

## ЧИСЛЕННЫЕ МЕТОДЫ И АНАЛИЗ ДАННЫХ

### Детерминированная прогнозная модель управления сигналами светофоров в интеллектуальных транспортных и геоинформационных системах

В.В. Мясников<sup>1,2</sup>, А.А. Агафонов<sup>1</sup>, А.С. Юмаганов<sup>1</sup>

<sup>1</sup> Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, Россия, г. Самара, Московское шоссе, д. 34;

<sup>2</sup> ИСОИ РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, 443001, Россия, г. Самара, ул. Молодогвардейская, д. 151

#### Аннотация

В работе предлагается метод адаптивного управления сигналами/фазами светофоров в интеллектуальных транспортных и геоинформационных системах, основанный на детерминированной прогнозном модели. Под детерминированной прогнозном моделью в работе понимается набор явных аналитических закономерностей и /или операций, связывающих информацию о движении транспортных средств в окрестности конкретного перекрёстка, с данными о прогнозируемом «потоке» транспортных средств через перекрёсток за одну конкретную фазу светофорного цикла. Предлагаемый метод управления основывается на выборе фазы светофорного цикла, прогнозируемый поток для которой оказывается максимален. Таким образом, метод обеспечивает управление сигналами / фазами светофоров на основе данных о движении транспорта, включая данные с подключенных и автономных транспортных средств. Экспериментальные исследования были проведены в системе микроскопического моделирования транспортных потоков SUMO. Представлено сравнение предложенного метода с решениями, обладающими лучшими в своём классе показателями эффективности: эмпирическими алгоритмами управления и методами управления на основе обучения с подкреплением. Показано преимущество предложенного метода, определены направления дальнейших исследований.

**Ключевые слова:** анализ данных, интеллектуальная транспортная система, управление сигналами светофора, детерминированная модель, обучение с подкреплением, подключенные и автономные транспортные средства.

**Цитирование:** Мясников, В.В. Детерминированная прогнозная модель управления сигналами светофоров в интеллектуальных транспортных и геоинформационных системах / В.В. Мясников, А.А. Агафонов, А.С. Юмаганов // Компьютерная оптика. – 2021. – Т. 45, № 6. – С. 917-925. – DOI: 10.18287/2412-6179-CO-1031.

**Citation:** Myasnikov VV, Agafonov AA, Yumaganov AS. A deterministic predictive traffic signal control model in intelligent transportation and geoinformation systems. Computer Optics 2021; 45(6): 917-925. DOI: 10.18287/2412-6179-CO-1031.

#### Введение

Транспортные системы являются неотъемлемой частью человеческой деятельности. Постоянный рост дорожного трафика, особенно в крупных городах, приводит к значительному увеличению затрат (времени, стоимости) на осуществление корреспонденций участниками дорожного движения, а также к увеличению вредных выбросов в атмосферу и ухудшению экологической обстановки. Существуют оценки такого негативного влияния в различных странах: на основании данных [1] в США потери от неэффективности транспортной инфраструктуры, вызывающей заторы, оцениваются в 160 млрд. долларов при избыточных затратах 3,1 млрд. галонов топлива; в Европейском Союзе оценка потерь составляет 1% ВВП [2]. Важность решения указанной проблемы в России подчёркивает утверждённая Указом Президента Российской Феде-

рации от 1 декабря 2016 г. № 642 Стратегия научно-технологического развития Российской Федерации, в частности, направление «Связанность территории Российской Федерации за счёт создания интеллектуальных транспортных и телекоммуникационных систем, а также занятия и удержания лидерских позиций в создании международных транспортно-логистических систем, освоении и использовании космического и воздушного пространства, Мирового океана, Арктики и Антарктики».

Для улучшения ситуации во многих странах используются различные стратегии: замена классических транспортных средств на гибриды и электромобили, территориальное и временное зонирование доступных территорий передвижения и стоянки, развитие альтернативных средств передвижения, развитие средств совместного использования транспортных средств (каршеринг, информационные системы попутчиков),

оптимизация существующей транспортной инфраструктуры с целью повышения эффективности её использования и др. Среди названных путей решения направление, связанное с увеличением пропускной способности транспортной сети путём оптимизации систем управления сигналами/фазами светофоров, является одним из наиболее дешёвых и позволяющих повысить эффективность использования транспортной инфраструктуры в кратчайшие сроки [3]. Данный факт, наряду с изложенными выше, подчёркивает актуальность проблемы управления сигналами/фазами светофоров в интеллектуальных транспортных системах и системах «умный город» [4, 5].

Работа построена следующим образом. В первом параграфе представлен краткий обзор существующих методов управления сигналами/фазами светофоров в транспортных системах, указаны основные тенденции современного развития и ключевые проблемы, введены некоторые основные понятия и обозначения. Во втором параграфе представлены недостающие понятия, описан предлагаемый адаптивный метод управления сигналами/фазами светофоров, основанный на детерминированной прогнозной модели. В третьем параграфе представлены результаты экспериментов, выполненных в системе микроскопического транспортного моделирования SUMO и направленных на сравнение эффективности предложенного метода с наилучшими существующими. Завершают работу параграфы: заключение, благодарности и список использованных источников.

### 1. Современное состояние исследований

Системы адаптивного управления движением транспортных средств с помощью сигналов/тактов светофоров используются с начала 80-х годов. С этого момента широкое распространение на практике получили системы SCOOT [6–7] и SCATS [8], которые использовали информацию о числе «проехавших» мимо соответствующего детектора транспортных средств и на основании заранее разработанного эмпирического плана переключали на «следующую» фазу светофора в рамках определённого цикла работы светофора. Подобная жёсткая схема адаптации оказалась недостаточно эффективной, что привело к появлению других методов управления. Дополнительной причиной этому является также значительно возросшее количество источников данных, которые могут использоваться для решения задачи координированного управления сигналами/фазами светофоров: данные с камер видеонаблюдения, траектории движения из навигационных приложений, информация от транспортных средств, обменивающихся данными с объектами инфраструктуры (т.н. подключенные транспортные средства – CV и автономные транспортные средства – CAV). Одновременно с ростом числа источников в последнее время лавинообразно возрос и объём доступных для анализа данных, что позволяет для решения задач управ-

ления светофорами в интеллектуальных транспортных системах использовать методы машинного обучения и обработки «больших данных» [4, 5]. Ниже представлен краткий обзор основных подходов и передовых методов, используемых для управления сигналами/тактами светофоров. Предварительно дадим несколько определений, некоторые из которых известны из правил дорожного движения [9].

Под *перекрёстком* понимается «место пересечения, примыкания или разветвления дорог на одном уровне, ограниченное воображаемыми линиями, соединяющими соответственно противоположные, наиболее удалённые от центра перекрёстка начала закруглений проезжих частей». Под  *проезжей частью*  дороги понимается «элемент дороги ... предназначенный для движения безрельсовых транспортных средств». Наконец, под *полосой движения* понимается «любая из продольных полос проезжей части, обозначенная или не обозначенная разметкой и имеющая ширину, достаточную для движения автомобилей в один ряд». На рис. 1 показан перекрёсток как пересечение двух дорог с разделёнными проезжими частями, каждая из которых имеет по три полосы движения.

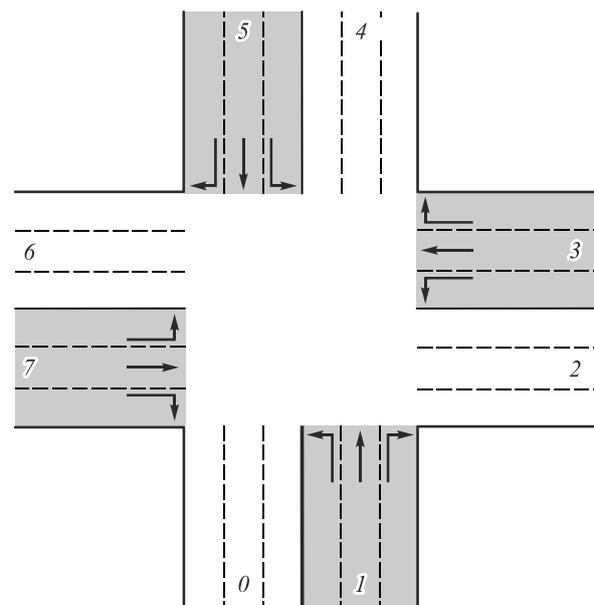


Рис. 1. Пример перекрёстка двух дорог: 8 проезжих частей, каждая из которых имеет по 3 полосы движения

Движение транспорта по перекрёстку осуществляется в соответствии с разметкой, которая определяет допустимые направления движения, и сигналами светофора(ов), которые определяют возможность перемещения по соответствующему направлению в текущий момент времени. Например, обозначение движения парой «1б» означает поворот налево с проезжей части, помеченной цифрой «1», на проезжую часть, помеченную цифрой «б». Данное движение, с учётом присутствующей разметки, допустимо только с левой полосы первой проезжей части.

Под *фазой светофора* будем понимать множество непротиворечивых сигналов светофора на перекрёстке,

то есть таких сигналов, которые не могут приводить к аварийной ситуации. Например, непротиворечивыми являются сигналы следующего множества {56, 50, 14, 12}, а противоречивыми – {14, 72}. Очевидно, что множество  $P$  возможных фаз светофоров на конкретном перекрёстке зависит от числа дорог, проезжих частей, полос и структуры перекрёстка. В частности, для перекрёстка на рис.1 можно определить 8 фаз, которые будут использоваться в данной работе:

$$P = \{ \{56, 50, 14, 12\}, \{34, 36, 71, 72\}, \{16, 52\}, \{30, 74\}, \{12, 14, 16, 70\}, \{56, 50, 52, 34\}, \{34, 36, 30, 12\}, \{70, 72, 74, 56\} \}.$$

Фазы светофорного регулирования могут выбираться как произвольно, так и в рамках некоторой предопределённой последовательности фаз – *светофорного цикла*. Одной из наиболее простых систем управления – контроллеров – является широко распространённый способ Uniform, подразумевающий переключение фаз светофора через заранее фиксированные промежутки времени из детерминированного светофорного цикла. Если выбор фазы светофорного регулирования происходит на основании данных движения транспортных средств, то говорят о *системе адаптивного управления дорожными сигналами*. При этом следует различать системы, где светофорный цикл предопределён (то есть адаптивно выбирается только момент переключения, как в вышеуказанных системах SCOOT и SCATS) и где выбор следующей фазы производится *без ограничений*. В данной работе интерес представляет именно последний тип систем управления. Системы подобного типа можно условно разделить на *эмпирические детерминированные алгоритмы* и *алгоритмы на основе машинного обучения*. Последние для управления движением активно используют *методы обучения с подкреплением* (англ.: RL – reinforcement learning). Среди них в последнее время наибольшее внимание уделяется методам с использованием глубоких нейронных сетей (DNN).

1.1. Эмпирические детерминированные методы адаптивного управления сигналами светофоров

К эмпирическим детерминированным методам адаптивного управления сигналами светофора могут быть отнесены:

- метод Webster [10],
- метод MaxPressure [11],
- метод SOTL (англ.: Self-organizing traffic lights) [12–13].

Детальное описание и сравнение указанных методов адаптивного управления сигналами/фазами светофора было представлено в работе [14]. Указанная работа *продемонстрировала преимущество метода MaxPressure, сравнение с которым и будет проведено в настоящей работе*. Для определённости дадим его краткое описание.

В качестве ключевой характеристики в методе MaxPressure выступает величина «давления» – *Pressure* – рассчитываемая для данной фазы *phase* светофора, действующей в текущий момент, следующим образом [14]:

$$Pressure(phase) = \sum_{l \in L_{phase}^{income}} n_l - \sum_{l \in L_{phase}^{outcome}} n_l. \tag{1}$$

Здесь  $L_{phase}^{income}, L_{phase}^{outcome}$  – множество входящих и исходящих полос движения на перекрёстке, для которых допустимо движение при фазе *phase*,  $l$  – обозначение некоторой полосы движения, а  $n_l$  – число транспортных средств в текущий момент на соответствующей полосе движения. В качестве новой активной фазы светофора метод выбирает ту, для которой величина «давления» оказывается максимальной. При этом переключение фаз допускается с интервалом не чаще некоторого заданного временного интервала  $g_{min}$ . С учётом сделанных обозначений псевдокод используемого в экспериментах метода MaxPressure имеет следующий вид [11, 14]:

*Входные данные:*

- $g_{min}$  – допустимый минимальный временной интервал переключения,
- $t_p$  – время активности текущей действующей фазы светофора,
- $P$  – множество фаз светофора.

*Выходные данные*

- *phase* – фаза светофора, действующая с текущего момента.

**procedure** MaxPressure( $g_{min}, t_p, P$ )

if ( $t_p < g_{min}$ ) then

$t_p = t_p + 1$

else

$t_p = 0$

return argmax( {Pressure(*phase*) for *phase* in  $P$  } )

**end procedure**

Более детальное рассмотрение представленных и других ставших уже «классическими» эмпирических методов адаптивного и неадаптивного управления сигналами светофоров можно найти в обзорах [15–17].

1.2. Методы адаптивного управления сигналами светофоров на основе машинного обучения

В последние два десятилетия для адаптивного управления сигналами/фазами светофоров были использованы различные методы искусственного интеллекта и машинного обучения: методы на основе нечёткой логики [18], интеллекта роя [19], генетические алгоритмы [20–22], RL-методы [23–32] и др. Среди них решения на основе RL-методов являются, по мнению многих авторов, наиболее перспективными.

Для удобства изложения мы использовали следующую структуру настоящего подпараграфа: вначале представлено описание принципов методов обучения с подкреплением (далее – RL-методов) в типовой нотации, далее дано краткое описание наиболее известных RL-методов; в заключении настоящего подпараграфа указаны особенности использования RL-методов для управления сигналами / фазами светофоров.

Итак, пусть  $S$  обозначает множество состояний некоего объекта (для задачи управления сигналами/фазами светофора под состоянием может пониматься расположение и характеристики движения транспортных средств в окрестности перекрёстка или какие-либо производные числовые характеристики от указанных данных);  $A$  – множество возможных действий. Упрощённо, задача состоит в построении такого отображения  $S \rightarrow A$ , которое для каждого конкретного состояния  $s \in S$  давало бы то действие  $a \in A$ , которое необходимо предпринять в указанном состоянии в некотором смысле оптимальным образом. Построение указанного отображения производится путём обучения, то есть анализа некоторой последовательности вида:  $s_{start} = s_0, a_0, s_1, a_1, \dots, s_{N-2}, a_{N-2}, s_{N-1} = s_{terminal}$ .

Учитывая возможную статистическую природу перехода состояний при конкретных действиях, для решения указанной проблемы обычно используют аппарат Марковских процессов принятия решений:

$$\langle S, A, P, R, \gamma \rangle. \quad (2)$$

Здесь  $P$  определяет вероятности перехода из состояния в состояние при конкретных действиях в некоторый момент времени  $t$ , т.е.:

$$P_a(s_{prev}, s_{next}) = P_a(s_{t+1} = s_{next} | s_t = s_{prev}, a_t = a);$$

величина  $r_t \equiv R_a(s_t, s_{t+1})$  характеризует «награду» за переход из состояния  $s_t$  в состояние  $s_{t+1}$  под действием  $a$ ; а коэффициент  $\gamma \in [0, 1)$  используется для формальной записи целевой функции RL-задачи в виде:

$$R = \sum_{t=0}^{\infty} r_t \gamma^t. \quad (3)$$

В рамках представленного формализма (2) решение заключается в нахождении т.н. политики/стратегии  $\pi(s, a) \in R[0, 1]$ , которая бы численно характеризовала в каждый момент времени и состояния  $s_t$ , и предпочтительность соответствующего действия  $a_t$ . Типовое решение в этом случае заключается в построении т.н. Q-функции, которая задаётся как ожидаемое значение целевой функции (3) для конкретной пары состояние-действие:  $Q^\pi(s, a) = E\{R | s_t = s, a_t = a, \pi\}$ , и удовлетворяет уравнению Беллмана. Основными RL-методами (исторически) являются:

- метод Q-learning [33], разработанный в 1989 году и используемый в случае небольшого числа состояний и действий и заключающийся в итерационном формировании оценок  $Q(s, a)$  величин в соответствующей таблице;
- метод REINFORCE, предложенный в 1992 году и реализующий градиентный метод над параметрически заданной политикой  $\pi$  [34] (метод относится к классу методов TD – temporal difference);
- алгоритм Actor-Critic, предложенный в 1999 [35] и использующий независимо функцию политики ( $\pi$ -функцию) и функцию оценок (Q-функцию).

Функция политики именуется как *актёр*, потому что она используется для выбора действий, а Q-функция известна как *критик*, потому что она «критикует» (оценивает) действия, совершённые актёром. Такое разбиение позволило улучшить получаемые решения (метод также относится к классу методов TD);

– Double Q-learning [36], предложенный в 2011 году и являющийся модификацией базового алгоритма: он использует две Q-функции, одна из которых определяет действие, а вторая – оценку его «полезности»;

– метод Deep Q-learning или DQN, предложенный в 2014 году фирмой DeepMind (дочерняя фирма корпорации Google) и использующий в качестве отображения Q глубокую сверточную нейронную сеть;

– развитие метода Actor-Critic: алгоритмы A2C (англ. A2C – Advantage Actor Critic) и его асинхронная версия A3C [37], предложенные в 2016 году. Одним из последних и наиболее совершенных на момент написания данной работы считается метод Soft Actor-Critic, разработанный в 2018 году [38].

В данной работе, по аналогии с работой [30] по управлению сигналами светофора, для сравнения с предложенным методом мы используем один из наиболее современных A2C RL-метод [37], в котором DNN используются как для приближения политики, так и для оценки ценности, т.е. Q-функции.

*Применение RL-методов для решения проблемы адаптивного управления сигналами светофоров* требует, наряду с выбором конкретного метода и при использовании структуры DNN сети, информации о том, *как описывается состояние* дорожного движения в текущий момент времени и *как определяется «награда»*. Собственно, указанные выше работы по использованию RL-методов для адаптивного управления светофором [23–31] именно этим и отличаются. Среди возможных вариантов описания *состояния* можно выделить [3]: длину очереди, время ожидания, количество ТС, величину задержки, скорость, длительность фазы светофора и саму фазу, загруженность, относительные позиции ТС и др. Среди возможных вариантов задания величины «награды» можно выделить [3]: длину очереди, время ожидания, изменение задержки, скорость, количество остановок, пропускную способность, частоту смены сигнала, давление и др.

Для обучения RL-методов используют следующие системы моделирования: AIMSUN, CORSIM, MATSim, Paramics, VISSIM, GLD, SUMO, CityFlow, AIM [3, 39]. В настоящей работе использована система микромоделирования движения ТС с открытым кодом SUMO (Simulation of Urban MObility) как наиболее распространённая.

Более детальное рассмотрение представленных и других методов адаптивного управления сигналами светофоров с использованием RL-методов можно найти в обзорах: для ранних публикаций, когда подход с использованием RL не стал трендом, – в [15–17], для работ в последние два-три года – в обзорах [3, 30–32].

**2. Метод адаптивного управления сигналами/фазами светофоров, основанный на детерминированной прогнозной модели**

В рамках настоящей работы под детерминированной прогнозной моделью мы будем понимать набор явных аналитических закономерностей и /или операций, связывающих входные данные, то есть информацию о движении транспортных средств (включая транспортные средства типа CAV и CV) в окрестности конкретного перекрёстка, с данными о прогнозируемом «потоке» (то есть количестве ТС) транспортных средств через перекрёсток за одну конкретную фазу светофорного цикла: *PredictedFlow(phase)*. Собственно, сам предлагаемый метод управления тривиален – он выбирает ту фазу, прогнозируемый поток для которой оказывается максимален:

```

procedure MaxPredictedFlow( $g_{min}, t_p, P$ )
  if ( $t_p < g_{min}$ ) then
     $t_p = t_p + 1$ 
  else
     $t_p = 0$ 
  return  $\text{argmax}(\{ \text{PredictedFlow}(phase) \text{ for } phase \text{ in } P \})$ 
end procedure
    
```

Очевидно, ключевым моментом предложенного метода является то, как именно рассчитывается поток ТС за интервал  $g_{min}$  через перекрёсток в текущий момент времени, то есть вид функции *PredictedFlow(phase)*. Набор используемых в предложенной модели аналитических закономерностей базируется на простейших принципах физики (механики), связывающих возможное перемещение транспортного средства с его возможным ускорением и физическими ограничениями среды, вызванными, например, впереди стоящими ТС. Эти закономерности используют ряд параметров (максимальная скорость, максимальное ускорение, время задержки при начале движения и т.п.), которые на практике могут быть получены на основании оценок параметров движения реальных транспортных средств. В среде моделирования они оказываются известными.

2.1. Оценка потока для заданной фазы

Имея оценку времени  $t(c)$ , требуемого для достижения перекрёстка ТС-м  $c$  из множества ТС  $C_l$  на конкретной  $l$ -й полосе, искомая оценка потока имеет вид (другие обозначения см. в выражении (1)):

$$PredictedFlow(phase) = \sum_{l \in I_{phase}^{income}} \sum_{a \in C_l} I(t(c) < g_{min}),$$

здесь  $I(val)$  – индикатор, принимающий значение «1» при  $val = True$ , и значение «0» – иначе.

Способ оценки  $t(c)$  представлен ниже.

2.2. Детерминированная прогнозная модель: оценка интервала времени  $t(a)$ , необходимого для достижения перекрёстка

Для упрощения изложение ведётся для фиксированного ТС, поэтому аргумент  $c$ , обозначающий это ТС, в функции  $t(c)$  и других мы опускаем. Дополнительные обозначения:  $V_{max}$  – максимальная разрешённая скорость ТС;  $a_{max}$  – максимально возможное ускорение ТС;  $v$  – текущая скорость ТС;  $S$  – расстояние от ТС до следующего перекрёстка.

Время, за которое ТС достигнет перекрёстка  $t$ , складывается в общем случае из трёх составляющих:  $t_1$  – времени равноускоренного прямолинейного движения ТС (до момента достижения скорости  $V_{max}$ );  $t_2$  – времени равномерного прямолинейного движения;  $d$  – временной задержки в начале движения:  $t = t_1 + t_2 + d$ .

В зависимости от текущего значения скорости ТС  $v$  способ вычисления слагаемых различается.

Случай 1: Если  $v = 0$ , то предварительная оценка

$$t_1 = V_{max} / a_{max}, \tag{4}$$

то есть это интервал времени, за которое ТС достигнет скорости  $V_{max}$ , двигаясь равноускоренно. Поскольку ТС может достигнуть перекрёстка раньше, чем за время  $t_1$ , то это значение корректируется. А именно, пусть  $S_1$  – расстояние, которое преодолет ТС, двигаясь равноускоренно за время  $t_1$ . Имеем:

*Если*  $S_1 > S$ , тогда слагаемое  $t_2 = 0$ , а

$$t_1 = \sqrt{\frac{2S}{a_{max}}}, \quad d = \alpha \cdot S / (L + L_0). \tag{5}$$

Здесь  $L$  – средняя длина ТС,  $L_0$  – среднее расстояние между соседними ТС на полосе;  $\alpha$  – настраиваемый коэффициент, в экспериментах полагался  $\alpha = 1$ .

*Иначе*, то есть при  $S_1 \leq S$ , величина  $t_1$  вычисляется по (4),  $t_2 = (S - S_1) / V_{max}$ , а  $d$  вычисляется по (5).

Случай 2: Если  $v > 0$ , тогда время, за которое ТС достигнет скорости  $V_{max}$ , двигаясь равноускоренно:

$$t_1 = \frac{(V_{max} - v)}{a_{max}}; \text{ и } S_1 = v * t_1 + \frac{a_{max} * t_1^2}{2}. \tag{6}$$

*Если*  $S_1 > S$ , тогда слагаемое  $t_2 = 0$ , при

$$t_1 = \frac{-v + \sqrt{v^2 + 2a_{max} \cdot S}}{a_{max}}, \text{ и } d = \begin{cases} \frac{S}{L + L_0}, & \text{если } V_{max} - v > \Delta \\ 0, & \text{иначе} \end{cases}. \tag{7}$$

Здесь  $\Delta$  – настраиваемый параметр, полагавшийся в экспериментах  $\Delta = 5$ .

Иначе, то есть при  $S_1 \leq S$ , величины  $t_1$  и  $S$  вычисляются по (6),  $d$  вычисляется по (7), а

$$t_2 = \frac{S - S_1}{V_{\max}}$$

Выбор значений коэффициентов  $\alpha$  и  $\Delta$  обусловлен параметрами среды симуляции SUMO, используемыми в ходе проведения экспериментальных исследований.

### 3. Экспериментальные исследования

Цель экспериментов – сравнение эффективности предложенного метода MaxPredictedFlow и двух наиболее эффективных методов адаптивного управления сигналами светофора: эмпирического MaxPressure и RL-метода на базе современного алгоритма A2C. Причины выбора были указаны выше. В качестве данных для описания состояния в RL-методе на базе A2C использовался следующий набор:

- текущий сигнал светофора, представленный в виде бинарного вектора;
- нормированный вектор, состоящий из количества ТС на каждой из регулируемых светофором полос;
- нормированный вектор, состоящий из количества стоящих на месте ТС, для которых  $v(c) = 0$ , на каждой из регулируемых светофором полос.

Таким образом, для восьмистатусного светофора, изображённого на рис. 1, состояние перекрёстка описывается в виде вектора размерности 32 ( $= 8 + 12 + 12$ ).

Функция награды для RL-метода на базе A2C в экспериментах имела следующий вид:  $R = w_{t-1} - w_t$ , где  $w_t$  – сумма длительностей ожиданий ТС на регулируемых светофором полосах в момент времени  $t$ .

Для возможности сопоставления результатов с потенциально наихудшим значением параллельно оценивалась эффективность метода Uniform,  $g_{\min}$  во всех экспериментах и для всех методов полагался равным 10 с. Экспериментальные исследования были проведены в среде симуляции SUMO. Движение транспортных средств в SUMO может быть описано двумя способами. При использовании *первого способа* задаётся маршрут и время начала поездки для каждого ТС отдельно. При *втором способе* движение транспортных средств задаётся с использованием *потоков*: указывается маршрут поездки/потока и вероятность появления ТС в потоке (период появления ТС в потоке не превышает 1 секунды). В ходе проведения всех экспериментальных исследований использовался *второй способ описания движения*.

**Эксперимент 1.** Экспериментальные исследования проводились при шаге симуляции, равном 1 секунде, и при общем времени симуляции 300 секунд. Для проведения экспериментальных исследований было разработано три сценария (см. табл. 1).

Сценарий «1 × 1». Транспортная сеть представляет собой один регулируемый светофором перекрёсток,

как показано на рис. 1. Каждая проезжая часть длиной 500 метров состоит из трёх полос движения. Для каждой проезжей части, ведущей к перекрёстку, заданы три потока транспортных средств, при этом вероятности использования центральных полос (проезд «напрямую») выше, чем для остальных. Выбор проезжих частей равновероятен.

Сценарий «2 × 2». Транспортная сеть имеет вид решётки и включает в себя 4 перекрёстка, регулируемых светофорами. Каждая проезжая часть длиной 150 метров состоит из трёх полос, как на рис. 1. Движение транспортных средств задано в виде четырёх потоков: два горизонтальных с разными пунктами отправления и единым пунктом назначения, два вертикальных непересекающихся потока. Все светофоры используют один и тот же метод.

Табл. 1. Критерий эффективности управления дорожным движением с использованием сигналов светофоров

Тип сети	Критерий	Uniform	Max Pressure	A2C	Предл. метод
1 × 1	Mean	16,450	<b>9,706</b>	14,604	10,180
	CKO	1,570	0,869	1,911	<b>0,803</b>
	Min	14,288	<b>8,350</b>	11,972	8,614
	Max	18,756	<b>11,358</b>	18,850	11,569
2 × 2	Mean	54,229	17,945	10,815	<b>6,413</b>
	CKO	5,390	3,055	1,880	<b>0,912</b>
	Min	42,891	13,392	8,430	<b>5,163</b>
	Max	63,819	22,681	14,486	<b>7,675</b>
5 × 5	Mean	117,752	93,694	66,549	<b>63,422</b>
	CKO	<b>0,726</b>	2,636	1,109	1,235
	min	116,121	87,577	65,272	<b>61,121</b>
	max	118,961	97,936	69,088	<b>64,812</b>

Сценарий «5 × 5». Транспортная сеть имеет вид решётки и включает в себя 25 перекрёстков, регулируемых светофорами с помощью одного и того же метода. Каждая проезжая часть длиной 200 метров состоит из трёх полос, как на рис. 1. Для каждой проезжей части, смежной с центральными и крайними перекрёстками на границе транспортной сети, формируются два потока: первый с конечным пунктом, расположенным на противоположной стороне решётки; конечный пункт второго – соседний для конечного пункта первого, лежащий на противоположной стороне (центральный или крайний).

Для каждого из методов эксперименты с моделированием/обучением проводились по 10 раз, для каждого эксперимента оценивался критерий эффективности управления светофором: среднее время движения всех транспортных средств в сети. Статистические показатели этого критерия (среднее, минимум, максимум и CKO) для всех сравниваемых методов и типов сетей представлены в табл. 1.

Как видно из представленных в таблице результатов, по среднему, минимальному и максимальному среднему времени корреспонденций в сети *предложенный метод MaxPredictedFlow превосходит предыдущий RL-метод на базе A2C во всех случаях*, а метод

MaxPressure – в двух из трёх случаев, проигрывая только в тривиальном случае с одним перекрёстком!

**Эксперимент 2.** Отдельно исследовалась зависимость суммарного и среднего времени ожидания ТС на светофорах в зависимости от числа ТС в сети. Результаты представлены на рис. 2, единица по оси абсцисс соответствует 1200 ТС. Тип сети – «5 × 5».

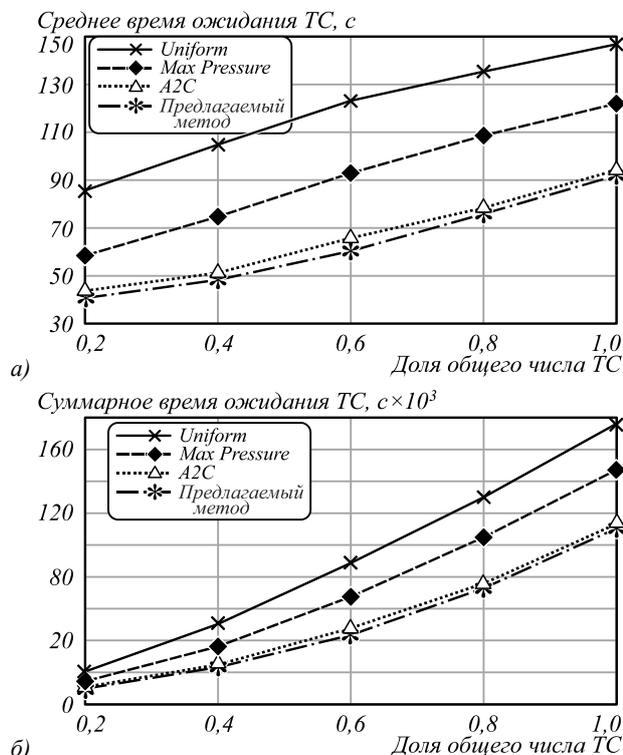


Рис. 2. Время ожидания ТС на светофорах в зависимости от числа ТС в сети: а) среднее время ожидания, б) суммарное время ожидания

Как видно из представленных результатов, предложенный метод везде превосходит все методы-конкуренты, включая RL-метод на базе A2C!

### Выводы и результаты

В работе предложен метод адаптивного управления сигналами/фазами светофоров в интеллектуальных транспортных системах, основанный на детерминированной прогнозном модели. Показано его однозначное преимущество перед наиболее эффективными существующими решениями: методом MaxPressure и RL-методов на базе A2C.

Представляются перспективными следующие направления развития настоящей работы:

- 1) совершенствование детерминированной модели расчёта  $t(c)$ ;
- 2) использование более сложной «непараметрической» модели прогнозирования  $t(c)$ , в том числе нейросетевой или на базе DNN;
- 3) расширение списка методов-конкурентов для более объективного сравнения;
- 4) сравнение эффективности методов на транспортной сети существующего мегаполиса.

### Благодарности

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

### Литература

1. The economist. The cost of traffic jams [Electronical Resource]. – URL: <https://www.economist.com/blogs/economist-explains/2014/11/economist-explains-1> (request date 20.08.2021).
2. Schrank, D. TTI's 2012 urban mobility report [Electronical Resource] / D. Schrank, B. Eisele, T. Lomax. – 2012. – URL: <https://static.tti.tamu.edu/tti.tamu.edu/documents/umr/archive/mobility-report-2012.pdf> (request date 20.08.2021).
3. Wei, H. A survey on traffic signal control methods [Electronical Resource] / H. Wei, G. Zheng, V. Gayah, Z. Li // arXiv Preprint. – 2020. – URL: <http://arxiv.org/abs/1904.08117> (request date 20.08.2021).
4. Агафонов, А.А. Исследование численного метода резервирования маршрутов в геоинформационной задаче маршрутизации автономных транспортных средств / А.А. Агафонов, В.В. Мясников // Компьютерная оптика. – 2018. – Т. 42, № 5. – С. 912-920. – DOI: 10.18287/2412-6179-2018-42-5-912-920.
5. Агафонов, А.А. Анализ больших данных в геоинформационной задаче краткосрочного прогнозирования параметров транспортного потока на базе метода  $k$  ближайших соседей / А.А. Агафонов, А.С. Юмаганов, В.В. Мясников // Компьютерная оптика. – 2018. – Т. 42, № 6. – С. 1101-1111. – DOI: 10.18287/2412-6179-2018-42-6-1101-1111.
6. Hunt, P.B. The SCOOT on-line traffic signal optimisation technique / P.B. Hunt, D.I. Robertson, R.D. Bretherton, M.C. Royle // Traffic Engineering & Control. – 1982. – Vol. 23, Issue 4. – P. 190-192.
7. Hunt, P.B. SCOOT, a traffic responsive method of coordinating signals / P.B. Hunt, D.I. Robertson, R.D. Bretherton, M.C. Royle. – Laboratory report, Vol. 1014. – Transport and Road Research Laboratory, 1981. – 41 p.
8. Lowrie, P.R. SCATS—a traffic responsive method of controlling urban traffic / P.R. Lowrie. – Roads and Traffic Authority NSW, 1990. – 28 p.
9. Постановление Правительства РФ от 23.10.1993 № 1090 (ред. от 31.12.2020) "О Правилах дорожного движения" (с изм. и доп., вступ. в силу с 01.03.2021).
10. Webster, F. Traffic signal settings / F. Webster. – Technical Paper no.39. – Road Research Laboratory, 1958. – 44 p.
11. Varaiya, P. The max-pressure controller for arbitrary networks of signalized intersections / P. Varaiya. – In: Advances in dynamic network modeling in complex transportation systems. Vol. 2. Complex networks and dynamic systems / ed. by S.V. Ukkusuri, K. Ozbay. – New York: Springer, 2013. – P. 27-66. – DOI: 10.1007/978-1-4614-6243-9\_2.
12. Gershenson, C. Self-organizing traffic lights [Electronical Resource] / C. Gershenson // arXiv Preprint. – 2004. – URL: <https://arxiv.org/abs/nlin/0411066> (request date 20.08.2021).
13. Cools, S.B. Self-organizing traffic lights: A realistic simulation / S.B. Cools, C. Gershenson, B.D. Hooghe. – In: Advances in applied self-organizing systems / ed. by M. Prokopenko. – 2<sup>nd</sup> ed. – London: Springer-Verlag, 2013. – P. 43-55. – DOI: 10.1007/978-1-4471-5113-5\_3.
14. Genders, W. An open-source framework for adaptive traffic signal control [Electronical Resource] / W. Genders, S. Razavi // arXiv Preprint. – 2019. – URL: <https://arxiv.org/abs/1909.00395> (request date 20.08.2021).

15. **Li, L.** A survey of traffic control with vehicular communications / L. Li, D. Wen, D. Yao // IEEE Transactions on Intelligent Transportation Systems. – 2014. – Vol. 15, Issue 1. – P. 425-432. – DOI: 10.1109/TITS.2013.2277737.
16. **Papageorgiou, M.** Review of road traffic control strategies / M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, Y. Wang // Proceedings of the IEEE. – 2003. – Vol. 91, Issue 12. – P. 2043-2067. – DOI: 10.1109/JPROC.2003.819610.
17. **Roess, R.P.** Traffic engineering / R.P. Roess, E.S. Prassas, W.R. McShane. – Upper Saddle River: Prentice Hall, 2004. – 786 p.
18. **Gokulan, B.P.** Distributed geometric fuzzy multiagent urban traffic signal control / B.P. Gokulan, D. Srinivasan // IEEE Transactions on Intelligent Transportation Systems. – 2010. – Vol. 11, Issue 3. – P. 714-727. – DOI: 10.1109/TITS.2010.2050688.
19. **Teodorović, D.** Swarm intelligence systems for transportation engineering: Principles and applications / D. Teodorović // Transportation Research, Part C: Emerging Technologies. – 2008. – Vol. 16, Issue 6. – P. 651-667. – DOI: 10.1016/j.trc.2008.03.002.
20. **Mikami, S.** Genetic reinforcement learning for cooperative traffic signal control / S. Mikami, Y. Kakazu // Proceedings of the First IEEE Conference on Evolutionary Computation. – 1994. – P. 223-228. – DOI: 10.1109/ICEC.1994.350012.
21. **Lee, J.** Real-time optimization for adaptive traffic signal control using genetic algorithms / J. Lee, B. Abdulhai, A. Shalaby, E.-H. Chung // Journal of Intelligent Transportation Systems. – 2005. – Vol. 9, Issue 3. – P. 111-122. – DOI: 10.1080/15472450500183649.
22. **Ricalde, E.** Evolving adaptive traffic signal controllers for a real scenario using genetic programming with an epigenetic mechanism / E. Ricalde, W. Banzhaf // 16<sup>th</sup> IEEE International Conference on Machine Learning and Applications. – 2017. – P. 897-902. – DOI: 10.1109/ICMLA.2017.00-41.
23. **Wei, H.** IntelliLight: A reinforcement learning approach for intelligent traffic light control / H. Wei, G. Zheng, H. Yao, Z. Li // Proceedings of the 24<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2018. – P. 2496-2505. – DOI: 10.1145/3219819.3220096.
24. **Wei, H.** CoLight: Learning network-level cooperation for traffic signal control / H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, Z. Li // Proceedings of the 28<sup>th</sup> ACM International Conference on Information and Knowledge Management. – 2019. – P. 1913-1922. – DOI: 10.1145/3357384.3357902.
25. **Chen, C.** Toward a thousand lights: decentralized deep reinforcement learning for large-scale traffic signal control / C. Chen, H. Wei, N. Xu, G. Zheng, M. Yang, Y. Xiong, K. Xu, Z. Li // Proc AAAI Conf on Artificial Intelligence. – 2020. – Vol. 34, Issue 4. – P. 3414-3421. – DOI: 10.1609/aaai.v34i04.5744.
26. **Liu, Y.** Intelligent traffic light control using distributed multi-agent Q learning / Y. Liu, L. Liu, W.-P. Chen // 2017 IEEE 20<sup>th</sup> International Conference on Intelligent Transportation Systems (ITSC). – 2017. – P. 1-8. – DOI: 10.1109/ITSC.2017.8317730.
27. **Li, Z.** Networkwide traffic signal control optimization using a multi-agent deep reinforcement learning / Z. Li, H. Yu, G. Zhang, S. Dong, C.-Z. Xu // Transportation Research, Part C: Emerging Technologies. – 2021. – Vol. 125. – DOI: 10.1016/j.trc.2021.103059.
28. **Yau, K.A.** A survey on reinforcement learning models and algorithms for traffic signal control / K.A. Yau, J. Qadir, H.L. Khoo // ACM Computing Surveys. – 2017. – Vol. 50, Issue 3. – P. 1-38. – DOI: 10.1145/3068287.
29. **Mannion, P.** An experimental review of reinforcement learning algorithms for adaptive traffic signal control / P. Mannion, J. Duggan, E. Howley. – In: Autonomic road transport support systems / ed. by Th.L. McCluskey, A. Kotsialos, J.P. Müller, F. Klügl, O. Rana. – Switzerland: Springer International Publishing, 2016. – P. 47-66. – DOI: 10.1007/978-3-319-25808-9\_4.
30. **Chu, T.** Multi-agent deep reinforcement learning for large-scale traffic signal control / T. Chu, J. Wang, L. Codecà, Z. Li // IEEE Transactions on Intelligent Transportation Systems. – 2019. – Vol. 21, Issue 3. – P. 1086-1095. – DOI: 10.1109/TITS.2019.2901791.
31. **Gregurić, M.** Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data / M. Gregurić, M. Vujić, C. Alexopoulos, M. Miletić // Applied Sciences. – 2020. – Vol. 10, Issue 11. – DOI: 10.3390/app10114011.
32. **Qadri, S.S.S.M.** State-of-art review of traffic signal control methods: challenges and opportunities / S.S.S.M. Qadri, M.A. Gökçe, E. Öner // European Transport Research Review. – 2020. – Vol. 12. – 55. – DOI: 10.1186/s12544-020-00439-1.
33. **Watkins, C.J.C.H.** Learning from delayed rewards [Electronical Resource] / C.J.C.H. Watkins. – 1989. – URL: <http://www.cs.rhul.ac.uk/~chrisw/thesis.html> (request date 20.08.2021).
34. **Ronald, W.J.** Simple statistical gradient-following algorithms for connectionist reinforcement learning / W.J. Ronald // Machine Learning. – 1992. – Vol. 8. – P. 229-256. – DOI: 10.1007/BF00992696.
35. **Sutton, R.S.** Policy gradient methods for reinforcement learning with function approximation / R.S. Sutton, D. McAllester, Y. Mansour // Proceedings of the 12<sup>th</sup> International Conference on Neural Information Processing Systems. – 1999. – P. 1057-1063.
36. **Van Hasselt, H.** Double Q-learning / H. Van Hasselt // Advances in Neural Information Processing Systems 23 (NIPS 2010). – 2010. – P. 2613-2622.
37. **Mnih, V.** Asynchronous methods for deep reinforcement learning / V. Mnih, A.P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, K. Kavukcuoglu // Proceedings of the 33<sup>rd</sup> International Conference on Machine Learning. – 2016. – P. 1928-1937.
38. **Haarnoja, T.** Soft Actor-Critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor / T. Haarnoja, A. Zhou, P. Abbeel, S. Levine // Proceedings of the 35<sup>th</sup> International Conference on Machine Learning. – 2018. – P. 1861-1870.
39. **Martinez, F.J.** A survey and comparative study of simulators for vehicular ad hoc networks (VANETs) / F.J. Martinez, C.K. Toh, J.-C. Cano, C.T. Calafate, P. Manzoni // Wireless Communications and Mobile Computing. – 2011. – Vol. 11, Issue 7. – P. 813-828. – DOI: 10.1002/wcm.859.

#### Сведения об авторах

**Мясников Владислав Валерьевич**, в 2008 защитил диссертацию на соискание степени доктора физико-математических наук. Работает профессором кафедры геоинформатики и информационной безопасности Са-

марского национального исследовательского университета имени академика С.П. Королева и одновременно ведущим научным сотрудником в ИСОИ РАН – филиале ФНИЦ «Кристаллография и фотоника» РАН. Круг научных интересов: компьютерное зрение, распознавание образов, искусственный интеллект, машинное обучение и геоинформатика. Имеет более 200 публикаций, в том числе более 100 статей и три монографии (в соавторстве). E-mail: [ymyas@geosamara.ru](mailto:ymyas@geosamara.ru).

**Агафонов Антон Александрович**, 1988 года рождения. В 2011 году окончил Самарский государственный аэрокосмический университет (СГАУ), в 2014 защитил диссертацию на соискание степени кандидата технических наук. В настоящее время работает старшим научным сотрудником НИЛ-55 Самарского университета. Круг научных интересов включает геоинформационные технологии, транспортное моделирование, веб-технологии. Имеет 36 публикаций, из них 10 статей. E-mail: [ant.agafonov@gmail.com](mailto:ant.agafonov@gmail.com).

**Юмаганов Александр Сергеевич**, 1993 года рождения. В 2016 году окончил Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет), в 2020 защитил диссертацию на соискание степени кандидата технических наук. В настоящее время работает старшим лаборантом НИЛ-55 Самарского университета. Область научных интересов: программирование, распознавание образов. Имеет 19 публикаций, из них 4 статьи. E-mail: [yumagan@gmail.com](mailto:yumagan@gmail.com).

---

ГРНТИ: 47.05.17

Поступила в редакцию 25 августа 2021 г. Окончательный вариант – 7 сентября 2021 г.

---

---

# A deterministic predictive traffic signal control model in intelligent transportation and geoinformation systems

V.V. Myasnikov<sup>1,2</sup>, A.A. Agafonov<sup>1</sup>, A.S. Yumaganov<sup>1</sup>

<sup>1</sup>Samara National Research University, 443086, Samara, Russia, Moskovskoye Shosse 34;

<sup>2</sup>IPSI RAS – Branch of the FSRC “Crystallography and Photonics” RAS,  
443001, Samara, Russia, Molodogvardeyskaya 151

## Abstract

In this paper, we propose a traffic signal control method in intelligent transportation and geoinformation systems, based on a deterministic predictive model. The method provides adaptive control based on traffic data, including data from connected and autonomous vehicles. The proposed method is compared with the state-of-the-art traffic signal control solutions: empirical control algorithms and reinforcement learning-based control methods. An advantage of the proposed method is shown and directions of further research are outlined.

**Keywords:** data analysis, intelligent transportation system, traffic light control, deterministic model, reinforcement learning, connected and autonomous vehicles.

**Citation:** Myasnikov VV, Agafonov AA, Yumaganov AS. A deterministic predictive traffic signal control model in intelligent transportation and geoinformation systems. *Computer Optics* 2021; 45(6): 917-925. DOI: 10.18287/2412-6179-CO-1031.

**Acknowledgements:** The work was supported by the Russian Science Foundation under grant No.21-11-00321, <https://rscf.ru/en/project/21-11-00321/>.

## References

1. The economist. The cost of traffic jams. Source: <https://www.economist.com/blogs/economist-explains/2014/11/economist-explains-1>.
  2. Schrank D, Eisele B, Lomax T. TTI's 2012 urban mobility report. Source: <https://static.tti.tamu.edu/tti.tamu.edu/documents/umr/archive/mobility-report-2012.pdf>.
  3. Wei H, Zheng G, Gayah V, Li Z. A survey on traffic signal control methods. Source: <http://arxiv.org/abs/1904.08117>.
  4. Agafonov AA, Myasnikov VV. Numerical route reservation method in the geoinformatic task of autonomous vehicle routing. *Computer Optics* 2018; 42(5): 912-920. DOI: 10.18287/2412-6179-2018-42-5-912-920.
  5. Agafonov AA, Yumaganov AS, Myasnikov VV. Big data analysis in a geoinformatic problem of short-term traffic flow forecasting based on a  $k$  nearest neighbors method. *Computer Optics* 2018; 42(6): 1101-1111. DOI: 10.18287/2412-6179-2018-42-6-1101-1111.
  6. Hunt PB, Robertson DI, Bretherton RD, Royle MC. The SCOOT on-line traffic signal optimisation technique. *Traffic Eng Control* 1982; 23(4): 190-192.
  7. Hunt PB, Robertson DI, Bretherton RD, Royle MC. SCOOT, a traffic responsive method of coordinating signals. Laboratory report, Vol 1014. Transport and Road Research Laboratory; 1981.
  8. Lowrie PR. SCATS—a traffic responsive method of controlling urban traffic. Roads and Traffic Authority NSW; 1990.
  9. RF Government Resolution No 1090 of October 23, 1993 [In Russian]. Source: <https://base.garant.ru/1305770/>.
  10. Webster F. Traffic signal settings. Road Research Laboratory; 1958; 39.
  11. Varaiya P. The max-pressure controller for arbitrary networks of signalized intersections. In Book: Ukkusuri SV, Ozbay K, eds. *Advances in dynamic network modeling in complex transportation systems*. Vol 2. Complex networks and dynamic systems 2013: 27-66. DOI: 10.1007/978-1-4614-6243-9\_2.
  12. Gershenson C. Self-organizing traffic lights. Source: <https://arxiv.org/abs/nlin/0411066>.
  13. Cools SB, Gershenson C, Hooghe BD. Self-organizing traffic lights: A realistic simulation. In Book: Prokopenko M, ed. *Advances in applied self-organizing systems*. 2<sup>nd</sup> ed. London: Springer-Verlag; 2013: 43-55. DOI: 10.1007/978-1-4471-5113-5\_3.
  14. Genders W, Razavi S. An open-source framework for adaptive traffic signal control. Source: <https://arxiv.org/abs/1909.00395>.
  15. Li L, Wen D, Yao D. A survey of traffic control with vehicular communications. *IEEE Trans Intell Transp Syst* 2014; 15(1): 425-432. DOI: 10.1109/TITS.2013.2277737.
  16. Papageorgiou M, Diakaki C, Dinopoulou V, Kotsialos A, Wang Y. Review of road traffic control strategies. *Proc IEEE* 2003; 91(12): 2043-2067. DOI: 10.1109/JPROC.2003.819610.
  17. Roess RP, Prassas ES, McShane WR. *Traffic engineering*. Upper Saddle River: Prentice Hall; 2004.
  18. Gokulan BP, Srinivasan D. Distributed geometric fuzzy multiagent urban traffic signal control. *IEEE Trans Intell Transp Syst* 2010; 11(3): 714-727. DOI: 10.1109/TITS.2010.2050688.
  19. Teodorović D. *Swarm intelligence systems for transportation engineering: Principles and applications*. *Transp Res Part C Emerg Technol* 2008; 16(6): 651-667. DOI: 10.1016/j.trc.2008.03.002.
  20. Mikami S, Kakazu Y. Genetic reinforcement learning for cooperative traffic signal control. *Proc First IEEE Conf on Evolutionary Computation* 1994: 223-228. DOI: 10.1109/ICEC.1994.350012.
  21. Lee J, Abdulhai B, Shalaby A, Chung EH. Real-time optimization for adaptive traffic signal control using genetic algorithms. *J Intell Transp Syst* 2005; 9(3): 111-122. DOI: 10.1080/15472450500183649.
  22. Ricalde E, Banzhaf W. Evolving adaptive traffic signal controllers for a real scenario using genetic programming with an epigenetic mechanism. 16<sup>th</sup> IEEE Int Conf on Machine
-

- Learning and Applications 2017: 897-902. DOI: 10.1109/ICMLA.2017.00-41.
23. Wei H, Zheng G, Yao H, Li Z. IntelliLight: A reinforcement learning approach for intelligent traffic light control. Proc 24<sup>th</sup> ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining 2018: 2496-2505. DOI: 10.1145/3219819.3220096.
  24. Wei H, Xu N, Zhang H, Zheng G, Zang X, Chen C, Zhang W, Zhu Y, Xu K, Li Z. CoLight: Learning network-level cooperation for traffic signal control. Proc 28<sup>th</sup> ACM Int Conf on Information and Knowledge Management 2019: 1913-1922. DOI: 10.1145/3357384.3357902.
  25. Chen C, Wei H, Xu N, Zheng G, Yang M, Xiong Y, Xu K, Li Z. Toward a thousand lights: decentralized deep reinforcement learning for large-scale traffic signal control. Proc AAAI Conf on Artificial Intelligence 2020: 34(4): 3414-3421. DOI: 10.1609/aaai.v34i04.5744.
  26. Liu Y, Liu L, Chen WP. Intelligent traffic light control using distributed multi-agent Q learning. IEEE 20<sup>th</sup> Int Conf on Intelligent Transportation Systems (ITSC) 2017: 1-8. DOI: 10.1109/ITSC.2017.8317730.
  27. Li Z, Yu H, Zhang G, Dong S, Xu C-Z. Networkwide traffic signal control optimization using a multi-agent deep reinforcement learning. Transp Res Part C Emerg Technol 2021; 125. DOI: 10.1016/j.trc.2021.103059.
  28. Yau KA, Qadir J, Khoo HL. A survey on reinforcement learning models and algorithms for traffic signal control. ACM Comput Surv 2017; 50(3): 1-38. DOI: 10.1145/3068287.
  29. Mannion P, Duggan J, Howley E. An experimental review of reinforcement learning algorithms for adaptive traffic signal control. In Book: McCluskey ThL, Kotsialos A, Müller JP, Klügl F, Rana O, eds. Autonomic road transport support systems. Switzerland: Springer International Publishing; 2016: 47-66. DOI: 10.1007/978-3-319-25808-9\_4.
  30. Chu T, Wang J, Li Z. Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE Trans Intell Transp Syst 2019; 21(3): 1086-1095. DOI: 10.1109/TITS.2019.2901791.
  31. Gregurić M, Vujić M, Alexopoulos C, Miletić M. Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data. Appl Sci 2020; 10(11). DOI: 10.3390/app10114011.
  32. Qadri SSSM, Gökçe MA, Öner E. State-of-art review of traffic signal control methods: challenges and opportunities. Eur Transp Res Rev 2020; 12: 55. DOI: 10.1186/s12544-020-00439-1.
  33. Watkins CJCH. Learning from delayed rewards. Source: (<http://www.cs.rhul.ac.uk/~chrisw/thesis.html>).
  34. Ronald WJ. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach Learn 1992; 8: 229-256. DOI: 10.1007/BF00992696.
  35. Sutton RS, McAllester D, Mansour Y. Policy gradient methods for reinforcement learning with function approximation. Proc 12<sup>th</sup> Int Conf on Neural Information Processing Systems 1999; 1057-1063.
  36. Van Hasselt H. Double Q-learning. Advances in Neural Information Processing Systems 23 (NIPS 2010) 2010: 2613-2622.
  37. Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, Silver D, Kavukcuoglu K. Asynchronous methods for deep reinforcement learning. Proc 33<sup>rd</sup> Int Conf on Machine Learning 2016: 1928-1937.
  38. Haarnoja T, Zhou A, Abbeel P, Levine S. Soft Actor-Critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. Proc 35<sup>th</sup> Int Conf on Machine Learning 2018: 1861-1870.
  39. Martinez FJ, Toh CK, Cano JC, Calafate CT, Manzoni P. A survey and comparative study of simulators for vehicular ad hoc networks (VANETs). Wirel Commun Mob Comput 2011; 11(7): 813-828. DOI: 10.1002/wcm.859.

#### *Authors' information*

**Vladislav Valerievich Myasnikov**, received his DrSc degree in Physics & Maths (2008). Currently he works as a professor at the Geoinformatics and Information Security department in Samara National Research University and, at the same time, as a leading researcher at the IPSI RAS, a Branch of the Russian Academy of Sciences 'Crystallography and Photonics' RAS. The range of scientific interests: computer vision, pattern recognition and artificial intelligence, machine learning and geoinformatics. He has about 200 publications, including more than 100 articles and three monographs. <http://www.ssau.ru/staff/62061001-Myasnikov-Vladislav-Valerevich>. E-mail: [vmyas@geosamara.ru](mailto:vmyas@geosamara.ru).

**Anton Aleksandrovich Agafonov** graduated from Samara State Aerospace University (SSAU) at 2011, received his PhD in Technical Sciences at 2014. At present, he is a researcher at Samara University. The area of interests includes geoinformatics, transport modelling and web-technologies. He's list of publications contains 36 publications, including 10 scientific papers. E-mail: [ant.agafonov@gmail.ru](mailto:ant.agafonov@gmail.ru).

**Alexander Sergeevich Yumaganov** graduated from Samara National Research University (2016), received his PhD in Technical Sciences at 2020. At present, he is a laboratory assistant at Samara University. His interests include computer programming, pattern recognition. He has 19 publications, including 4 scientific papers. E-mail: [yumagan@gmail.com](mailto:yumagan@gmail.com).

*Received August 25, 2021. The final version – September 7, 2021.*