

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И АЛГОРИТМЫ УПРАВЛЕНИЯ ТРАНСПОРТНЫМИ СРЕДСТВАМИ НА ПЕРЕКРЕСТКАХ ДОРОГ

Шахсуварова Илона Витальевна

Аспирант, Санкт-Петербургский государственный
университет телекоммуникаций им. проф.

М.А. Бонч-Бруевича

shahsuvarova.ilona@mail.ru

MATHEMATICAL MODELS AND ALGORITHMS FOR CONTROLLING VEHICLES AT INTERSECTIONS

I. Shakhshvarova

Summary. At the present stage of social development, accompanied by global digitalization, driver assistance systems have become particularly relevant when driving, which allow monitoring the traffic situation with the help of artificial intelligence. Such systems have certainly made modern human-driven vehicles more efficient and safer. At the same time, the technology of unmanned vehicles (cars), despite the availability of appropriate technologies, has not yet become an everyday reality. This practical complexity is due to the fact that a new generation of unmanned systems needs algorithms capable of preventing traffic accidents and coping with complex traffic situations, just as a person does while driving a vehicle, taking into account an optimally complete list of intersection paths by various road users.

Based on this, it is necessary to develop such a model of neural network training that will allow networks to learn in real time, tactically building scenarios for crossing intersections in real time, that is, building a model for solving complex multi-purpose tasks. At the same time, to solve the modeling problem, it is not enough to simply develop a network model and scenario options for its training — in order to properly train an artificial neural network, it is necessary to choose an appropriate learning algorithm that will solve the task in an optimal way. Of the available learning algorithms, the most optimal seems to be the algorithm of a deeply deterministic policy gradient (DDPG). The specifics of this algorithm will ensure the stability and efficiency of the model. The Deep deterministic Policy Gradient (DDPG) algorithm combines the advantages of the Subject-Critic and DQN ((Deep Q-Learning) algorithms, choosing the best of them and thereby solving the complex problem of the continuous action space through experience reproduction and asynchronous updating, which becomes particularly relevant in those systems that are associated with unmanned driving vehicles.

Keywords: genetic algorithms, optimization problem, artificial neural networks, neural network learning algorithms, deeply deterministic policy gradient algorithm.

Аннотация. На современном этапе общественного развития, сопровождающимся глобальной цифровизацией, особую актуальность приобрели системы помощи водителю при вождении, позволяющие осуществлять контроль дорожно-транспортной ситуации посредством использования искусственного интеллекта. Такие системы, безусловно, сделали современные транспортные средства, управляемые человеком, более эффективными и безопасными. Вместе с тем, технология беспилотных транспортных средств (автомобилей), несмотря на наличие соответствующих технологий, всё ещё не стала повседневной реальностью. Такая практическая сложность связана с тем, что новое поколение беспилотных систем нуждается в алгоритмах, способных предотвращать дорожно-транспортные происшествия и справляться со сложными дорожными ситуациями подобно тому, как это делает человек, находясь за рулем транспортного средства с учетом оптимально полного перечня траекторий проезда перекрестков различными участниками дорожного движения.

Исходя из этого, необходима разработка такой модели обучения нейронных сетей, которая позволит сетям обучаться в режиме реального времени, тактически выстраивая сценарии пересечения перекрестков в режиме реального времени, то есть необходимо построение модели решения сложных многоцелевых задач. Вместе с тем, для решения задачи моделирования недостаточно просто разработать модель сети и варианты сценариев для ее обучения — для того, чтобы надлежащим образом обучить искусственную нейронную сеть, необходимо выбрать соответствующий алгоритм обучения, который позволит решить поставленную задачу оптимальным образом. Из имеющихся алгоритмов обучения наиболее оптимальным видится алгоритм глубоко детерминированного градиента политики (DDPG). Специфика работы данного алгоритма позволит обеспечить стабильность и эффективность работы модели. Алгоритм глубоко детерминированного градиента политики (DDPG) сочетает в себе преимущества алгоритмов «Субъект-критик» и DQN ((Deep Q-Learning), выбирая лучшее из них, решая тем самым сложную проблему пространства непрерывного действия через воспроизведение опыта и асинхронное обновление, что приобретает особую актуальность в тех системах, которые связаны с беспилотным управлением транспортными средствами.

Ключевые слова: генетические алгоритмы, задача оптимизации, искусственные нейронные сети, алгоритмы обучения нейронных сетей, алгоритм глубоко детерминированного градиента политики.

Вопросами развития технологий беспилотного управления транспортными средствами научное сообщество активно занимается последние два десятилетия, обосновывая тот факт, что они должны лежать в основе современного автомобилестроения. Вместе с тем, до настоящего момента не придумано идеальной технологии беспилотного управления, а все имеющиеся разработки в данной области имеют как свои преимущества, так и свои недостатки. Рассматривая многочисленные научные исследования, посвящённые внедрению технологий беспилотного управления, можно прийти к выводу о том, что в основе всех без исключения исследований лежит общая гипотеза о невозможности применения к технологиям беспилотного управления традиционных алгоритмов управления и уже известных математических моделей ввиду сложности самой системы беспилотного управления [4]. С указанным мнением стоит согласиться, поскольку суть системы беспилотного управления сводится в первую очередь к возможности принимать решения при управлении транспортным средством без воздействия человека. Иными словами, технология беспилотного управления транспортным средством — это технология, в основе которой лежит искусственный интеллект, базирующийся на различных алгоритмах обучения искусственных нейронных сетей (далее — ИНС).

Так, например, известна модель проезда перекрестков беспилотным транспортным средством, позволяющая избегать возможных столкновений на тех перекрестках, где у беспилотного транспортного средства отсутствует открытый обзор [10]. В основе такой модели, разработанной исследователями Массачусетского технологического института совместно с компанией Toyota, заложен алгоритм, оценивающий совокупность критических факторов, возникающих на перекрестках: визуальные препятствия, шум и скорость других автомобилей. После оценки факторов система беспилотного управления автоматизировано принимает соответствующее решение: остановиться, замедлить скорость, продолжить движение с установленной скоростью. Вместе с тем, вышеуказанная модель не ориентирована на ряд других сложных факторов риска, с которыми может столкнуться беспилотный автомобиль при проезде перекрестков — например, присутствие пешеходов на перекрестке или вблизи него.

Американскими исследователями разработаны также алгоритмы безопасного и быстрого проезда перекрестка дорог на основании имеющейся информации о специфике перекрестка (наличие или отсутствие светофоров, наличие информации о средней интенсивности движения и пр.) в транспортных средствах, где система управления автомобилем частично автоматизирована [5].

Российские исследователи И.А. Зикратов, Т.В. Зикратова и И.И. Виксин разработали алгоритм, позво-

ляющий беспилотному транспортному средству решать одну из типовых задач, стоящих перед беспилотным транспортным средством — находить наиболее оптимальный план безопасного проезда перекрестка [1]. В основе предложенной модели лежит алгоритм решения задачи с использованием межагентного общения и принятия решения, позволяющий более эффективно преодолевать перекрестки, в том числе за счет сокращения времени на принятие решения. Вместе с тем, следует отметить и недостатки такой модели, заключающиеся в том, что она будет эффективно работать только на полностью автоматизированных транспортных средствах только в том случае, если практически все они будут беспилотными или же в том случае, если беспилотный автомобиль будет эксплуатироваться в местах с невысокой интенсивностью движения. При этом следует отметить, что предложенная указанными исследователями модель может вполне эффективно расширить возможности автомобилей с гибридной системой управления (водитель и адаптивный круиз контроль).

Китайскими исследователями предложено множество алгоритмов обучения искусственных нейронных сетей для целей проезда беспилотных транспортных средств на перекрестках. Среди таких алгоритмов — алгоритм, основанный на обучении модели «Субъект-критик» с приближением гауссовского процесса для решения задач с непрерывными пространствами состояний и действий [8]. Также китайскими исследователями предложена скрытая Марковская модель для прогнозирования намерений других транспортных средств и построения модели принятия решений для транспортных средств на перекрестках [7].

Популярными являются также комбинированные автономные и онлайн-методы машинного обучения для создания персонализированной модели принятия решений, которая могла бы имитировать характеристики поведения водителя [2, 3], использование теории грубых множеств для извлечения правил принятия решений различными водителями.

Вместе с тем, в вышеобозначенных исследованиях не учитывались общие сценарии взаимодействия различных ситуаций на дороге, например, одновременное появление на перекрестке незапланированного препятствия (выскочившее на дорогу животное и проезд другого водителя на красный цвет), поэтому они могут быть приняты только для краткосрочного прогнозирования траектории. В решении сложных задач необходимо использование таких алгоритмов обучения нейронных сетей, которые могут одновременно позволять обучать в режиме реального времени на основе информации об уже прошедших, не взаимосвязанных друг с другом событиях.

Для решения многоцелевых задач стандартные алгоритмы обучения не подходят, поскольку не удовлетворяют потребности многоцелевых ограничений. При этом известно, что решение сложных многоцелевых задач возможно при использовании генетического алгоритма недоминантной сортировки (NSGA-II), разработанного в 2000 году Кальяномой Дебом, Амритом Пратапом и Самиром Агарвалом [6], и алгоритма глубоко детерминированного градиента политики (DDPG). Однако первый алгоритм (NSGA-II) в последних исследованиях в области решения сложных оптимизационных задач был подвергнут серьезной критике ввиду нестабильности и недостаточной эффективности при решении сложных задач управления, требующих воспроизведения опыта и асинхронного обновления ИНС [9].

Алгоритм глубоко детерминированного градиента политики (DDPG) в данной связи видится одним из наиболее перспективных алгоритмов для решения сложных задач, таких, как проезд перекрестков. Алгоритм глубоко детерминированного градиента политики (DDPG) одновременно изучает Q-функцию и политику, используя данные, полученные вне политики, и уравнение Беллмана для изучения Q-функции, при этом саму Q-функцию алгоритм использует для изучения политики, в связи с чем он прекрасно подходит для использования в средах с пространствами непрерывного действия (движение транспортного средства).

Алгоритм глубоко детерминированного градиента политики (DDPG) является улучшенным алгоритмом «Субъект-критик» («actor-critic»). В алгоритме «Субъект-критик» функция «субъект» генерирует действие с учетом текущего состояния. «Критик» оценивает функцию значения действия на основе выходных данных от «Субъекта», а также текущего состояния. Ошибки TD (temporal-difference), полученные от критика, приводят к его обучению в сети, а затем сеть «Субъектов» обновляется на основе градиента политики.

Алгоритм глубоко детерминированного градиента политики (DDPG) сочетает в себе преимущества алгоритмов «Субъект-критик» и DQN (Deep Q-Learning). Алгоритм DQN — это комбинация глубокого обучения, глубокого обучения и обучения с подкреплением обучения для достижения сквозного революционного алгоритма от восприятия к действию. Иными словами, алгоритм глубоко детерминированного градиента политики (DDPG) из концепции DQN выбирает целевую сеть и оценочную сеть как для «Субъекта», так и для «критика». Более того, политика алгоритма глубоко детерминированного градиента политики (DDPG) уже не стохастическая, а детерминированная. Это означает, что единственное реальное действие выводится из сети «Субъектов» вместо того, чтобы сообщать вероятность различных дей-

ствий. Критическая сеть обновляется на основе уравнения (1.1.):

$$L = \frac{1}{N} \sum_i^N (Q(s_t, a_t | \theta^Q) - y_i)^2 \quad (1.1.)$$

В уравнении (1.1.)

$$y_i = r_i + \gamma^Q (s_{t+1}, a_t | \theta^Q) \quad (1.2.)$$

В уравнении (1.1.) значение Q, оцененное целевой сетью, а N указывает общее количество мини-пакетов. Сеть «Субъектов» обновляется с помощью градиентного термина, представленного следующим уравнением:

$$\nabla_{\theta^\mu} J = \frac{1}{N} \sum_i^N \nabla_a Q(s, a | \theta^Q) |_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s | \theta^\mu) |_{s=s_i} \quad (1.3.)$$

В представленном уравнении (1.3.) Q(s, a | θ^Q) от сети оценка «критика».

Важно отметить, что алгоритм глубоко детерминированного градиента политики (DDPG) решает также проблему пространства непрерывного действия (что приобретает особую актуальность в тех системах, которые связаны с беспилотным управлением транспортными средствами) посредством воспроизведения опыта и асинхронного обновления.

При вышеописанных заданных условиях обновления сетей целевого «критика» и целевого «Субъекта» будет иметь следующий вид:

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'} \quad (1.4.)$$

Как уже было отмечено, алгоритм глубоко детерминированного градиента политики (DDPG) — это алгоритм вне политики. Иными словами, буфер воспроизведения может содержать старые события, даже если они могли быть получены с использованием устаревшей политики. Причина в том, что уравнению Беллмана все равно, какие особенности переходов используются, или как были выбраны действия, или что происходит после перехода, потому что оптимальная Q-функция должна удовлетворять уравнению Беллмана для всех возможных переходов. Таким образом, любые переходы, которые когда-либо испытывали при обучении ИНС, являются возможными при попытке соответствовать аппроксиматору Q-функции с помощью минимизации среднеквадратичной ошибки Беллмана.

Алгоритмы Q-обучения для аппроксиматоров функций, таких, как DQN (и все его варианты) и алгоритм глубоко детерминированного градиента политики (DDPG), в значительной степени основаны на минимизации этой функции потерь среднеквадратичной ошибки Беллмана. При этом применительно к алгоритму глубоко детерми-

нированного градиента политики (DDPG) есть один основной момент, касающийся вычисления максимума над действиями в цели.

Обычно вычисление максимума над действиями в цели является проблемой в пространствах непрерывных действий. Алгоритм глубоко детерминированного градиента политики (DDPG) решает такую задачу используя целевую сеть политики для вычисления действия, которое приблизительно максимизируется $Q_{\phi_{target}}$. Целевая сеть политик определяется тем же способом, что и целевая Q-функция: путем усреднения параметров политики в ходе обучения:

$$L(\phi, \mathcal{D}) = \mathop{\text{E}}_{(s, a, r, s', d) \sim \mathcal{D}} \left[(Q_{\phi}(s, a) - (r + \gamma(1 - d)Q_{\phi_{target}}(s', M_{\phi_{target}}(s'))))^2 \right] \quad (1.5.)$$

В представленном уравнении $M_{\phi_{target}}$ — целевая политика.

Алгоритм глубоко детерминированного градиента политики (DDPG) обучает детерминированную политику ИНС вне политики. Политика детерминирована, то есть имеет риск того, что не весь широкий спектр действий можно использовать, чтобы найти полезные обучающие сигналы. Чтобы лучше изучить политику при использовании алгоритма глубоко детерминированного градиента политики (DDPG), добавляют различные препятствия (шумы) во время обучения. В качестве примера можно привести некоррелированный гауссов шум среднего нуля. Чтобы облегчить получение более качественных обучающих данных, допускается уменьшение масштаба шума в ходе обучения.

Во время тестирования для того, чтобы увидеть, насколько хорошо ИНС использует то, что она узнала в процессе обучения, к действиям шум не добавляется.

Изучение политики в алгоритме глубоко детерминированного градиента политики (DDPG) довольно простое. Если необходимо изучить детерминированную политику $M_{\pi}(s)$, то дается действие, максимизирующее $Q_{\phi}(s, a)$. Поскольку пространство действий непрерывно и мы предполагаем, что Q-функция дифференцируема по отношению к действию, мы можем просто выполнить градиентное восхождение (только по отношению к параметрам политики) для решения следующего уравнения:

$$\max_{\pi} \mathop{\text{E}}_{(s \sim \mathcal{D})} [Q_{\phi}(s, M_{\pi}(s))] \quad (1.6.)$$

Вышеописанный алгоритм еще раз доказывает тот факт, что многозадачные модели принятия решений с алгоритмом глубоко детерминированного градиента политики (DDPG) более стабильны и эффективны, поскольку данный алгоритм сочетает в себе преимущества алгоритмов «Субъект-критик» и DQN ((Deep Q-Learning), выбирая лучшее из них, что позволяет сделать однозначный вывод о возможности адаптации алгоритма глубоко детерминированного градиента политики (DDPG) к использованию в моделях, нацеленных на решение сложных многоцелевых задач, таких, как поездки перекрестков беспилотными транспортными средствами.

Подводя итог, отметим, что в настоящий момент невозможно разработать оптимально полный перечень траекторий проезда перекрестков другими участниками дорожного движения. С учетом того обстоятельства, что современные возможности хоть и позволяют проводить обучение нейронных сетей, предусмотреть все возможные траектории участников дорожного движения достаточно проблематично. Исходя из этого, необходима разработка такой модели обучения нейронных сетей, которая позволит сетям обучаться в режиме реального времени, тактически выстраивая сценарии пересечения перекрестков в режиме реального времени, то есть выполнять построение модели решения сложных многоцелевых задач. Вместе с тем, для решения задачи моделирования недостаточно просто разработать модель сети и варианты сценариев для ее обучения — для того, чтобы надлежащим образом обучить ИНС, необходимо выбрать соответствующий алгоритм обучения, который позволит решить поставленную задачу оптимальным образом. Из имеющихся алгоритмов обучения наиболее оптимальным видится алгоритм глубоко детерминированного градиента политики (DDPG), поскольку специфика его работы позволяет обеспечивать стабильность работы модели эффективнее, чем многие иные модели для проезда перекрестков, так как данный алгоритм сочетает в себе преимущества алгоритмов «Субъект-критик» и DQN ((Deep Q-Learning), выбирая лучшее из них, в связи с чем решается проблема пространства непрерывного действия посредством воспроизведения опыта и асинхронного обновления, что приобретает особую актуальность в тех системах, которые связаны с беспилотным управлением транспортными средствами.

ЛИТЕРАТУРА

1. Зикратов И.А., Вискнин И.И., Зикратова Т.В. Оптимизация проезда опасных участков дорог беспилотными транспортными средствами // Научно-технический вестник информационных технологий, механики и оптики. — 2019. — № 1. С. 155–165.
2. Зикратов И.А., Вискнин И.И., Зикратова Т.В. Мультиагентное планирование проезда перекрестка дорог беспилотными транспортными средствами // Научно-технический вестник информационных технологий, механики и оптики. — 2016. — № 5. — С. 839–849.

3. Медведков Д.И. Полный факторный эксперимент в исследовании работы блока управления проезда перекрестка беспилотным транспортным средством // Научно-технический вестник информационных технологий, механики и оптики. — 2018. — № 4. — С. 669–670.
4. Усков, А.А. Интеллектуальные технологии управления. Искусственные нейронные сети и нечеткая логика. / А.А. Усков, А.В. Кузьмин. — М.: Горячая линия-Телеком, 2004. — 143 с.
5. Carlino D., Boyles S.D., Stone P. Auction-based autonomous intersection management // Proc. 16 th Int. IEEE Conf. on Intelligent Transportation Systems (ITSC2013). — Hague, Netherlands, 2013. — P. 529–534.
6. Deb K., Pratap A., Agarwal S., Meyarivan T. A fast elitist non-dominated sorting genetic algorithm for multiobjective optimization: NSGA-II // in Marc Schoenauer, et al., editors, Parallel Problem Solving from Nature (PPSN VI), pages 849–858, Springer, 2000.
7. Liu et al., Research on intelligent merging decision-making of unmanned vehicles based on reinforcement learning, in 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 91–96, Changshu, Suzhou, China, July 2018.
8. X. Chen, M. Jin, M. Yi-song, and Q. Zhang, Driving decision-making analysis of car-following for autonomous vehicle under complex urban environment, Journal of Central South University, vol. 24, pp. 1476–1482, 2017.
9. Yahui, Wang & Ling, Shi & Cai, Zhang & Liuqiang, Fu & Xiangjie, Jin. (2020). NSGA-II algorithm and application for multi-objective flexible workshop scheduling. Journal of Algorithms & Computational Technology.
10. В МТИ разработали алгоритм, помогающий беспилотным авто проходить перекрёстки. Электронный ресурс. Режим доступа: <https://habr.com/ru/news/t/474730/> (Дата обращения 01.06.2022 г.).

© Шахсуварова Илона Витальевна (shahsuvarova.ilona@mail.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»



Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича