

ОПТИМІЗАЦІЯ ДОРОЖНЬОГО РУХУ НА ПЕРЕХРЕСТЯХ ЗАСОБАМИ ГЛИБИННОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Мокрик Я.Л.¹; Бойко Н.І., к.е.н., доц.²

^{1, 2} Національний університет «Львівська політехніка»

^{1, 2} Україна, Львів

¹ 0000-0002-2338-855X; ² 0000-0002-6962-9363

Анотація. Роботу присвячено розробці середовища симуляції дорожнього руху та порівнянню алгоритмів навчання з підкріпленням для керування автомобілями та світлофорами на симульованому перехресті. Результатами дослідження є метрики розглянутих алгоритмів (DQN, PPO) і порівняння їх з деякими модифікованими алгоритмами в різних сценаріях дорожнього руху.

Ключові слова: машинне навчання; глибинне навчання з підкріпленням; Q-навчання; автономні транспортні засоби; світлофорне регулювання.

Вступна частина. Стрімкий ріст автомобільної промисловості та збільшення кількості транспортних засобів на дорогах створюють чимало зацікавленості в темі автономного керування транспортними засобами. Упродовж останніх кількох років кількість аварій на некерованих перехрестях, тобто тих, де нема світлофорів або знаків, була значно більшою порівняно з сигналізованими перехрестями [1]. Автори одного дослідження зауважили, що висока кількість аварій на дорогах є актуальною і глобальною проблемою, щороку майже 1,35 мільйонів людей помирають або стають інвалідами в результаті транспортних аварій [2].

Для того, щоб зрозуміти існуючі підходи до вирішення цих проблем, під час цієї роботи ми розглянули існуючі літературні джерела для порівняння середовищ симуляції дорожнього руху, а також порівняння існуючих алгоритмів керування автомобілями та світлофорами на перехрестях. Було взято до уваги 17 досліджень, з яких три були присвячені розробці середовищ для тренування моделей машинного навчання, а саме DriverGym [3], CriSGen [4] і дослідження мікроскопічної симуляції транспорту [5]. Також було взято до уваги середовища тренування моделей машинного навчання, які використовувались в інших дослідженнях, наприклад CARLA [6] і SUMO [7]. Два дослідження присвячені проблемі керування світлофорами для оптимізації руху транспорту через сигналізоване перехрестя, а решта 12 розглядають різні методи керування автономними транспортними засобами в симульованих середовищах.

Головними недоліками існуючих середовищ симуляції транспорту виявились можливість проводити велику кількість симуляцій з обмеженими ресурсами для тренування багатьох моделей, можливість керування параметрами симуляцій та створення конкретних критеріїв оцінки алгоритмів. Тому розроблене власне середовище для тренування моделей машинного навчання та їх порівняння. Основними недоліками існуючих алгоритмів є складність адаптації алгоритму до більш складних умов, тобто відсутність можливості узагальнення вміння керування автомобілями в сценаріях, які не були враховані в симуляції.

Мета роботи полягає в побудові методу оптимізації руху дорожнього транспорту на перехресті з допомогою навчання з підкріпленням. Досягнення цієї мети здійснене шляхом розв'язання певних теоретичних і прикладних проблем, а саме: дослідження існуючих алгоритмів машинного навчання для керування транспортними засобами, програмна

реалізація середовища симуляції дорожнього руху, розробка алгоритмів, які спроможні оптимізувати дорожній рух у середовищі та порівняння розроблених алгоритмів.

Основна частина. Перш за все, наведемо короткий опис розробленого середовища тренування моделей і методику порівняння алгоритмів. Середовище дає змогу тренувати алгоритми на різних сценаріях дорожнього руху, від спрощених до складніших, що дозволяє оцінити вміння моделей узагальнювати навчені можливості. Це допомагає зрозуміти, як вони працюватимуть в реальному світі. Оцінки роботи таких алгоритмів здійснені за допомогою різних метрик, наприклад AQL (Average Queue Length, середня довжина черги) і AWT (Average Waiting Time, середній час очікування). Середовище дозволяє тренувати два типи моделей машинного навчання: ті, що керують транспортом, і ті, що керують світлофорами.

Для керування транспортом модель отримує вхідні дані, що містять координати автомобіля на перехресті, відстань до інших автомобілів, їх кількість і фаз світлофорів на даний час. Функція винагороди для моделей, що керують автомобілями, враховує відстань автомобіля до цілі, тобто до сторони перехрестя, до якої авто прямує, а також будь-які порушення правил дорожнього руху: перетин перехрестя при червоному світлі, перетин меж доріг, перетин зустрічної смуги, довга зупинка при зеленому світлі, тощо. Простір дій керування авто є дискретним, воно складається з чотирьох можливих дій: прискорення, гальмування, поворот руля вліво і вправо. Окрім керування авто, розглядаються також алгоритми, які керують світлофорами на перехресті. У таких сценаріях, рух автомобілів зафіксовано, як і раніше, але жоден з них не керується агентом. Керування світлофорами реалізується одним із двох способів: керування з дискретним простором дій, де алгоритм обирає один із можливих фаз світлофора кожні 1000 кроків симуляції, і керування з неперервним простором дій, де модель обирає довжину наступної фази світлофора.

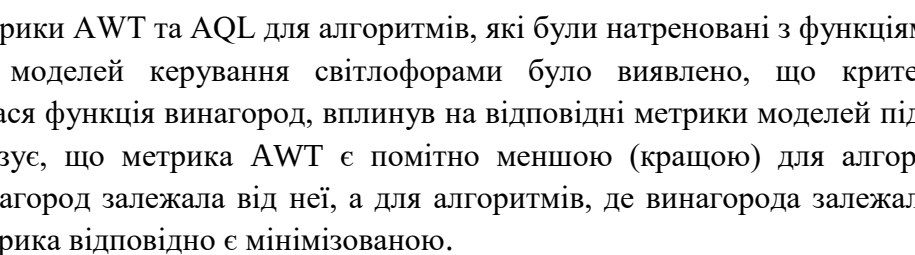
Табл. 1. Результати тренування DQN і PPO на найскладнішому сценарії руху транспорту

Алгоритм	LR	Ентропія	Найкраща винагорода	Кроки до найкращої винагороди
DQN	0.01	-	$2.67 * 10^4$	797,000
DQN	0.001	-	$3.56 * 10^4$	2,150,000
DQN	1.0 (спад.)	-	$3.54 * 10^4$	2,440,000
DQN	0.1 (спад.)	-	$5.22 * 10^4$	2,280,000
PPO	0.0003	0.1	$1.64 * 10^4$	30,000
PPO	0.0003	0.01	$2.64 * 10^4$	770,000
PPO	0.0003	0.001	$3.95 * 10^4$	2,700,000
PPO	0.0003	0.0001	$1.86 * 10^4$	1,380,000

Під час дослідження було використано різні методи безмодельного навчання з підкріпленням. Серед них алгоритми DQN (Deep Q Network, глибинна Q-мережа) і PPO (Proximal Policy Optimization) [8]. Експерименти керування авто проведено на трьох сценаріях руху автомобілів на перехресті: від простіших, де немає іншого транспорту, до ускладнених, де є транспорт, який рухається з усіх напрямків перехрестя. У табл. 1 наведено значення функції винагороди тренування алгоритмів DQN і PPO на найскладнішому сценарії. Ці експерименти проведені з різними значеннями гіперпараметрів LR (Learning Rate, темп навчання) і коефіцієнту ентропії. PPO показав найкращі результати. Більше того,

Bar chart comparing AWT (Average Wait Time) and AQL (Average Query Latency) execution times for various queries. The Y-axis represents time in seconds, ranging from 0 to 35. The X-axis lists 24 queries. AWT is shown in blue bars and AQL in orange bars. AWT times are generally higher than AQL times, with some queries showing significantly higher AWT times (e.g., q10, q11, q12, q13, q14, q15, q16, q17, q18, q19, q20, q21, q22, q23, q24).

Query	AWT (s)	AQL (s)
q1	12.5	10.5
q2	14.5	5.0
q3	13.5	7.0
q4	12.5	12.5
q5	13.5	8.5
q6	13.5	5.0
q7	15.5	11.0
q8	8.0	10.5
q9	5.5	3.0
q10	6.5	3.0
q11	1.5	1.0
q12	3.5	3.0
q13	35.0	1.5
q14	35.0	1.0
q15	12.5	0.5
q16	36.0	0.5
q17	4.5	3.0
q18	4.5	3.5
q19	1.5	3.0
q20	35.0	1.0
q21	26.0	0.5
q22	21.0	0.5
q23	37.0	0.2
q24	0.2	0.2



Висновки. Під час дослідження проведено аналіз сучасної літератури на тему керування автономними транспортними засобами і світлофорами на дорогах. Виявлено, що основними недоліками існуючих підходів є складність адаптації моделей до реальних ситуацій і мала кількість чітких метрик для порівняння моделей між собою. Розроблене середовище дозволяє порівнювати роботу агентів машинного навчання з підкріпленням на різних сценаріях дорожнього руху, від спрощених до більш складних, і надає метрики AQL і AWT для їх оцінки. Виявлено, що алгоритми PPO з неперервним простором дій дають кращі результати тренування, які менше страждають від проблеми катастрофічного забування, особливо з малим значенням гіперпараметру коефіцієнту ентропії (0.001-0.0001). Серед алгоритмів DQN ті, які мали високе початкове значення гіперпараметру темпу навчання (1.0-0.1), яке лінійно спадає до нуля впродовж тренування, отримали кращі результати порівняно з алгоритмами DQN зі статичним гіперпараметром темпу навчання на сценаріях водіння автомобіля. Показано позитивну кореляцію між вибором метрики для функції винагород і валідаційними метриками результатів керування світлофорами AWT і AQL.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Tran, Q.-D., & Bae, S.-H. (2021). An Efficiency Enhancing Methodology for Multiple Autonomous Vehicles in an Urban Network Adopting Deep Reinforcement Learning. *Applied Sciences*, 11(4), Article 4. <https://doi.org/10.3390/app11041514>
- [2] Ahmed, S. K., Mohammed, M. G., Abdulqadir, S. O., El-Kader, R. G. A., El-Shall, N. A., Chandran, D., Rehman, M. E. U., & Dhama, K. (2023). Road traffic accidental injuries and deaths: A neglected global health issue. *Health Science Reports*, 6(5), e1240. <https://doi.org/10.1002/hsr2.1240>
- [3] Kothari, P., Perone, C., Bergamini, L., Alahi, A., & Ondruska, P. (2021). *DriverGym: Democratising Reinforcement Learning for Autonomous Driving* (arXiv:2111.06889). arXiv. <https://doi.org/10.48550/arXiv.2111.06889>
- [4] Nonnengart, A., Klusch, M., & Müller, C. (2020, Листопад 24). *CriSGen: Constraint-based Generation of Critical Scenarios for Autonomous Vehicles*.
- [5] Bacchiani, G., Molinari, D., & Patander, M. (2019). *Microscopic Traffic Simulation by Cooperative Multi-agent Deep Reinforcement Learning* (arXiv:1903.01365). arXiv. <https://doi.org/10.48550/arXiv.1903.01365>
- [6] Gutiérrez-Moreno, R., Barea, R., López-Guillén, E., Araluce, J., & Bergasa, L. M. (2022). Reinforcement Learning-Based Autonomous Driving at Intersections in CARLA Simulator. *Sensors*, 22(21), Article 21. <https://doi.org/10.3390/s22218373>
- [7] 'Eclipse SUMO – Simulation of Urban Mobility' / <https://eclipse.dev/sumo/>. (б. д.). Eclipse SUMO - Simulation of Urban MObility. Вилучено 02, Листопад 2023, із <https://www.eclipse.dev/sumo/>
- [8] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal Policy Optimization Algorithms* (arXiv:1707.06347). arXiv. <https://doi.org/10.48550/arXiv.1707.06347>

Yaroslav Mokryk, Nataliya Boyko

Optimizing Traffic at Intersections with Deep Reinforcement Learning

Abstract. The work is dedicated to developing a traffic simulation environment and comparing deep reinforcement learning algorithms for the purpose of controlling autonomous vehicles and traffic lights at a simulated intersection. The results of the research include the metrics of the studied algorithms (DQN, PPO) and their comparison with certain modified algorithms in different traffic scenarios.

Keywords: Machine learning; Deep reinforcement learning; Q-learning; Autonomous vehicles; Traffic light control.