптических датчиков (жёлтый цвет)

Симуляционный модуль должен учитывать фазы, циклы и программы управления светофорных объектов. Конфигурационный файл TrafficLightConfig.xml содержит информацию об активируемых фазах и активируемых направлениях для выбранных светофорных объектов.

В симуляционном модуле движение транспортных средств должно учитывать минимальную безопасную дистанцию, максимальную разрешенную скорость и коэффициент торможения транспортных средств. Данные параметры могут оказывать существенное влияние при описании загруженных участков дорожной сети.

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

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

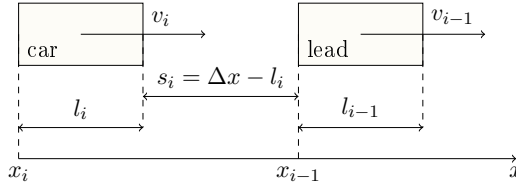


Рис. 3. Иллюстрация модели IDM

в классических обозначениях имеет вид:

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

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

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

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 (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 4, зависит, как правило, от двух следующих транспортных средств на текущей полосе движения и соответственно на целевой полосе движения.

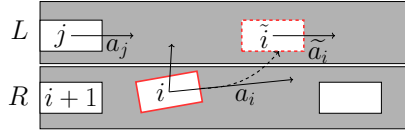


Рис. 4. Модель смены полосы MOBIL

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 :

$$R \rightarrow L : (\tilde{a}_i - a_i) + p((\tilde{a}_{i+1} - a_{i+1}) + (\tilde{a}_j - a_j)) \geq \delta,$$

где $p \in [-\infty; \frac{1}{2}] \cup [1; +\infty]$ — вручную задаваемый коэффициент вежливости, символ $\tilde{}$ 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 (1) и изменяются последовательно от лидирующего транспортного средства к последующему, то описать такое замедление можно формулой: $\tilde{a}_j \geq -b_{\text{safe}}$.

3. Модуль валидации

3.1. матмодель

3.2. описание модуля

В модуле валидации решается задача имитационного моделирования интенсивности движения транспортного потока. Решение задачи моделирования зависимых распределений среднесуточных интенсивностей состоит из двух этапов: этапа предобработки и этапа имитационного моделирования.

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.

Рассмотрим второй подход для введения в модель зависимых случайных величин. 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 2б. 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 Склера [?] совместную функцию распределения можно представить копулой C

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

Далее будем использовать двухпараметрическую копулу Маршалла-Олкина [?, ?] с коэффициентами $0 \leq \alpha, \beta \leq 1$, $\theta_1 = \alpha, \theta_2 = \beta$

$$(3) \quad C(u_1, u_2, \dots, u_n) = \prod_{i=1}^n u_i^{1-\theta_i} \cdot \min(u_1^{\theta_1}, u_2^{\theta_2}, \dots, u_n^{\theta_n}).$$

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 Красноярск 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 ЕМ-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 ЕМ алгоритма и вид маргинальных плотностей распределений случайных величин X и Y приведены на рисунке 5а) и рисунке 5б) соответственно.

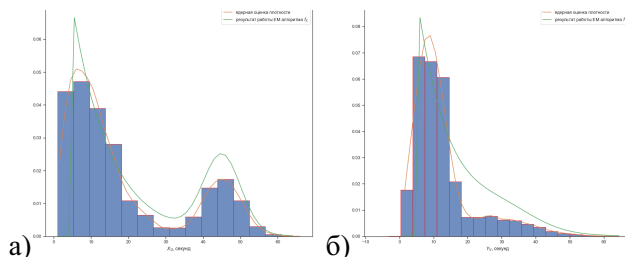


Рис. 5. Результаты этапа предобработки: оценка плотностей маргинальных плотностей: а)случайной величины X ; б) случайной величины Y

Гистограмма выборки, полученной моделированием копулой Маршалла-Олкина с параметрами $\alpha = 0.9, \beta = 0.25$ представлена на рисунке 6.

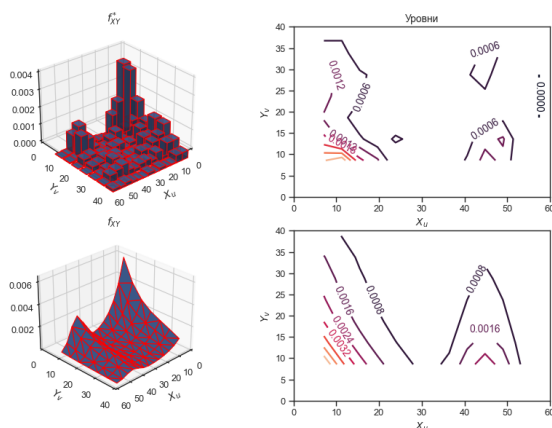


Рис. 6. Сравнение эмпирической f_{XY}^* и теоретической f_{XY} функции плотности распределения (X, Y) при моделировании значений (U, V) для копулы Маршалла-Олкина с параметрами $\alpha = 0.9, \beta = 0.25$

3.3. Модуль адаптивного управления светоформными объектами

На основании данных, полученных при наблюдении в модуле симуляции, формируется двумерная выборка $\mathcal{X} = \{(s_i, a_i)\}_{i=1}^N$ объемом N порядка 10^6 . В результате управления δ^* , принятого из соображений увеличения значения функции оценки эффективности (4) с учетом выбранной стратегии агентов, 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 $\hat{p}(s, a)$ вычисля-

ются оценки политики агента $\hat{\pi}(a|s)$ для каждого $s \in \mathcal{S}$

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

Наряду с политиками агента, при обработке интенсивностей записываются массивы $r_{a^{(k)}} = \{r(s_0, a^{(k)}), r(s_1, a^{(k)}), r(s_2, a^{(k)}), \dots\}$, $k = 0, 1$. Элементы этих массивов $r(s_t, a^{(k)})$ вычислены как время нахождения машин на активируемых фазой $a^{(k)} \oplus s_t$ полосах.

4. 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 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.

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 $z \in Zones = \{z^{(0)}, z^{(1)}, \dots, z^{(m)}\}$, $m \in \mathbb{N}$. 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 $I(s', t) = \{i \mid t_i < t, i\psi_t z, z\phi s'\}$.

Приведем рассуждения, исходя из которых считается функция вознаграждения. Для каждой полосы определено число машин на отрезке дороги, начинающемся с детектора и заканчиваемся 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 $R_t = \sum_{i \in I(s', t)} (t - t_i)$.

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 .

5. Задача мультиагентного обучения с подкреплением для сети 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 и переходит в следующее состояние $s_{t+1} = s'$ с вероятностью $p(s' \mid s_t, a_t)$.

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

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

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 $\langle S, A, \mathbb{P}, r \rangle$ для управления светофорным объектом, активная в начальный момент времени фаза светофорного объекта s_0 .
- Найти: управление светофорного объекта $\delta^* = \{a_t^*\}_{0 \leq t < \infty}$, которое доставит максимум функции оценки его эффективности (4).

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

Предложение 1. [?] В задаче управления фазами светофорного объекта уравнение Вальда—Беллмана имеет вид

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

Перепишем формулу (5) в итерационной записи, называемой Q -обучение. Функция суммарных вознаграждений при оптимальном управлении на шаге t имеет вид

$$V^* \left(\{s_{t'}, \delta\}_{t'=0}^{t'=t} \right) = \max_{a \in A} Q_t(s_t, a),$$

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

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

Решение задачи поиска оптимального совокупного управления светофорными объектами дорожной сети ищется, чтобы увеличить максимальное совокупное вознаграждение, определяемое функцией Q . справедливо слудующее утверждение + убрать предложение

Предложение 2. [?] Для задачи поиска оптимального управления светофорным объектом с любым количеством фаз справедливы следующие утверждения

- существует единственное точное решение;
- оценка точности приближенного решения на n -ом шаге итерации

$$\rho(Q_n, Q_0) \leq \frac{\gamma^n \rho(Q_1, Q_0)}{1 - \gamma},$$

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 $\rho(q, w) = \max_{j \in \mathbb{N}, j \leq |\mathcal{A}|+|\mathcal{S}|} |q_j - w_j|$;

- приближенное решение находится согласно формулам

$$(6) \quad V^*(s) = \max_{a \in \mathcal{A}} \lim_{l \rightarrow +\infty} Q_l(s, a),$$

$$(7) \quad a_t(s) = \arg \max_{a' \in \mathcal{A}} Q_l(s, a').$$

Схема подсчета функции оценки эффективности управления управления светофорными объектами представлена на рисунке 7. сказать пару фраз про нечеткий контроллер

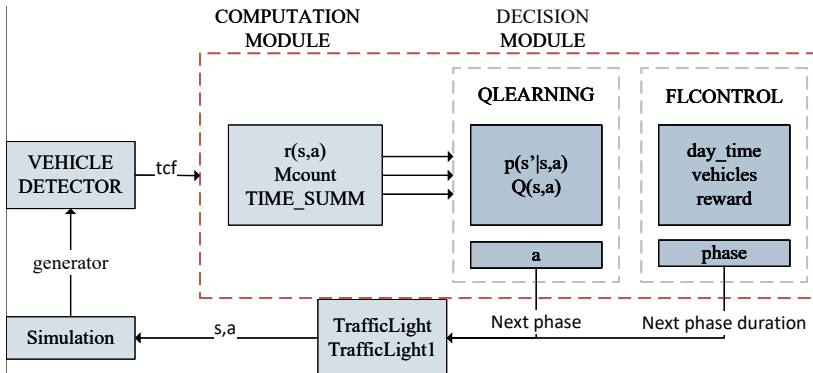


Рис. 7. Схема алгоритма MARLIN24 управления светофорными объектами

При совокупном управлении светофорными объектами в результате вызова процедуры *generator* в имитационной среде (*Simulation*) создаются машины в количествах, приближенных к реальным значениям. Далее 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 TL . На следующем шаге симуляции машины удаляются из коллекции *tcf*, при проезде через зону z . В течении периода времени *period* во вспомогательном модуле выбора управления (*DECISION MODULE*) вызывается модуль *QLEARNING*, реализующий управление согласно выбранной стратегии совокупного управления. На основе выходных данных модуля принимается решение о переключении фазы светофоров (*Next phase*).

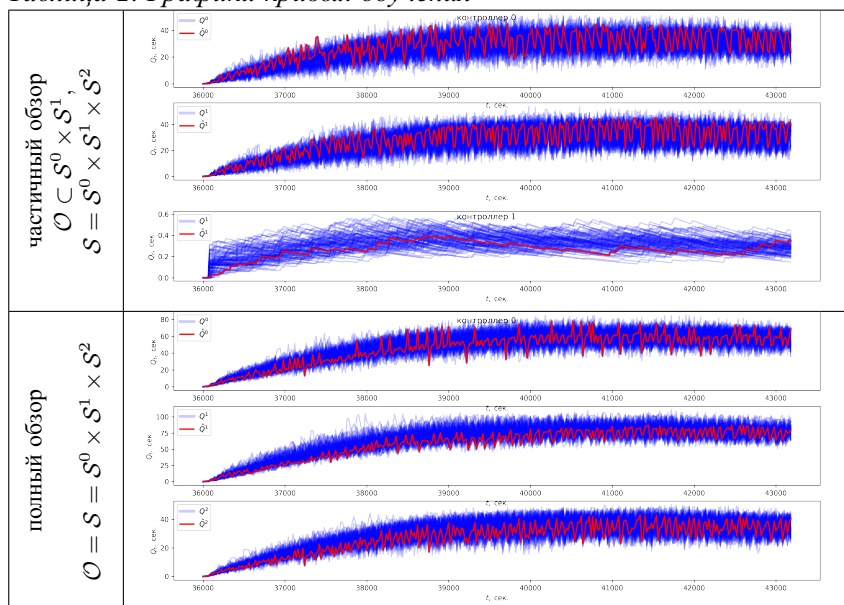
5.1. Вычислительные эксперименты и обсуждение

Для исследования представленных в работе моделей были проведены серии вычислительных экспериментов. Экспери-

менты 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 приведены в таблице 2.

Таблица 2. Графики кривых обучения



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 светофорно-

го 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.

описать словами ограниченный обзор

Сравнение комплекса MARLIN24 и АСУДД24 в работе [?] показало сопоставимые результаты (таблица 3).

Таблица 3. Сравнение показателей эффективности управления для различных моделей

Целевая функция	Ед. изм.	АСУДД24	MARLIN24	улучшение
Средняя задержка	<u>сек.</u> 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).

Viktor 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 ...*