Адаптивное управление светофорным объектом на основе машинного обучения

П.В. Остапенко[1], К.А. Султантемирова[1], О.Н. Сапрыкин[1]

[1]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. В данной статье рассматриваются основные причины возникновения транспортных заторов на дорогах города. Особое внимание обращается на современные методы адаптивного управления светофором как способ снизить время ожидания на регулируемом перекрестке. В статье излагаются современные подходы, основанные на применении методов искусственного интеллекта, а также известные проблемы данных методов. Авторами предложен метод оптимизации работы светофора на основе модифицированного алгоритма машинного обучения Q-learning. Метод апробирован на имитационной модели перекрестка Аврора-Партизанская города Самара.

1. Введение

Проблема городских заторов является очень серьезной по многим причинам. Современные города покрыты сетью автодорог, которые призваны обеспечивать экономические и социальные потребности населения. А для этого необходимо, чтобы они обладали высокой пропускной способностью. Но с ростом городского населения и увеличением уровня автомобилизации данная проблема стала колоссальной. Во многих мегаполисах все чаще стали наблюдаться транспортные коллапсы. Транспортное средство, стоящее в пробке, наносит вред не только водителю, теряющему свое время, но также наносит урон окружающей среде, выбрасывая в атмосферу вредные выхлопные газы [1]. Проблемы, связанные с заторами на дорогах, можно свести к следующим:

- увеличение транспортных расходов, связанных с дорожно-транспортными задачами, что негативно влияет на национальную производительность и конкурентоспособность;
- увеличение выбросов CO_2 от транспортных средств, из-за увеличения времени простоя;
- общественное недовольство по поводу отсутствия эффективного управления движением в качестве пункта назначения поездки время увеличивается [2].

До недавнего времени, основным методом увеличения пропускной способности было строительство и расширение имеющихся дорог, что приводило к большим финансовым и временным затратам. Лишь в конце двадцатого века, с развитием информационных и компьютерных технологий стали создаваться первые интеллектуальные транспортные системы управления светофорами, призванные минимизировать ожидание водителя на перекрестке. Данные методы были основаны на оптимизации и управлении дорожным движением в узловых точках дорожной сети (на перекрестках). В данной статье предложен метод оптимизации работы светофора на основе машинного обучения.

2. Обзор литературы

2.1. Методы Адаптивного управления светофором

С появлением первых светофоров возникла потребность в регулировании светофорного цикла для того, чтобы всем участникам дорожного движения не приходилось подолгу ждать своей очереди на перекрестке. Поначалу светофор регулировался вручную полицейским с помощью пульта управления, встроенного в корпус светофора. В связи с развитием электроники, появилась возможность снабжения светофоров таймерами и реле. Теперь сигналы переключались автоматически. Для каждого светофора было свое расписание переключений, а также особое время работы циклов в часы пик.

Один из первых адаптивных методов управления светофором, разработанный в начале семидесятых годов Лабораторией Транспортных Исследований (TRL) стал SCOOT (Split, Cycle and Offset Optimization Technique). Метод основан на регулировании светофорного цикла путем расчета показателя насыщенности фаз (процент используемого зеленого сигнала), а также коэффициента ожидания [3]. В конце семидесятых годов Sims A.G. и Dobinson K.W., взяв на вооружение методику «насыщения», разработали алгоритм под названием SCATS (Sydney Co-ordinated Traffic Control System) [4]. Перекрестки в SCATS объединены в системы и подсистемы по географическому признаку (по ведущим магистралям, или по районам). Данные от контроллеров передаются региональному компьютеру, а он в свою очередь передает информацию в городской центр управления движением. Системы SCATS и SCOOT зарекомендовали себя во многих городах мира, и по сей день используются для оптимизации дорожного движения.

Описанные выше методы имеют свои ограничения. Они работают над улучшением сразу всей контролируемой сети и делают это не напрямую, а через изменение циклов светофоров. В 2008 году компания Rhythm Engineering разработала систему InSync [5], которая лишена данных ограничений. Разработчики сделали ставку на поиск оптимальных локальных решений на конкретных перекрестках. С помощью этого предполагается найти оптимальное решение для всей дорожной сети. Для этого на каждом перекрестке устанавливаются камеры и технологии распознавания образов. InSync вычисляет, сколько времени машины ожидают зеленого сигнала. Данная система так же предусматривает «Сетевой» режим. Для пропуска очереди автомобилей, фазы светофоров корректируются, чтобы создать «зеленую волну». Решение о переключение в данный режим принимается оператором.

2.2. Методы управления светофором на основе обучения с подкреплением

Представленные выше методы требуют заранее определенной модели среды. То есть они работают эффективно, когда дорожное движение на перекрестке стабильно с течением времени. Но в реальных условиях, трафик имеет стохастическую природу. В связи с этим, для управления дорожным движением в недавнее время стали использоваться методы машинного обучения, в частности – обучение с подкреплением. Методы, основанные на искусственном интеллекте, все чаще применяются для управления светофорами, благодаря своей возможности адаптироваться к изменениям трафика [1]. Методы управления дорожным сигналом светофора на основе обучения с подкреплением представлены в следующих статьях [6-9].

2.2.1. MDP

Проблема обучения с подкреплением обычно формулируется как Марковский процесс принятия решений (MDP), который считается стандартом при формализации проблем, связанных с обучением последовательному принятию решений [10]. MDP может быть представлен с использованием функции R вознаграждения, набора состояний S, набора действий A и переходной функции T [10], то есть группы <S, A, T, R>. Когда в любом состоянии $s \in S$ выбор действия $a \in A$ приведет к тому, что среда войдет в новое состояние $s' \in S$ с вероятностью $T(s, a, s') \in (0, 1)$ и даст вознаграждение $r = R(s, a, s')$. Политика π определяет поведение агента в его среде. Политики обеспечивают отображение состояний на действия, которыми руководствуется агент при выборе наиболее подходящего действия для

данного состояния. Цель любого MDP - найти лучшую политику (ту, которая дает наибольшую общую награду) [11]. Оптимальная политика для MDP обозначается π^*.

RL можно классифицировать на две категории: без моделей (например, Q-Learning, SARSA) и на основе моделей (например, Dyna, Prioritized Sweeping). Для успешной реализации подходов, основанных на моделях, необходимо знать функцию перехода T [10], которую может быть трудно или даже невозможно определить даже в относительно простых областях. Напротив, в подходе без модели это не является обязательным требованием. Исследование требуется для безмодельных подходов, которые выбирают базовый MDP для получения знаний о неизвестной модели. Использование подхода, основанного на моделях, в весьма стохастической проблемной области, такой как управление трафиком, также добавляет ненужные дополнительные сложности по сравнению с безмодельным подходом [12]. Наш анализ в этой главе будет сосредоточен в основном на модельных подходах к этой проблемной области по причинам, изложенным выше.

2.2.2. Q-Learning

Одним из наиболее популярных подходов RL, используемых сегодня, является Q-Learning [13]. Это не политический, не модельный алгоритм обучения. Было показано, что Q-обучение сходится к оптимальным значениям действия с вероятностью 1, если все действия многократно выбираются во всех состояниях, а значения действия представлены дискретно [13]. В Q-Learning значения Q обновляются в соответствии с приведенным ниже уравнением:

$$Q(s_j, a_j) \leftarrow Q(s_j, a_j) + a_j\,((R_j + \gamma_j max\, Q(S_{j+1}, a) - Q(s_j, a_j)) \quad (1)$$

где:

- Действие (a): A это набор из всех возможных действий, которые агент может сделать. Светофор может изменить длительность цикла, в зависимости от длины очереди.
- Коэффициент скидки (γ): коэффициент скидки умножается на будущие вознаграждения, которые будут обнаружены агентом для снижения влияния на выбор действий агента. Это делает будущие вознаграждения менее ценными по сравнению с немедленными вознаграждениями; таким образом, это создает своего рода краткосрочный гедонизм для агента. Если γ равен 0.8 и вознаграждение равно 10 очкам после трех шагов, настоящая ценность вознаграждения составит 0.8^3 x 10. Коэффициент скидки равный 1 означает, что будущее вознаграждение не отличается от немедленного.
- Состояние (s): состояние – это конкретная и непосредственная ситуация, в которой находится агент. Это может быть конкретное место или момент, непосредственная конфигурация, которая ставит агента по отношению к другим важным вещам, таким как инструментам, препятствиям, врагам или призам. Это может быть конкретная ситуация, возвращенная с помощью окружающей среды или другая ситуация в будущем.
- Вознаграждения (R): вознаграждение – это обратная связь, с помощью которой мы оцениваем успех или неудачу действий агента. Вознаграждение для светофора – уменьшение времени ожидания машин на перекрестке. Из любого заданного состояния, агент отправляет выходные данные с помощью своих действий в окружающую среду, и окружающая среда возвращает новое состояние агента (которое получается в результате действия в предыдущим состоянии), а также вознаграждения, если таковые имеются. Вознаграждения могут быть немедленными или отсроченными. Они эффективно оценивают действия агента.
- Q отражает пары состояние-действие для наград.

Например, Saad Touhbi , Mohamed Ait Babram и другие [1] в своей работе применили алгоритм Q-Learning для адаптивного управления сигналом светофора. Управление RL было разработано для изолированного многофазного перекрестка с использованием микроскопического симулятора трафика Paramics. Новизна работы заключается в методологии, которая использует новое обобщенное пространство состояний с различными известными определениями вознаграждений. Результаты этого исследования продемонстрировали преимущество использования RL по сравнению с контроллерами. Помимо этого исследования алгоритм Q-Learning был применен в следующих работах [14-16].

2.2.3. Сравнительный анализ

Методы управления светофором на основе обучения с подкреплением показывают лучшие результаты, по сравнению с контроллерами движения [7, 15]. Так, El-Tantawy, S., Abdulhai, B., Abdelgawad в своей статье [7] установили, что контроллер на основе RL экономит 48% средней задержки автомобиля по сравнению с оптимизированным предварительно настроенным контроллером и полностью активированным контроллером. Адаптивное управление светофором на основе RL обеспечивает следующую экономию: средняя задержка (27%), длина очереди (28%) и l коэффициенты выбросов CO2 (28%).

3. Методы

3.1. Моделирование транспортных потоков

Для того, чтобы смоделировать перекресток, мы использовали микросимулятор дорожного городского движения SUMO (Simulation of Urban MOility). В этом плане мы последовали за другими исследователями [17]. Пакет SUMO позволяет строить различные типы дорожных сетей, добавлять автомобили различных типов, строить их маршруты, а также устанавливать светофорные объекты и датчики. Пакет SUMO поставляется с TraCI, который предоставляет доступ к моделированию дорожного движения, позволяет извлекать значения моделируемых объектов и манипулировать их поведением онлайн. TraCI содержит в себе пакет tools / traci, который позволяет взаимодействовать с SUMO с использованием Python [18].

Процедура проведения моделирования (рис. 1) включает в себя три этапа – подготовку транспортной сети и сценария моделирования, собственно моделирование, составление отчетов и обработку результатов.

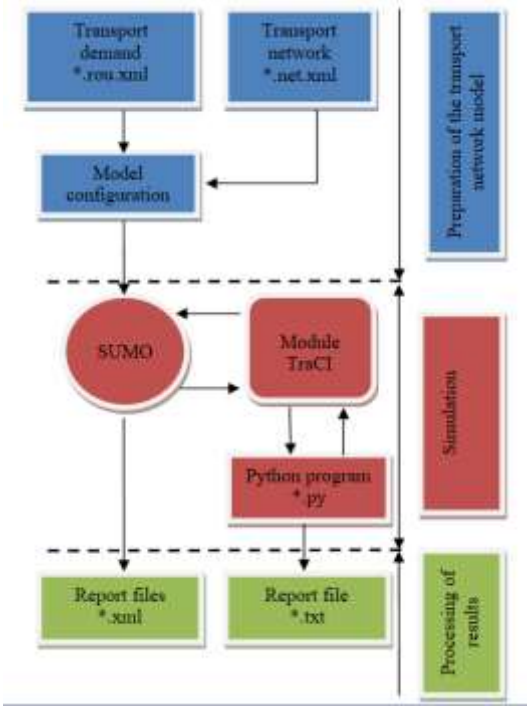

Рисунок 1. Схема функционирования системы моделирования.

3.2. Метод адаптации светофорного цикла

Введем функцию Q, отражающую ценность каждого возможного действия, агента a (в нашем случае – светофора) для текущего состояния моделирования s , в котором он находится

$$Q(s, a) \quad (5)$$

Процесс обучения – итерационное уточнение функции Q на каждом шаге. Величина максимальной возможной награды на следующем шаге определяется как:

$$max_a Q(s_{j+1}, a) \quad (6)$$

состояние зависит от длины светофорной фазы, размера очереди и времени ожидания на перекрестке. Имеет следующую зависимость:

$$(\text{light phases}) * [(\text{queue sizes}) * (\text{waiting times})]^{(\text{edges})} \quad (7)$$

Величина награды, которую получит агент, обозначим переменной r_t. Вознаграждение записывается по следующей формуле:

$$r = \sum_{\text{edge } i=1}^{4} \beta_q (\text{queue sizq})_i^{\theta q} + \beta_w (\text{waiting time})_i^{\theta w} \quad (8)$$

Введем дисконтирующий коэффициент γ, снижающий ценность будущих наград для светофора по сравнению с немедленными. Таким образом, формулу для функции Q запишем в следующем виде:

$$Q(s_j, a_j) \leftarrow Q(s_j, a_j) + a_j \left((R_j + \gamma_j \max Q(S_{j+1}, a) - Q(s_j, a_j) \right) \quad (9)$$

4. Результаты

Самара является крупным транспортным узлом в России, через который пролегают кратчайшие пути из Центральной и Западной Европы в Сибирь, Среднюю Азию и Казахстан. В городе имеются проблемы с заторами на дорогах. В часы пик часто наблюдаются транспортные коллапсы, когда движение во всем городе парализовано.

Одной из причин, приводящих к этому, является не оптимальная работа светофоров, которые работают по установленным циклам. Только в 2018 году, в преддверии чемпионата мира по футболу, главную магистраль города – Московское шоссе оснастили системой адаптивных светофоров. Тем не менее, большинство перекрестков нуждается в улучшении светофорного цикла и адаптации светофоров под существующий трафик. Большие пробки образуются на дорогах, ведущих в город по утрам, когда люди, живущие на окраинах, спешат на работу в центр города. Вечером наблюдается та же картина, но в обратном направлении.

Для своего исследования мы выбрали перекресток на пересечении улиц Партизанская и Аврора. На данном перекрестке наблюдается большое скопление машин, особенно в часы пик. Это связано с тем, что через это пересечение проезжает большое количество людей, проживающих на окраинах города, но работающих в центральной его части. Нам кажется, что светофор работает не оптимально и его работу можно улучшить.

Рисунок 2. Вид смоделированного перекрестка в программе SUMO.

Для того чтобы моделирование было максимально точным и приближённым к реальности, необходимо посчитать основные показатели исследуемого перекрёстка, такие как: длина цикла светофорного объекта, интенсивность движения транспортных средств на перекрёстке за 60 минут. Подсчет ведется в часы пик самого загруженного дня недели. Данные следует привести к формату программы SUMO и записать в файл модели. Полученные данные необходимо преобразовать в формат SUMO и занести в файлы модели. Составные части УДС занести в файл транспортной сети. Интенсивность транспортного потока на перекрёстке за четверть часа

привести к часовой, затем вычислить периодичность появления транспортного средства за час, и занести в файл транспортного спроса. Данные о работе светофорного объекта преобразовать в формат, содержащий информацию о количестве фаз, их последовательности в цикле, длительности основных тактов, длительности промежуточных тактов, общей длительности светофорного цикла, и занести в файл транспортного спроса.

Проанализировав результаты верификации и убедившись, что светофорный цикл настроен неправильно, приходим к идее внедрения адаптивного светофорного регулирования на данном участке УДС. На каждую полосу движения были поставлены детекторы E1 [17]. Детектор, который используется для измерения занятости полосы, потока ТС, времени прохождения ТС через детектор, скорости ТС, количества автомобилей, вошедших в детектор, вклада транспортных средств во взаимодействие с другими агентами в модели. Детектор вычисляет значения, путем определения времени транспортного средства, входящего и выходящего из детектора. Из достоинств данного типа детекторов можно отметить возможность измерения некоторых параметров потока ТС, которые недоступны для других типов детекторов.

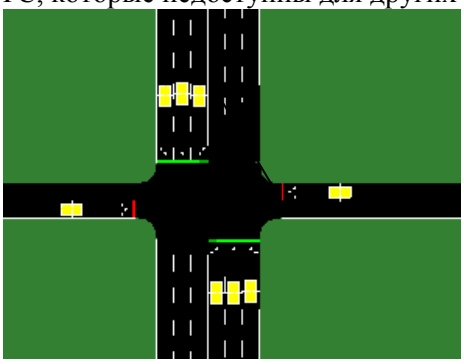

Рисунок 3. Детекторы транспортного потока в программе SUMO.

Далее был разработан модуль на языке программирования Python, позволяющий в реальном времени снимать показания датчиков и использовать их для оперативного регулирования светофорным объектом.

5. Заключение

В данной работе были рассмотрены современные методы адаптивного управления светофорным объектом. В программе SUMO была создана модель перекрестка Аврора-Партизанская, для которого предложен алгоритм работы светофора на основе нейронной сети. Основные проблемы метода заключаются в:

- точность моделирования зависит от точности данных, вносимых в модель, которые могут меняться с течением времени;
- выработанный светофорный цикл может не согласовываться с существующими светофорными ГОСТами.

В дальнейшем мы планируем сравнить данный метод оптимизации с другими методами. Также будет рассматриваться более широкая сеть, и учитываться больше факторов среды.

6. Литература

[1] Touhbia, S. Adaptive Traffic Signal Control: Exploring Reward Definition For Reinforcement Learning / S. Touhbia, M. Ait Babrama, T. Nguyen-Huu, N. Marilleau, M.L. Hbid, C. Cambierb, S. Stinckwich // The 8th International Conference on Ambient Systems, Networks and Technologies, 2017.

[2] Fernando, B. A Review of Current Traffic Congestion Management in the City of Sydney For Infrastructure Australia / B. Fernando, E. Gray, J. Kellner, 2013.

[3] Hunt, P.B. Winton: SCOOT - a traffic responsive method of coordinating signals / P.B. Hunt, D.I. Robertson, R.D. Bretherton // TRL Laboratory Report. – 1981. – Vol. 1014.

[4] Sims, A.G. The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits / A.G. Sims, K.W. Dobinson // IEEE Transactions on vehicular technology. – 1980. – Vol. 29(2). – P. 130-137.
[5] In Sync [Electronic resource]. – Access mode: https://trafficbot.rhythmtraffic.com.
[6] Bazzan, A.L.C. Opportunities for multiagent systems and multiagent reinforcement learning in traffic control // Autonomous Agents and Multi-Agent Systems. – 2009. – Vol. 18(3). – P. 342-375.
[7] El-Tantawy, S. Design of reinforcement learning parameters for seamless application of adaptive traffic signal control / S. El-Tantawy, B. Abdulhai, H. Abdelgawad // Journal of Intelligent Transportation Systems. – 2014. – Vol. 18(3). – P. 227-245.
[8] Mannion, P. An experimental review of reinforcement learning algorithms for adaptive traffic signal control / P. Mannion, J. Duggan, E. Howley // Autonomic Road Transport Support Systems – Springer, 2016. – P. 47-66.
[9] Yau, K.L.A. A survey on reinforcement learning models and algorithms for traffic signal control / K.L.A. Yau, J. Qadir, H.L. Khoo, M.H. Ling, P. Komisarczuk // ACM Computing Surveys (CSUR). – 2017. – Vol. 50(3). – P. 34.
[10] Wiering, M. Reinforcement Learning: State-of-the-Art / M. Wiering, M. van Otterlo – Springer, 2012.
[11] Salkham, A. A collaborative reinforcement learning approach to urban traffic control optimization / A. Salkham, R. Cunningham, A. Garg, V. Cahill // Web Intelligence and Intelligent Agent Technology. IEEE/WIC/ACM International Conference on. – 2008. – Vol. 2. – P. 560-566. DOI 10.1109/WIIAT.2008.88.
[12] El-Tantawy, S. Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (marlin-atsc): Methodology and large-scale application on downtown toronto / S. El-Tantawy, B. Abdulhai, H. Abdelgawad // Intelligent Transportation Systems, IEEE Transactions on. – 2013. – Vol. 14(3). – P. 1140-1150. DOI 10.1109/TITS.2013.2255286.
[13] Watkins, C. Technical note: Q-learning / C. Watkins, P. Dayan // Machine Learning. – 1992. – Vol. 8(3-4). – P. 279-292. DOI 10.1023/A:1022676722315.
[14] Abdulhai, B. Reinforcement learning for true adaptive traffic signal control / B. Abdulhai, R. Pringle, G. Karakoulas // Journal of Transportation Engineering. – 2003. – Vol. 129(3). – P. 278-285. DOI 10.1061/ (ASCE)0733-947X(2003)129:3(278).
[15] Arel, I. Reinforcement learning-based multi-agent system for network traffic signal control / I. Arel, C. Liu, T. Urbanik, A. Kohls // Intelligent Transport Systems, IET. – 2010. – Vol. 4(2). – 128-135. DOI 10.1049/iet-its.2009.0070.
[16] Balaji, P. Urban traffic signal control using reinforcement learning agents / P. Balaji, X. German, D. Srinivasan // Intelligent Transport Systems, IET. – 2010. – Vol. 4(3). – P. 177-188. DOI 10.1049/iet-its. 2009.0096.
[17] Covell, M. Micro-auction-based traffic-light control: Responsive, local decision making / M. Covell, Sh. Baluja, R. Sukthankar // IEEE 18th International Conference on Intelligent Transportation Systems (ITSC), 2015. – P. 558-565.
[18] Krajzewicz, D. Recent Development and Applications of SUMO – Simulation of Urban Mobility / D. Krajzewicz, J. Erdmann, M. Behrisch, L. Bieker // International Journal On Advances in Systems and Measurements. – 2012. – Vol. 5(3-4). – P. 128-138.

Adaptive traffic light control based on machine learning

P.V. Ostapenko[1], K.A. Sultantemirova[1], O.N. Saprykin[1]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. This article discusses the main causes of traffic congestion on the city roads. Particular attention is paid to modern methods of adaptive traffic control as a means to reduce the waiting time at a signaled crossing. The article outlines modern approaches based on the use of artificial intelligence methods, as well as the known issues of these methods. The authors propose a traffic light optimization method that is based on a modified Q-learning machine learning algorithm. The method was tested on a simulation model of the Aurora-Partizanskaya intersection in the city of Samara.