ICCEPM 2024 https://dx.doi.org/10.6106/ICCEPM.2024.0799
The 10th International Conference on Construction Engineering and Project Management
Jul. 29-Aug.1, 2024, Sapporo

Q-learning for tunnel excavation schedule

Shuhan YANG[1], Ke DAI[1], Zhihao REN[1*], Jung In KIM[1], Bin XUE[2], Dan WANG[2], Wooyong JUNG[3]

[1] *Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR, E-mail address: shuhayang5-c@my.cityu.edu.hk; kedai2-c@my.cityu.edu.hk; z.ren@my.cityu.edu.hk; jungikim@cityu.edu.hk*
[2] *School of Public Policy and Administration, Chongqing University, China, E-mail address: bxue@cqu.edu.cn; wangxiaodan@cqu.edu.cn*
[3] *Department of Nuclear Power Plant Engineering, KEPCO International Nuclear Graduate School, South Korea, E-mail address: trustjung@gmail.com*

Abstract: Construction planners for hard rock tunnel projects often encounter practical challenges caused by inherent uncertainties in ground conditions and resource constraints. Therefore, planners cannot rapidly generate optimal excavation schedules for the shortest project durations with a given equipment fleet by considering the uncertainties in ground conditions. Although some schedule optimization methods exist, they are not tailored for resource-constrained hard rock tunnel projects. To overcome these limitations, the authors specified a formal Q-learning-based schedule optimization methodology for resource-constrained hard rock tunnel projects. States are defined according to the locations of tunnel faces under excavation. Actions consist of multiple and comprehensive heuristic-based rules, which are efficient methods for resource allocation. Rewards are the time intervals required between current states and next states. After that, the methodology is validated using a case study. The generated Q tables indicate (1) best actions under different states and (2) the shortest remaining durations when the project starts from specific (state, action) pairs. The results demonstrate that the optimal schedules can be obtained by applying the proposed methodology. Furthermore, it is beneficial for planners to rapidly assign optimal rules for each state under one ground condition scenario. The results further show the potential to consider the uncertainties in ground conditions using the information of possible ground condition scenarios provided.

Key words: Q-learning, schedule optimization, hard rock tunnel, resource allocation policies

1. INTRODUCTION

In hard rock tunnel construction projects, planners face practical challenges due to the constraints imposed by limited resources and the inherent uncertainties in ground conditions, which may lead to cost overruns, schedule delays, and associated risks [1]. Thus, construction planners (CPs) are required to generate optimal schedules that not only accommodate the uncertainties in ground conditions and resource constraints, but also ensure efficient resource allocation.

Some studies have been carried out on the resource-constrained project scheduling problems [2][3][4][5][6], and heuristic-based policies are proven to be the efficient solutions for dealing with such problems [2][7]. Yet optimal policies may differ significantly among different projects [4][5]. Similarly, optimal policies can vary with project states in the process of construction. CPs are expected to identify a series of optimal policies assigned for each state, defined as a schedule for tunnel construction projects with resource constraints and uncertain ground conditions.

Some construction process optimization methods exist for linear construction projects, including linear programming (LP), integer programming (IP), dynamic programming (DP), genetic algorithm (GA), and simulation-based optimization [8][9][10][11]. However, these methods are not tailored for hard rock tunnel projects while considering uncertainties in ground conditions. A DP-based method has been developed to enable CPs to automatically generate excavation schedules in preconstruction and construction [12], but it fails to consider the impact of resource constraints. In addition, all of the mentioned models have limitations on selecting an optimal policy while considering multiple and comprehensive policies for different project states. Q-learning, a popular reinforcement learning (RL) approach, has the capacity to indicate optimal policies to guide agents' actions in given states. However, RL methods have not been tailored for resource-constrained tunnel construction scheduling problems. Therefore, the main practical problem is that CPs cannot yet rapidly generate excavation schedules with minimum durations (i.e., assign an optimal policy for each state) with a given equipment fleet while taking uncertain ground conditions into consideration.

This paper first reviews the existing studies to identify the research gap. A Q-learning-based construction process optimization method is then proposed and validated by a case study.

2. POINTS OF DEPARTURE

This section includes reviews of three aspects: (1) some common resource allocation policies, (2) existing construction process optimization methods, and (3) RL for scheduling. Our research question and objectives are then presented.

Resource allocation policies are efficient tools for dealing with the resource-constrained project scheduling problems. Various policies have been proposed with different emphases on resource allocation (e.g., activity duration, resource requirements, resource waiting time). They include but are not limited to First Come First Served (FCFS), Late Come First Served (LCFS), Short Operation First (SOF), Maximum Operation First (MOF), Minimum Total Work Content (MINTWK), Maximum Total Work Content (MAXTWK), etc. [2][3][4][5][6][7]. Because the optimal policies can change from one project state to another, it would be more comprehensive to include more policies to obtain shorter total project durations.

Some construction process optimization methods have been applied to linear projects. An IP model has been applied to a comprehensive earthmoving system in road construction optimization to minimize costs while considering earthwork deadlines [13]. In addition, several DP-based optimization methods have been developed for scheduling linear construction projects [14][15]. A practical optimization method has been developed to analyze time–cost trade-off by formalizing a GA procedure [16]. A simulation-based optimization model has been developed to optimize fleet selection for earthmoving operations by taking into account linear indirect project costs [17]. In addition, under given resource constraints for general construction projects, many researchers have developed optimization methods to determine optimal activity start times [18][19][20]. However, they are not tailored for hard rock tunnels with the consideration of uncertain ground conditions. A schedule generation and estimation methodology based on the DP method has been developed for hard rock tunnel projects, yet it failed to consider resource constraints [12]. In addition, all of them have limitations on the selection of optimal policies at decision-making points among multiple and comprehensive policies for different states.

RL has been widely used in production scheduling, because the intelligent agent in RL can provide a series of decisions (e.g., activity sequencing, resource allocation) after learning [21]. Recently, RL has also been applied in civil engineering, especially for complex problems [21]. For example, an n-step Q-network traffic signal controller has been trained by applying RL with function approximation to reduce vehicle queues [22]. An RL-based framework for strategy development for conventional tunneling is presented to excavate the rockmass in an optimal way from different excavation sequences [23]. A practical control framework for a building energy model has been proposed based on deep RL to reduce heat demand [24]. An RL-based energy management system has also been proposed to maximize the profit in residential energy sales [25]. Although RL shows the capacity to provide better solutions at each decision-making point, act with the environment, and rapidly find optimal actions when applied in similar environments, it has not been tailored for the construction scheduling problem for resource-constrained hard rock tunnels by considering uncertain ground conditions.

Our research question is as follows: ***How can construction planners formally and rapidly generate optimal schedules for resource-constrained hard rock tunnel projects by using the Q-learning method with multiple resource allocation policies involved?*** The research objectives of this paper are to (1) formalize a methodology for the schedule optimization of hard rock tunnel projects while considering uncertainties in ground conditions and resource constraints and (2) compare the schedule optimization results after Q-learning (with multiple policies involved) and the schedule simulation results under all involved single policies.

3. Q-LEARNING-BASED SCHEDULE OPTIMIZATION METHODOLOGY FOR RESOURCE-CONSTRAINED HARD ROCK TUNNEL

The research team adapts a Q-learning algorithm to solve the schedule optimization problem of resource-constrained hard rock tunnel projects. The definition of Q-learning is shown as Formula (1), where α and γ denote learning rate and discount factor respectively [26].

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha\left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\right] \quad (1)$$

The specification of the Q-learning algorithm in hard rock tunnel projects is shown in Figure 1. The environment in the tunnel scheduling problem should relate to the construction progress. To consider the uncertainties in ground conditions, multiple reference ground condition (GC) scenarios will be provided. A representative GC scenario will be generated based on the scenarios, which will be another part of the environment for Q-learning. Before learning, the Q is required to be initialized. If the states are terminals, the Q values are 0. Otherwise, the Q values are initialized to a fixed sufficiently large negative number. The definition of state in this specified Q-learning method ought to indicate the current excavation progress. For example, it can be a vector consisting of the locations of all under-excavated tunnel faces. During the process of Q-learning, for each episode, it starts from the same unexcavated initial state and goes through some iterations to update the excavation progress until the construction of the entire project is completed. For each iteration at time t, the agent (i.e., CPs) firstly selects an action A_t via Ɛ-greedy algorithm to decide whether the current iteration is for exploitation or exploration, in which actions are resource allocation policies to deal with the limitations brought by resource constraints. With the given state S_t and action A_t, the environment will be interacted with, which means the construction progress will be updated. After that, the environment will return the new state S_{t+1} and corresponding R_{t+1}, where the reward R_{t+1} equals the negative value of time interval required to move from old to new states. At the end of each iteration, the Q and time are updated.

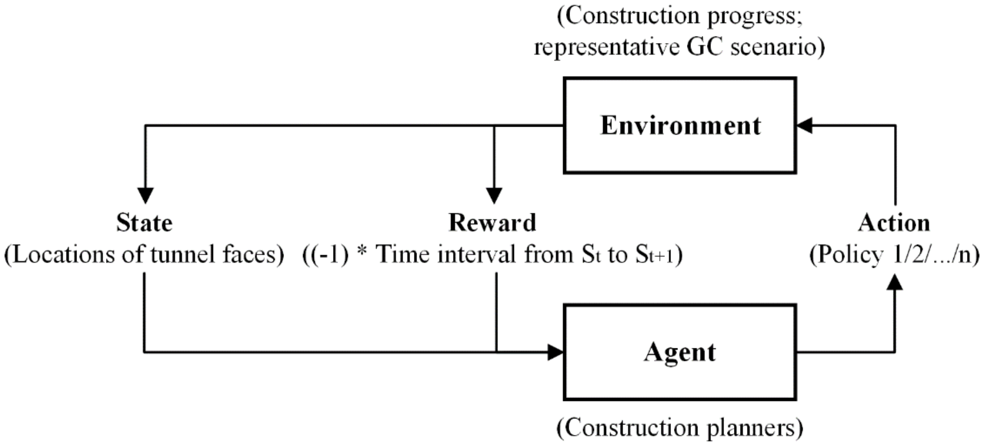
(Construction progress; representative GC scenario)
Environment
State
(Locations of tunnel faces)
Reward
((-1) * Time interval from St to St+1)
Action
(Policy 1/2/.../n)
Agent
(Construction planners)

Figure 1. Specification of Q-learning algorithm for resource-constrained hard rock tunnels

In the specified Q-learning methodology, the Q value of each (state, action) pair after learning is expected to represent the minimum remaining time required to finish the entire excavation from the current state after taking the action at this moment (expressed as a negative value). Therefore, the learning rate and discount factor are both set to 1 in this method.

After the learning of each episode, a state transition path will be recorded with the information of (1) states, (2) actions, and (3) rewards (i.e., time intervals between states). Q values of the (state, action) pairs on the path will be updated backward from the terminal state. The Q value of a pair will be updated if it is larger than its original value. If the maximum Q value of any state becomes larger during the process, the Q values of all prior (state, action) pairs, which can lead to this state, also need to be one-step updated (if required). After the backward focusing of planning computations, the Q values of all (state, action) pairs can indicate the minimum remaining durations to finish all excavations of the entire project from the current "state" after taking the corresponding "action" (except for the pairs of which the Q values are still the initialized). This method makes the optimization process more efficient.

4. CASE STUDY

The research team validated the proposed Q-learning-based schedule optimization methodology via a real case, which is a 6km long hard rock tunnel project in Korea. The layout of the simplified tunnel structure is shown in Figure 2. It includes 2 main tunnels, each of which is cut by a shaft. The total lengths of main tunnel 1, main tunnel 2, and the shaft are 6190m, 6175m, and 866m, respectively. It is assumed that the main tunnels can be two-way excavated, and the shaft can only be one-way excavated. Figure 2 also shows the identifications of the tunnel phases (from P1 to P9) assigned to different tunnel sections. After the completion of the shaft, the excavation of P2, P3, P6, and P7 can be started.

There are 5 types of GCs (i.e., very good, good, fair, poor, very poor) and 1 representative GC scenario generated via geostatistical method. The excavation method for the project is the drilling and blasting method. The advance rates without considering resource constraints in very good, good, fair, poor, and very poor GCs are 2.55, 2.32, 2.06, 1.79, and 1.28 meter per day, respectively. The project includes 3 zones. Zone 1 covers the construction sites of P1 and P5, zone 2 covers P9, P2, P3, P6 and P7, and zone 3 covers P4 and P8. Resources are shared among phases in the same zone. Only if the penetration of a main tunnel happens will the related zones be integrated and resources belonging to them be shared together. Three sets of equipment are allocated to zones 1, 2, and 3 evenly at the beginning of construction. The advance rates of the phases while considering resource constraints are assumed to be known information obtained from another paper about schedule evaluation methodology. The process for obtaining these advance rates is omitted here because it is not the focus of this paper.

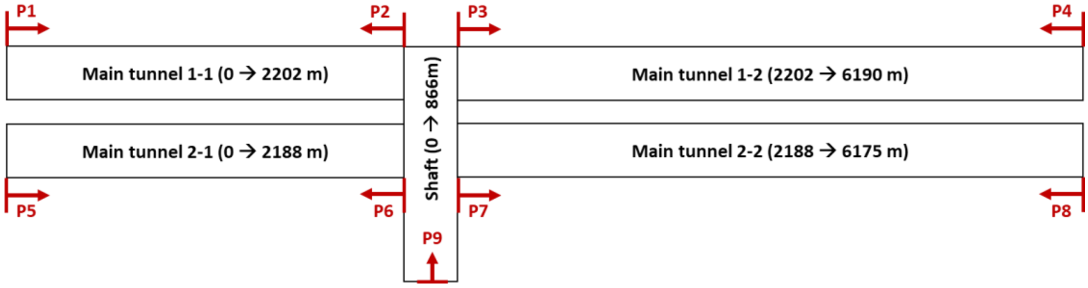
P1
P2
P3
P4
Main tunnel 1-1 (0 → 2202 m)
Main tunnel 1-2 (2202 → 6190 m)
Shaft (0 → 866m)
Main tunnel 2-1 (0 → 2188 m)
Main tunnel 2-2 (2188 → 6175 m)
P5
P6
P7
P8
P9

Figure 2. Layout of the simplified tunnel project

The state is defined as a vector as shown in the equation in function (2), where loc_i denotes the longitudinal location (rounded to the nearest integer) of the current tunnel face of P_i. The initial state is [0, 2202, 2202, 6190, 0, 2188, 2188, 6175, 0]. The terminal states are $[loc_1, loc_1, loc_3, loc_3, loc_5, loc_5, loc_7, loc_7, 866]$.

$$\begin{cases} S = [loc_1, loc_2, loc_3, loc_4, loc_5, loc_6, loc_7, loc_8, loc_9] \\ 0 \le \text{loc}_1 \le \text{loc}_2 \le 2202 \\ 2202 \le \text{loc}_3 \le \text{loc}_4 \le 6190 \\ 0 \le \text{loc}_5 \le \text{loc}_6 \le 2188 \\ 2188 \le \text{loc}_7 \le \text{loc}_8 \le 6175 \\ 0 \le \text{loc}_9 \le 866 \end{cases} \tag{2}$$

Five actions (i.e., resource allocation policies) are considered in this case: FCFS, SOF, MOF, MINTWK, and MAXTWK. In addition, agents follow the Ɛ-greedy algorithm to take actions, where Ɛ is set to 0.4.

The reward is defined as the time interval moving from the current state to the next state (expressed as a negative value).

There are 1000 episodes in total. In each episode, the excavation progress is updated every 100 days until the end. In other words, agents need to select policies for every 100-day interval, and resources will be allocated under the selected policy in the following 100 days. Table 1 shows the optimal path found after Q-learning with the information on the current state, optimal action, reward, and next state, in which state 1369 is the terminal state of this path. From Table 1, it can be calculated that the total project duration after Q-learning equals (-1) * accumulated rewards (i.e., time intervals between states). Therefore, the total project duration after Q-learning is 1144.5 days. Table 2 shows the summary of Q-learning results in 11 episodes among all 1000 episodes. The Q value in the initial state indicates the minimum remaining duration required to finish the construction of entire project. It decreases from 1172.46 to 1144.50 days and becomes increasingly stabilized. There are 2205 states in total after all learning. Furthermore, the number of states keeps increasing as the process of Q-learning proceeds. Around 150 new states will arise every 100 episodes (from episodes 700 to 1000). The expansion of the state space is due to the strict definition of state (a 9-dimension vector) to distinguish different states.

Table 1. Optimal path found by the Q-learning-based schedule optimization methodology

Current state		Action (optimal)	Reward	Next state
ID	Locations of tunnel faces			ID

0	[0,2202,2202,6190,0,2188,2188,6175,0]	all policies	-100	1
1	[191,2202,2202,6009,189,2188,2188,5983,401]	all policies	-100	2
2	[354,2202,2202,5776,340,2188,2188,5756,804]	FCFS/MOF	-100	3
3	[491,2104,2309,5576,476,2086,2273,5537,866]	MINTWK	-100	76
76	[657,1990,2422,5354,670,1970,2385,5342,866]	SOF	-100	659
659	[885,1854,2573,5133,888,1936,2514,5115,866]	MAXTWK	-100	988
988	[1109,1709,2733,4922,1108,1921,2660,4903,866]	MAXTWK	-100	1317
1317	[1329,1559,2870,4704,1309,1921,2829,4701,866]	MINTWK	-100	1330
1330	[1480,1480,3027,4491,1521,1772,2973,4483,866]	MINTWK	-100	1366
1366	[1480,1480,3252,4289,1642,1642,3176,4257,866]	all policies	-100	1367
1367	[1480,1480,3484,4105,1642,1642,3408,4044,866]	all policies	-100	1368
1368	[1480,1480,3713,3894,1642,1642,3643,3843,866]	SOF	-44.5	1369
1369	[1480,1480,3808,3808,1642,1642,3740,3740,866]	-	-	-

Table 2. Summary of Q-learning results

Episode ID	Q value under initial state	No. States
1	-1172.46	13
100	-1159.37	355
200	-1151.46	605
300	-1151.03	831
400	-1151.03	1062
500	-1150.82	1262
600	-1144.51	1550
700	-1144.51	1773
800	-1144.50	1914
900	-1144.50	2071
1000	-1144.50	2205

The result after Q-learning-based schedule optimization is compared with the simulation results under 5 single policies (Table 3). The optimized project duration after Q-learning is shorter than the simulation results under all 5 single policies. Q-learning shows an advantage from 1.39% to 8.15% over them. Especially for FCFS, which is the most commonly used resource allocation policy for construction projects, the Q-learning result is 2.38% shorter than FCFS.

Table 3. Comparison between Q-learning result and simulation results under single policies

Single policy		Q-learning result (day)	Q-learning vs single policy
Name	Simulation result (day)		
FCFS	1172.46	1144.50	2.38%
SOF	1160.63		1.39%
MOF	1171.45		2.30%
MINTWK	1174.42		2.55%
MAXTWK	1246.06		8.15%

5. CONCLUSIONS AND FUTURE WORKS

This paper has proposed Q-learning-based schedule optimization methodology for resource-constrained hard rock tunnel projects by considering uncertainties in ground conditions. In the specified Q-learning method, state is defined as a vector to indicate current excavation progress, action includes all considered resource allocation policies, and reward is the time interval moving from the current state to the next state. Due to the setting of values of the learning rate and discount factor, the Q value of each (state, action) pair reflects the minimum remaining duration to finish the excavation of the entire project. During the Q-learning process, the backward focusing of planning computations is applied to make the learning more efficient. After that, it is validated via a real case study. The optimal path is generated after Q-learning. In addition, the Q-learning results show advantages ranging from 1.39% to 8.15% over the simulation results under 5 single policies. Specifically, the Q-learning result is 2.38% shorter than FCFS, which is the most common resource allocation policy. The proposal of the formal Q-learning-based schedule optimization methodology can help CPs rapidly generate excavation schedules of minimum durations under a given equipment fleet while considering uncertainties in ground conditions and appropriately assign optimal policy for the state with hints from Q tables.

There are still some limitations of this study. It can be observed from the case study that the total number of states steadily increases in the process of Q-learning with the number of episodes. This finding stems from the relatively strict definition of state in this paper. In reality, two highly similar states have really close remaining durations to finish all excavations, yet they are identified and treated as 2 independent states in this context. The vast state space can lead to many state–action pairs not being sufficiently explored, which may overlook better solutions. Therefore, future works will focus on the reduction of the state space and the improvement of optimization efficiency (e.g., through function approximation methods).

ACKNOWLEGEMENTS

The support of the General Research Fund of the University Grants Committee of Hong Kong (#21212019) and City University of Hong Kong (#9239027) is gratefully acknowledged.

REFERENCES

[1] C. Paraskevopoulou and A. Benardos, “Assessing the construction cost of Greek transportation tunnel projects.,” Tunnelling and Underground Space Technology. vol. 38, pp. 497–505, 2013.

[2] I. Kurtulus and E.W. Davis, “Multi-Project Scheduling: Categorization of Heuristic Rules Performance.,” Management Science. vol. 28, no. 2, pp. 161–172, 1982.

[3] S. Tsubakitani and R.F. Deckro, “A heuristic for multi-project scheduling with limited resources in the housing industry.,” European Journal of Operational Research. vol. 49, no. 1, pp. 80–91, 1990.

[4] D.M. Tsai and H.N. Chiu, “Two heuristics for scheduling multiple projects with resource constraints.,” Construction Management and Economics. vol. 14, no. 4, pp. 325–340, 1996.

[5] Y. Wang, Z. He, L.P. Kerkhove, and M. Vanhoucke, “On the performance of priority rules for the stochastic resource constrained multi-project scheduling problem.,” Computers and Industrial Engineering. vol. 114, no. October, pp. 223–234, 2017.

[6] H.J. Chen, G. Ding, J. Zhang, and S. Qin, “Research on priority rules for the stochastic resource constrained multi-project scheduling problem with new project arrival.,” Computers and Industrial Engineering. vol. 137, no. August, p. 106060, 2019.

[7] P.H. Chen and S.M. Shahandashti, “Hybrid of genetic algorithm and simulated annealing for multiple project scheduling with multiple resource constraints.,” Automation in Construction. vol. 18, no. 4, pp. 434–443, 2009.

[8] B. Said, “EARTHWORK ALLOCATIONS WITH LINEAR UNIT COSTS.,” vol. 114, no. 4, pp. 641–655, 1989.

[9] M. Marzouk and O. Moselhi, “Multiobjective Optimization of Earthmoving Operations.,” Journal of Construction Engineering and Management. vol. 130, no. 1,

pp. 105–113, 2004.
[10] K. El-Rayes and A. Kandil, "Time-Cost-Quality Trade-Off Analysis for Highway Construction.," Journal of Construction Engineering and Management. vol. 131, no. 4, pp. 477–486, 2005.
[11] P.G. Ipsilandis, "Multiobjective Linear Programming Model for Scheduling Linear Repetitive Projects.," Journal of Construction Engineering and Management. vol. 133, no. 6, pp. 417–424, 2007.
[12] J.I. Kim, M. Fischer, and C. Kam, "Generation and evaluation of excavation schedules for hard rock tunnels in preconstruction and construction.," Automation in Construction. vol. 96, no. October, pp. 378–397, 2018.
[13] A.K.W. Jayawardane and F.C. Harris, "FURTHER DEVELOPMENT OF INTEGER PROGRAMMING IN EARTHWORK OPTIMIZATION.," vol. 116, no. 1, pp. 18–34, 1990.
[14] N.N. Eldin and A.B. Senouci, "Scheduling and control of linear projects.," Canadian journal of civil engineering. vol. 21, no. 2, pp. 219–230, 1994.
[15] A. Myat, M. Paing, and N.L. Thein, "OPTIMIZING RESOURCE UTILIZATION FOR REPETITIVE CONSTRUCTION PROJECTS.," vol. 4, no. 3, pp. 85–99, 2012.
[16] T. Hegazy, "Optimization of construction time-cost trade-off analysis using genetic algorithms.," Canadian Journal of Civil Engineering. vol. 26, no. 6, pp. 685–697, 1999.
[17] A. Alshibani and O. Moselhi, "Fleet selection for earthmoving projects using optimization-based simulation.," Canadian Journal of Civil Engineering. vol. 39, no. 6, pp. 619–630, 2012.
[18] P. Ghoddousi, E. Eshtehardian, S. Jooybanpour, and A. Javanmardi, "Multi-mode resource-constrained discrete time-cost-resource optimization in project scheduling using non-dominated sorting genetic algorithm.," Automation in Construction. vol. 30, pp. 216–227, 2013.
[19] L.D. Long and A. Ohsato, "Fuzzy critical chain method for project scheduling under resource constraints and uncertainty.," International Journal of Project Management. vol. 26, no. 6, pp. 688–698, 2008.
[20] M. Lu, H.C. Lam, and F. Dai, "Resource-constrained critical path analysis based on discrete event simulation and particle swarm optimization.," Automation in Construction. vol. 17, no. 6, pp. 670–681, 2008.
[21] N.S. Kedir, S. Somi, A.R. Fayek, and P.H.D. Nguyen, "Hybridization of reinforcement learning and agent-based modeling to optimize construction planning and scheduling.," Automation in Construction. vol. 142, no. July, p. 104498, 2022.
[22] W. Genders and S. Razavi, "Asynchronous n-step Q-learning adaptive traffic signal control.," Journal of Intelligent Transportation Systems: Technology, Planning, and Operations. vol. 23, no. 4, pp. 319–331, 2019.
[23] G.H. Erharter, T.F. Hansen, Z. Liu, and T. Marcher, "Reinforcement learning based process optimization and strategy development in conventional tunneling.," Automation in Construction. vol. 127, no. December 2020, p. 103701, 2021.
[24] Z. Zhang, A. Chong, Y. Pan, C. Zhang, and K.P. Lam, "Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning.," Energy and Buildings. vol. 199, pp. 472–490, 2019.
[25] H. Berlink, N. Kagan, and A.H. Reali Costa, "Intelligent Decision-Making for Smart Home Energy Management.," Journal of Intelligent and Robotic Systems: Theory and Applications. vol. 80, pp. 331–354, 2015.
[26] R.S. Sutton and A.G. Barto, "Reinforcement Learning: An Introduction.," MIT press, 2018.