НАЗВАНИЕ СТАТЬИ[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. Введение

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

[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*).

ной 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.

Системы, управляющие светофорными объектами, подразделяют на те, которые корректируют сигналы светофоров в реальном времени и реагируют на текущую дорожную обстановку — АСУД (адаптивные системы управления дорожным движением) и неадаптивные — те, которые работают согласно фиксированному плану управления. Неадаптивные системы светофоров переключают фазы светофоров через заранее заданное фиксированное время. В таблице 1 представлены наиболее известные АСУД.

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
авторы					

В последние годы возрос интерес к теории мультиагентного программирования, в частности, в текущей работе исследуется модель сети управляемых светофоров. В данной модели целью управления является уменьшение времени проезда транспортных

средств через выбранные участки дорожной сети. Как альтернатива к применяемым на практике АСУД, в работе предложен метод адаптивного управления, использующий мультиагентное обучение с подкреплением (Multiagent Reinforcement Learning for Integrated Network, MARLIN)[6], основанные на марковских процессах принятия решений, которые могут и не оперировать интенсивностью проезда через поперечное сечение дороги.

Во введении дается краткий обзор наиболее известных систем адаптивного управления светофорными объектами. Приводятся особенности реализации.

В параграфе 2 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 3 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 4 приведено решение задачи машинного обучения с подкреплением для нескольких агентов. Приведены критерий оптимального управления Вальда-Беллмана в предложении 1, итерационная запись для функции оценки эффективности управления V, а также приведено предложение 2 о существовании и единственности оптимального решения. В параграфе 4 выведены формулы для поиска оптимального решения (8) и (9).

В параграфах 5, 6, 7 приведено описание комплекса программных средств MARLIN24. Для исследования задачи управления светофорными объектами был разработан комплекс программных средств MARLIN24 на языке Python3. Структура комплекса программных средств MARLIN24 приведена на рисунке 1.

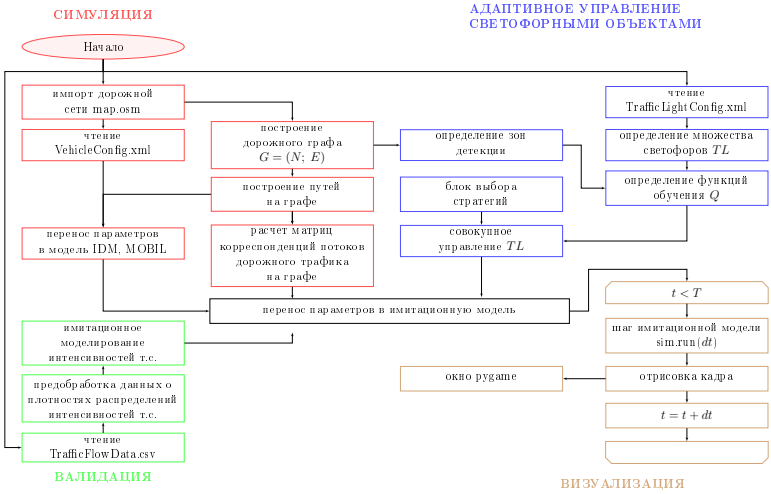


Рис. 1. Структура комплекса программных средств MARLIN24

В рамках вычислительных экспериментов в **параграфе 8** были построены кривые обучения функции оценки эффективности управления, а также дана интерпретация полученному оптимальному управлению.

2. Описание задачи для одного агента

Как правило, обучение с подкреплением используется для одного агента в среде, чтобы максимизировать его долгосрочную награду. В работе исследуется модель обучения с подкреплением нескольких агентов (светофорных объектов) в стохастической среде. Обучение с подкреплением сводится к оптимальному сопоставлению агентом действия a состоянию среды s. Предложенная математическая модель процесса выбора фазы светофора учитывает текущее расположение светофоров и их загрузку и позволяет сформулировать оптимизационные задачи, целью которых является минимизация задержки движения автомобилей. Отметим, что структура мультиагентной системы обеспечивает наиболее эффективное распараллеливание всей задачи на подзадачи.

Введём следующие обозначения: $\mathcal{S}$ — дискретное пространство состояний, $\mathcal{A}$ — дискретное пространство действий агента, $\delta\colon\ \mathcal{S} \to \mathcal{A}$ — управление сменой фаз светофора. Фазы светофора меняются последовательно. Полагаем, что множество $\mathcal{S}$ есть кольцо классов вычетов по модулю n — $\mathbb{Z}_n = \langle \mathbb{Z}, +, \cdot \rangle$, где $|\,\mathcal{S}\,| = n$ — количество классов $\mathbb{Z}_n$.

Характеристика кольца вычетов $\mathbb{Z}_n$ равна n и в данной модели отражает число смен фазы светофора до её возвращения к начальному значению, следовательно, число действий $|\mathcal{A}|$ до возвращения в начальное состояние светофора равно n.

Если агент находится в состоянии s, то при выборе действия $a = \delta(s)$ новое состояние s' определяется формулой

$$s' = (a \oplus s) = (a + s) \bmod |\mathcal{S}|. \tag{1}$$

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.

Элементы матрицы переходов между состояниями $\mathbb{P}$ определяются вероятностями вида $p(s' \mid s, a)$, где s — текущее состояние среды, a — текущее действие среды, s' — следующее состояние среды и определяется по формуле (1). Состояние s' всецело определяется текущим состоянием s и выбранным действием a.

Опишем поведение агентов с помощью марковского процесса принятия решений $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, r \rangle$ [10].

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 $\delta = \{a_t, 0 \leqslant t < \infty\}$ получается как кумулятивная функция:

$$V(\{s_t, \delta_t\}) = \lim_{T \to \infty} V(\mathcal{T}) = \sum_{t=0}^{\infty} \gamma^t r\left(s_{t+1} \mid s_t, \delta_t\right) = \\ = \sum_{t=0}^{\infty} \gamma^t r\left(s_t, \delta_t\right) = \lim_{T \to \infty} \mathbb{E}_{\mathcal{T}} \sum_{t=0}^{T} \gamma^t R_t, \tag{2}$$

где величина γ, $0 < \gamma < 1$, называется коэффициентом переоценки и показывает во сколько раз уменьшается отложенное вознаграждение за один временной шаг [3]. Переоценка задает приоритет получения награды в ближайшее время перед получением той же награды через некоторое время. Математически смысл условия $0 < \gamma < 1$ состоит в том, чтобы гарантировать ограниченность функционала V. Под отложенным до момента времени t вознаграждением принято понимать число $\gamma^t r_t$. Далее предел в выражении (2) будем опускать.

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 (2).

3. Описание задачи для нескольких агентов

Расширим задачу на сеть из K светофорных объектов. Утверждения сформулированные выше справедливы для K агентов. Далее для удобства записи будем использовать множество из двух агентов. Имеем множество из двух агентов, для каждого из которых описаны множество состояний $\mathcal{S}^0, \mathcal{S}^1$ и множество решений $\mathcal{A}^0, \mathcal{A}^1$. Обозначим $\mathbf{s}_t = \{s_t^1, s_t^2\} \in \mathcal{S}^0 \times \mathcal{S}^1$ как совокупное состояние среды в момент времени t, $\mathbf{a}_t = \{a_t^0, a_t^1\} \in \mathcal{A}^0 \times \mathcal{A}^1$ как совокупное управление в момент времени t.

Отметим, что смена фазы любым агентом приводит к изменению общего состояния среды $\mathbf{s}$.

Пример 1. В таблице 2 представлен пример распределения условных вероятностей перехода из состояния $\mathbf{s}$ в состояние $\mathbf{s}'$ с учетом выбора действия $\mathbf{a}$.

Таблица 2. Распределение $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{a})$

$\mathbf{a}$	$\mathbf{s}$ \ $\mathbf{s}'$	$\mathbf{s}^{(0)}$	$\mathbf{s}^{(1)}$	$\mathbf{s}^{(2)}$	$\mathbf{s}^{(3)}$
$\mathbf{a}^{(0)}$	$\mathbf{s}^{(0)}$	1	0	0	0
	$\mathbf{s}^{(1)}$	0	1	0	0
	$\mathbf{s}^{(2)}$	0	0	1	0
	$\mathbf{s}^{(3)}$	0	0	0	1
$\mathbf{a}^{(1)}$	$\mathbf{s}^{(0)}$	0	1	0	0
	$\mathbf{s}^{(1)}$	0	0	1	0
	$\mathbf{s}^{(2)}$	0	0	0	1
	$\mathbf{s}^{(3)}$	1	0	0	0

$\mathbf{a}$	$\mathbf{s}$ \ $\mathbf{s}'$	$\mathbf{s}^{(0)}$	$\mathbf{s}^{(1)}$	$\mathbf{s}^{(2)}$	$\mathbf{s}^{(3)}$
$\mathbf{a}^{(2)}$	$\mathbf{s}^{(0)}$	0	0	1	0
	$\mathbf{s}^{(1)}$	0	0	0	1
	$\mathbf{s}^{(2)}$	1	0	0	0
	$\mathbf{s}^{(3)}$	0	1	0	0
$\mathbf{a}^{(3)}$	$\mathbf{s}^{(0)}$	0	0	0	1
	$\mathbf{s}^{(1)}$	1	0	0	0
	$\mathbf{s}^{(2)}$	0	1	0	0
	$\mathbf{s}^{(3)}$	0	0	1	0

Также переходы между состояниями, соответствующие рассматриваемой цепи, можно графически изобразить в виде стохастического графа на рисунке 2, где используется нумерация в двоичной системе счисления согласно формуле (3).

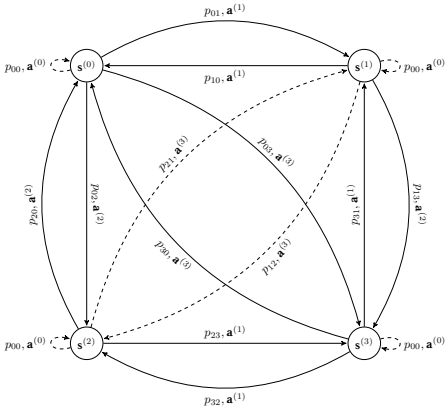


Рис. 2. Стохастический граф управляемого процесса смены фаз. Здесь s_l отражает совокупное состояние светофоров l (l =0b00, 0b01, 0b10, 0b11). При действии $\boldsymbol{a}^{(m)}$ (m =0b00, 0b01, 0b10, 0b11) процесс перейдет в состояние $n = l \oplus m$, с вероятностью p_{ln}.

•

Для случая, когда рассматривается сеть светофорных объектов с двумя состояниями, следующую фазу светофора можно найти с помощью операции «исключающего или»:

$$\mathbf{s}' = \mathbf{a} \oplus \mathbf{s}. \tag{3}$$

Введем обозначение двоичного литерала из языка программирования Python3. Согласно ей, в записи перед номером ставится префикс $0b$ (binary). Например, число 2 в двоичной записи будет представлено как $0b10$. В случае для действия $\mathbf{a} = 2$ можно провести следующие рассуждения: с заменой системы счисления и сопоставлением вектору из $\mathcal{A}^0 \times \mathcal{A}^1$. Действия $a^0 = 1$ и $a^1 = 0$ однозначно задают совместное действие $\mathbf{a} = \left(a^0, a^1\right) = 0b10$.

В момент времени t функция наград $r(\mathbf{s}_t, \mathbf{a}_t)$ определяется как сумма функций наград каждого агента по отдельности $r(\mathbf{s}_t, \mathbf{a}_t) = r(s_t^0, a_t^0) + r(s_t^1, a_t^1)$, которые в свою очередь рассчитываются согласно формуле (2). В таком случае функция оценки

эффективности управления примет вид

$$V(\mathbf{s}) = \sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t, \mathbf{a}_t), \tag{4}$$

где $0 \leqslant \gamma \leqslant 1$ — коэффициент переоценки.

Управление фазами двух светофорных объектов находится как решение задачи обучения с подкреплением (Multiagent Reinforcement Learning for Integrated Network, MARLIN) и формулируется следующим образом.

Дано: марковский процесс принятия решения $\langle \mathcal{S}^0 \times \mathcal{S}^1, \mathcal{A}^0 \times \mathcal{A}^1, \mathbb{P}, r \rangle$ для управления дорожной сетью из двух светофорных объектов, активные в начальный момент времени фазы $\mathbf{s}_0$.

Найти: управление светофорными объектами $\delta^* = \{\mathbf{a}_t^*\}_{0 \leqslant t < \infty}$, которое доставит максимум функции оценки эффективности (4).

4. Решение задачи

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

***Предложение* 1.** *[15] В задаче управления фазами светофорного объекта уравнение Вальда—Беллмана имеет вид*

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

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

$$V^*\left(\{\mathbf{s}_{t'}, \delta\}_{t'=0}^{t'=t}\right) = \max_{\mathbf{a} \in \mathcal{A}} Q_t(\mathbf{s}_t, \mathbf{a}),$$

$$Q_t(\mathbf{s}_t, \mathbf{a}) = \sum_{\mathbf{s}_{t+} \in \mathcal{S}} p(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}) \Big(r(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}) + \gamma \max_{\mathbf{a}' \in \mathcal{A}} Q_{t-1}(\mathbf{s}_{t-1}, \mathbf{a}') \Big).$$

Для совместных действий и состояний итерационно заданная функция Q_t имеет вид:

$$Q_t(\mathbf{s}, \mathbf{a}) = \sum_{\mathbf{s}'} p(\mathbf{s}' \mid \mathbf{s}, \mathbf{a}) r(\mathbf{s}, \mathbf{a}) + \gamma \sum_{\mathbf{s}'} p(\mathbf{s}' \mid \mathbf{s}, \mathbf{a}) V^* \left(\mathbf{s}'\right).$$

Условимся, что записи $Q_t(\mathbf{s}, a, a')$ и $Q_t(\mathbf{s}, \mathbf{a})$ тождественны.

Считаем, что нам известно состояние среды s_{t+1} и оптимальное управление a_{t+1} на шаге $t+1$, соответствующий итерации $\hat{t}$, и условимся, что итерация Q идет по индексу $\hat{t}$, тогда

$$\begin{aligned} Q_t(\mathbf{s}, \mathbf{a}) &= \underbrace{p(s_{t+1}|s, a)}_{\alpha_{\hat{t}}} \Big(r_{t+1} + \gamma V^*(\mathbf{s}_{t+1}) \Big) + \\ &+ \underbrace{\sum_{\mathbf{s}' \in S/\mathbf{s}_{t+1}} p(\mathbf{s}'|\mathbf{s}, \mathbf{a})}_{1-\alpha_{\hat{t}}} \Big(r(\mathbf{s}'|\mathbf{s}, \mathbf{a}) + \gamma V(\mathbf{s}') \Big) = \\ &= \alpha_{\hat{t}} \Big(r_{\hat{t}} + \gamma \max_{\mathbf{s}'} Q_{\hat{t}}(\mathbf{s}_{t+1}, \mathbf{s}') \Big) + \Big(1 - \alpha_{\hat{t}}\Big) Q_{\hat{t}}(\mathbf{s}, \mathbf{a}) = \\ &= Q_{\hat{t}+1}(\mathbf{s}, \mathbf{a}). \end{aligned} \tag{6}$$

Сведем задачу поиска максимизирующего совместного действия агентов k, j к случаю для одного агента, совершающего действие $\mathbf{a}^{kj}$. Функция Q для агента k определяется как

$$\begin{aligned} Q^k_{\hat{t}}(\mathbf{s}, a^k) = \sum_{a^j \in \mathcal{A}^j} p(a^j \mid s, a^k) \Big(r(\mathbf{s}, \mathbf{a}^{kj}) + \\ + \gamma \sum_{\mathbf{s}'} p\left(\mathbf{s}, \mathbf{a}^{kj}\right) \max_{\mathbf{a}'} Q_{\hat{t}-1}(\mathbf{s}', \mathbf{a}') \Big). \end{aligned}$$

Оптимальное управление для фиксированного агента k будем искать как решение задачи MARL в виде

$$a^k_{\hat{t}} = \arg \max_{a^k \in \mathcal{A}^k} Q^k_{\hat{t}}(\mathbf{s}_{\hat{t}}, a^k). \tag{7}$$

Вероятность $p(a^j \mid s, a^k)$ — это вероятность того, что агент j выберет действие $\mathbf{a}^j$ с учётом текущего совместного состояния $\mathbf{s}$ и выбранного агентом k действия $\mathbf{a}^k$.

Решение задачи поиска оптимального совокупного управления светофорными объектами дорожной сети ищется, чтобы

увеличить максимальное совокупное вознаграждение, определяемое функцией Q по итерационной формуле (6), (7). Вероятность $p\left(\mathbf{s}' \mid \mathbf{s}, \mathbf{a}^{kj}\right)$ для такого максимизирующего действия оценивается параметром α. Для каждого многомерного решения верны утверждения (1), (2), а также, они верны в случае нескольких координированных агентов. Таким образом, задача совокупного управления несколькими агентами сводится к задаче управления одним агентом и справедливо утверждение об единственности решения.

***Предложение* 2.** *[16] 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*

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

где $Q_t \in \mathbb{R}_{\infty}^{|\mathcal{A}|+|\mathcal{S}|}$ — вектора значений $Q(s,a)$ на шаге t, $\forall q, w \in \mathbb{R}_{\infty}^{|\mathcal{A}|+|\mathcal{S}|}$ определена функция $\rho(q,w) = \max\limits_{1 \leqslant j \leqslant |A|+|S|} |q_j - w_j|$;

- *приближенное решение находится согласно формулам*

$$V^*(s) = \max_{a \in \mathcal{A}} \lim_{t \to +\infty} Q_t(s,a), \tag{8}$$

$$a_t(s) = \arg\max_{a' \in \mathcal{A}} Q_t(s,a'). \tag{9}$$

5. Модуль симуляции

Согласно схеме на рисунке 1, в модуле симуляции с помощью пакета osmnx из Open Street Map был импортирован мультиграф **map.osm**, задающий параметры и координаты в координатной системе WGS84/UTM (World Geodetic System 1984/Universal Transverse Mercator). Далее, по графу строится дорожная сеть, строятся маршруты транспорта с использованием пакета networkx[5]. Поскольку построение маршрутов транспорта

является весьма трудоемкой задачей, то для ускорения инициализации модели используется многопоточная реализация алгоритма построения маршрутов при помощи модуля threading[4] (рис. 1).

При построении дорожной, сети вершинам присваивается уникальный номер osmid. Каждая вершина представляет узел (перекресток, регулируемый пешеходный переход) или место существенного изменения характеристик дороги. Каждое ребро соответствует реальному участку дороги без перекрестков. В ребрах хранится информация о соединяемых ими вершинах, их координатах, длине дороги, количестве полос, названии улицы, а также других параметрах из табл. 3.

Таблица 3. Импортируемые параметры дорог из OSM

Параметр	**Принимаемые значения**	**Описание**
lanes	int	число полос.
highway	residential primary secondary terliary unclassified	внутри жилых зон, федерального, областного, местного значения, образующие соединительную сеть дорог
lanes	str	название улицы
oneway	bool	дороги, которые проходят внутри жилых зон
reversed	bool	направление движения дорог
length	float	Длинна дороги.

Для вручную отмеченных светофорных объектов были построены матрицы корреспонденции потоков дорожного трафика, отвечающие его за распределение по доступным направлениям, а также из конфигурационного файла TrafficLightConfig.xml была перенесена информация об активируемых фазах и активируемых направлениях светофорных объектов.

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 и остальные параметры, описанные в таблице 3.

В модели IDM[7] транспортные средства рассматриваются как индивидуальные сущности, обладающие своими характери-

стиками и поведением. Рассматриваемая модель 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 автомобиля, 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.

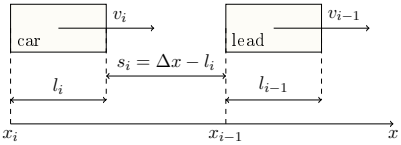


Рис. 3. Иллюстрация модели IDM

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

$$(10) \quad \begin{cases} \dfrac{dv_i}{dt} = \underbrace{a_i\left(1 - \left(\dfrac{v_i}{v_{0,i}}\right)^{\delta}\right)}_{a_{\text{free}}} - \underbrace{a_i\left(\dfrac{s^*\left(v_i, \Delta v_i\right)}{s_i}\right)^2}_{a_{\text{deceleration}}}, \\ s^*\left(v_i, \Delta v_i\right) = s_{0,i} + v_i T_i + \dfrac{v_i \Delta v_i}{\sqrt{2 a_i b_i}}. \end{cases}$$

При имитационном моделировании для нахождения значений скорости и ускорения будем пользоваться формулами, вытекающими из численного метода «пристрелки» [12]:

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

Шаг симуляции dt выбирается как шаг по времени при численном решении системы (10).

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). Данная модель перестроения вместе с моделью IDM была разработана Дириком Гельфандом, Мартином Трейбером и Арнольдом Кухнем [14] в 1999 году и предназначалась для анализа и улучшения эффективности движения автомобилей на перекрестках.

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.

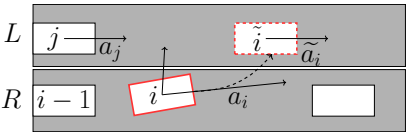


Рис. 4. Модель смены полосы MOBIL

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:

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

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

Также следует учитывать, что при перестроении i на соседнюю полосу транспортные средства j и следующие за ним не должны двигаться с коэффициентом торможения меньше, чем b_{safe}. Поскольку в модели IDM скорости, а, следовательно, и ускорения связаны формулой (10) и изменяются последовательно от

лидирующего транспортного средства к последующему, то описать такое замедление можно формулой: $\widetilde{a}_j \geqslant -b_{\text{safe}}$.

6. Модуль валидации

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 [18] при описании интенсивности транспортного потока использовалась статистка, описывающая количество машин, в работе [19] 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 двупараметрической копулы Маршалла-Олкина [21] для совместного распределения временных интервалов появления автомобилей.

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. По теореме Скляра [20] совместную функцию распределения можно представить копулой C

$$H_{XY}(x, y) = C(F(x), G(y)), \qquad \forall x, y \in \mathbb{R}. \tag{11}$$

Далее будем использовать двупараметрическую копулу Маршалла-Олкина [21, 8] с коэффициентами $0 \leqslant \alpha, \beta \leqslant 1$

$$C_{\alpha,\beta}(u, v) = uv \min\left(u^{-\alpha}, v^{-\beta}\right). \tag{12}$$

Этап предобработки состоит в оценивании плотности распределений случайных величин X и Y, описывающих число де-

тектируемых 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 [9]. Далее формулируется упрощающее предположение о том, что каждая из рассматриваемых случайных величин представима в виде смеси нормальных распределений. С использованием EM-алгоритма [11], 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):

этап 1: моделируем равномерно распределенные на отрезке $[0;\ 1]$ случайные величины l, w;

этап 2: если выполняется $\sum\limits_{m=1}^{M} p_m < l < \sum\limits_{m=1}^{M+1} p_m$, то функция распределения M-й компоненты смеси $F = F_M$;

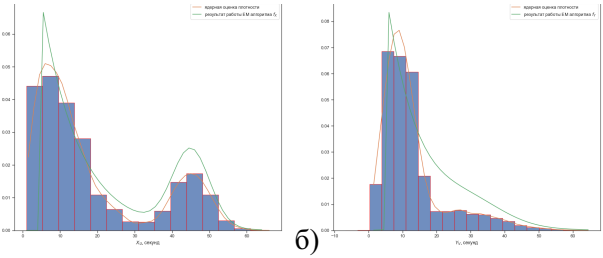
этап 3: аналогично этапу 2, для w определяем функцию распределения компоненты смеси G;

этап 4: моделируем случайные величины (U, V) согласно [20];

этап 5: восстанавливаем U-квантиль X_U распределения F;

этап 6: восстанавливаем V-квантиль Y_V распределения G.

Пример 2. 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) соответственно.



а) б)

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.

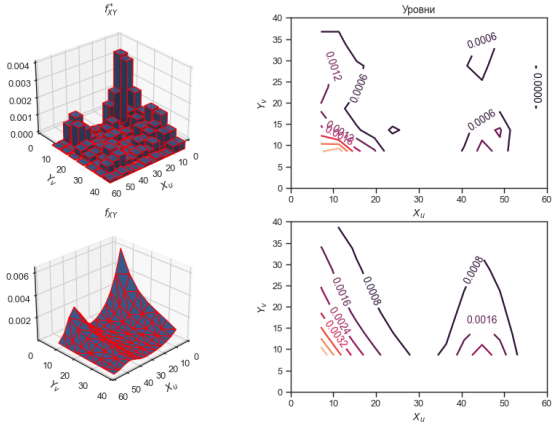


Рис. 6. Сравнение эмпирической f^*_{XY} и теоретической f_{XY} функции плотности распределения (X, Y) при моделировании значений (U, V) для копулы Маршалла-Олкина с параметрами $\alpha = 0.9, \beta = 0.25$

7. Модуль адаптивного управления светофорными 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 (2) 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 $\hat{p}(s,a)$ 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 $r_{a^{(k)}} = \left\{r(s_0, a^{(k)}), r(s_1, a^{(k)}), r(s_2, a^{(k)}), \dots\right\}$, $k = 0, 1$. Элементы этих массивов $r\left(s_t, a^{(k)}\right)$ вычислены как время нахождения машин на активируемых фазой $a^{(k)} \oplus s_t$ полосах. Опишем подробнее процесс подсчета $\hat{\pi}(a|s)$, опираясь на схему на рисунке 1.

Согласно схеме 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 разработанный 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 транспортного средства $i \in I \subset \mathbb{N}$ в зоне $z \in Zones = \left\{z^{(0)}, z^{(1)}, \dots, z^{(m)}\right\}, m \in \mathbb{N}$. Отметим, что при имитационном моделировании псевдослучайная интенсивность движения транс-

портных 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 $I(s', t) = \{i|\ \ t_i < t, i\psi_t z, z\phi s'\}$. d Пусть в момент времени t активна фаза светофора $s \in \mathcal{S}$ и управление $a = \delta(s) \in \mathcal{A}$, задержка на фазе $R = \sum\limits_{i \in I(s',t)} (t - t_i)$.

Далее для построенного множества светофорных объектов TL и зон детекции z определяются функция наград $r(s, a) = \mathbb{E}R$, число проехавших машин (Mcount), суммарное время проезда через детектируемые участки дорожной сети (TIME_SUMM) и обучающие функции Q.

Схема подсчета функции оценки эффективности управления управления светофорными объектами представлена на рисунке 7.

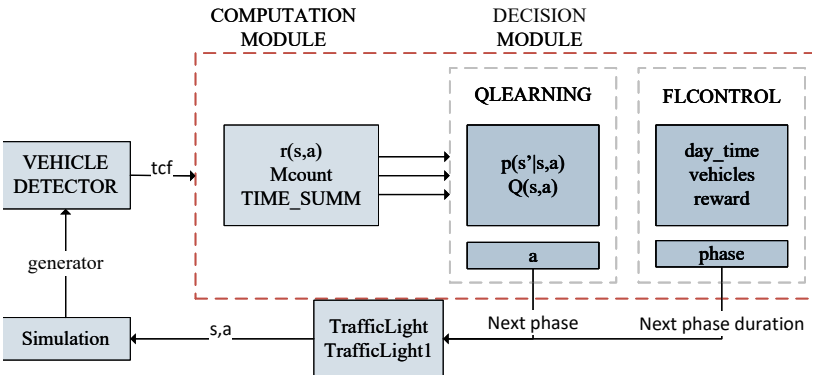


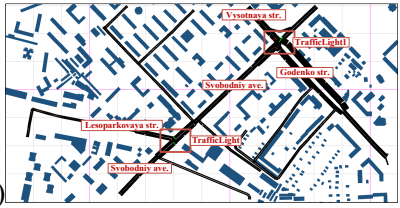
Рис. 7. Схема алгоритма MARLIN24 управления светофорными объектами

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).

8. 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.

а) 

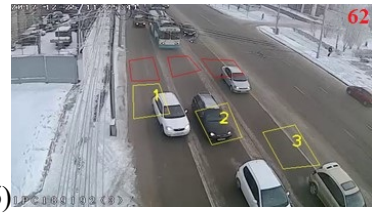
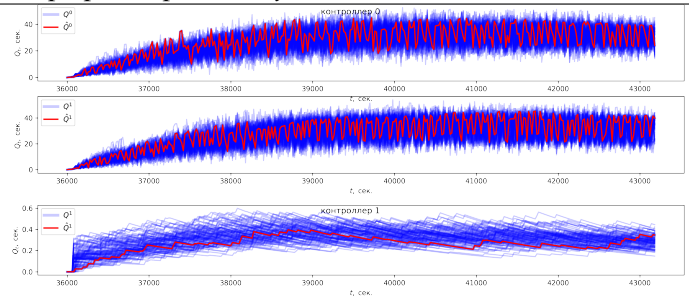
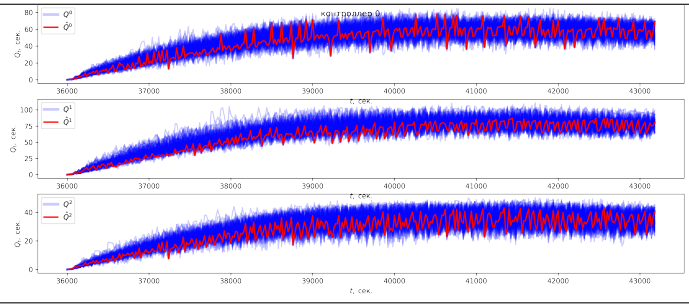
б) 

Рис. 8. а)модель рассматриваемого участка дорожной сети, реализованная в системе MARLIN24; б)зоны дороги (жёлтый цвет), фиксируемые оптическими датчиками

Вычислительный эксперимент состоит в

- демонстрации сходимости функции оценки эффективности управления Q к $\hat{Q}$ при равновесной по Нэшу стратегии в таблице 4;
- сравнении графиков кривой обучения Q^i для светофорных объектов $0, 1$ при условии ограниченного обзора пространства состояний $\mathcal{O}$ и при условии, когда оно совпадает с полным пространством состояний $\mathcal{S}$;
- сравнении показателей эффективности управления для АСУДД24 и MARLIN24.

Таблица 4. Графики кривых обучения

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

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 при машинном обучении время нахождения было усреднено до общего времени и составило

40 секунд. Сравнение комплекса MARLIN24 и АСУДД24 в работе [17] показало сопоставимые результаты (таблица 5). В рам-

Таблица 5. 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%

ках вычислительных экспериментов было проведено сравнение кривых обучения агентов на протяжении 100 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.

Литература

1. *Транспортная стратегия Российской Федерации, утверждена распоряжением Правительства Российской Федерации от 22 ноября 2008 года №1734-р* [Электронный ресурс]. – Режим доступа: http://mintrans.gov.ru. – Дата обращения: 25.06.2024.
2. *АСУДД «МИКРО–М»* [Электронный ресурс]. – Режим доступа: http://asud55.ru/archives/1346. – Дата обращения: 25.06.2024.
3. *Иванов, С. Конспект по обучению с подкреплением* [Электронный ресурс]. – Режим доступа: https://arxiv.org/abs/2201.09746. – Дата обращения: 25.06.2024.
4. *threading — Thread-based parallelism* [Электронный ресурс]. – Режим доступа: https://docs.python.org/3/library/threading.html. – Дата обращения: 25.06.2024.
5. *NetworkX* [Электронный ресурс]. – Режим доступа: https://networkx.org/. – Дата обращения: 25.06.2024.
6. EL-TANTAWY, S., ABDULHAI, B. *Towards multi-agent reinforcement learning for integrated network of optimal traffic controllers (MARLIN-OTC)* // Transportation Letters: The International Journal of Transportation Research. – 2010. – T. 2. – C. 89–110.
7. TREIBER, M., HENNECKE, A., HELBING, D. *Congested traffic states in empirical observations and microscopic simulations* // Transportation Physics Reviews E. – 2000. – T. 62. – C. 1805–1824.
8. QUESADA-MOLINA, J.J., RODR?GUEZ-LALLENA, J.A. *Bivariate copulas with quadratic sections* // Journal of Nonparametric Statistics. – 1995. – C. 323–337.
9. EPANECHNIKOV, V.A. *Non-Parametric Estimation of a Multivariate Probability Density* // Theory of Probability and

Its Applications. – 1969. – С. 153–158.
10. ГАСНИКОВ, А.В., ГОРБУНОВ, Э.А. и др. *Лекции по случайным процессам: учебное пособие.* – М.: МФТИ, 2019. – 208 с.
11. BISHOP, C.M. *Pattern Recognition and Machine Learning.* – Springer, 2007. – 738 с.
12. ЗАЛИЗНЯК, В.Е. *Численные методы. Основы научных вычислений: учебное пособие для бакалавров.* – М.: ЮРАЙТ, 2012. – 356 с.
13. ВОЙТИШЕК, А.В. *Основы метода Монте-Карло: Учебное пособие.* – Новосибирск: НГУ, 2010. – 108 с.
14. TREIBER, M., HELBING, D. *Realistische Mikrosimulation von Stra?enverkehr mit einem einfachen Modell* // 16. Symposium "Simulationstechnik ASIM 2002"Rostock. – 2002. – С. 514–520.
15. TISLENKO T.I., SEMENOVA D.V., SERGEEVA N.A., GOLDENOK E.E., KONONOVA N.V. *Multiagent Reinforcement Learning for Integrated Network: Applying to a Part of the Road Network of Krasnoyarsk City* // IEEE 16th International Conference on Application of Information and Communication Technologies (AICT). – 2022. – С. 1–5.
16. TISLENKO T.I., SEMENOVA D.V., SOLDATENKO A.A. *Modeling and Comparison of Different Management Approaches on the Intersections Network* // 2023 IEEE 26th International Conference, Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN). – 2023. – С. 25–29.
17. ТИСЛЕНКО Т.И. *О двух подходах адаптивного управления светофорными объектами участка дорожной сети г. Красноярска* // Информационные технологии и математическое моделирование (ИТММ-2023): Материалы XXII Международной конференции имени А.Ф. Терпугова. – Томск: Национальный исследовательский Томский государственный университет, 2023. – С. 273–278.
18. ТИСЛЕНКО Т.И. *Моделирование интенсивностей транс-*

портных 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.

19. XXXX *xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx* // XXXXXXXXXXXXXXXX. – 21.10.03. – .
20. NELSEN R.B. *An Introduction to Copulas* – New York: Springer, 2006. – 270 p.
21. MARSHALL A. *Copulas, Marginals, and Joint Distributions* // Lecture Notes-Monograph Series. Distributions with Fixed Marginals and Related Topics. – 1996. – No. 28. – С. 213–222.
22. АДОРНО Т.В. *К логике социальных наук* // Вопросы философии. – 1992. – №10. – С. 76–86.
23. БОЛЬШОВ П.П. *Название книги*. – М.: Изд-во «УБС», 2017. – 142 с.
24. ГЛУХОВ В.А. *Исследование, разработка и построение системы электронной доставки документов в библиотеке*: Автореф. дис. канд. техн. наук. – Новосибирск, 2000. – 18 с.
25. ЕСЬКОВ Д.Н., БОНШТЕДТ Б.Э., КОРЕШЕВ С.Н., ЛЕБЕДЕВА Г.И., СЕРЕГИН А.Г. *Оптико-электронный аппарат* // Патент России №2122745. – 1998. – Бюл. №33.
26. КОРНИЛОВ В.И. *Турбулентный пограничный слой на теле вращения при периодическом вдуве/отсосе* // Теплофизика и аэромеханика. – 2006. – Т. 13, №3. – С. 369–385.
27. КУЗНЕЦОВ А.Ю. *Консорциум – механизм организации подписки на электронные ресурсы* // Российский фонд фундаментальных исследований: десять лет служения российской науке. – М.: Научный мир, 2003. – С. 340–342.
28. ЛИТЧФОРД Е.У. *С Белой Армией по Сибири* [Электронный ресурс] // Восточный фронт Армии Генерала А.В. Колчака: сайт. – URL: http://east-front.narod.ru/memo/latchford.htm (дата обращения: 23.08.2007).
29. ЛОГИНОВА Л.Г. *Сущность результата дополнительно-*

го образования детей // Образование: исследовано в мире: междунар. науч. пед. интернет-журн. – 21.10.03. – URL: http://www.oim.ru/reader.asp?nomer=366 (дата обращения: 17.04.07).

30. *Официальные периодические издания : электронный путеводитель* / Рос. нац. б-ка, Центр правовой информации. [СПб.], 2005–2007. – URL: http://www.nlr.ru/lawcenter/izd/index.html (дата обращения: 18.01.2007).
31. Патент РФ №2000130511/28, 04.12.2000.
32. РАЙЗБЕРГ Б.А., ЛОЗОВСКИЙ Л.Ш., СТАРОДУБЦЕВА Е.Б. *Современный экономический словарь*. 5-е изд., перераб. и доп. – М.:ИНФРА-М, 2006. – 494 с.
33. *Рынок тренингов Новосибирска: своя игра* [Электронный ресурс]. – Режим доступа: http://nsk.adme.ru/news/2006/07/03/2121.html.
34. ТАРАСОВА В.И. *Политическая история Латинской Америки*: учеб. для вузов. 2-е изд. – М.: Проспект, 2006. – С. 305–412.
35. ФЕНУХИН В. И. *Этнополитические конфликты в современной России: на примере Северо-Кавказского региона*: дис. канд. полит. наук. – М., 2002. – С. 54–55.
36. *Философия культуры и философия науки: проблемы и гипотезы*: межвуз. сб. науч. тр. / Сарат. гос. ун-т; [под ред. С.Ф. Мартыновича]. – Саратов: Изд-во Сарат. ун-та, 1999. – 199 с.
37. *Экономика и политика России и государств ближнего зарубежья*: аналит. обзор, апр. 2007 / Рос. акад. наук, Ин-т мировой экономики и междунар. отношений. – М.: ИМЭМО, 2007. – 39 с.
38. CRAWFORD P.J., BARRETT T.P. *The reference librarian and the business professor: a strategic alliance that works* // Ref. Libr. – 1997. – Vol. 3, No. 58. – P. 75–85.
39. http://www.nlr.ru/index.html (дата обращения: 20.02.2007).

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: 150–200 words. Describes the standards of articles' formatting for "Large systems control" papers collection. Provides the examples for typical elements of an article. 150–200 words. Describes the standards of articles' formatting for "Large systems control" papers collection. Provides the examples for typical elements of an article. 150–200 words. Describes the standards of articles' formatting for "Large systems control" papers collection. Provides the examples for typical elements of an article. 150–200 words. Describes the standards of articles' formatting for "Large systems control" papers collection. Provides the examples for typical elements of an article.

Keywords: large systems control, electronic scientific publication, article formatting template.

УДК ...
ББК ...

Статья представлена к публикации
членом редакционной коллегии ...

Поступила в редакцию ...
Дата опубликования ...