A Survey of Machine Learning for Urban Decision Making: Applications in Planning, Transportation, and Healthcare

YU ZHENG, QIANYUE HAO, JINGWEI WANG, CHANGZHENG GAO, JINWEI CHEN, DEPENG JIN, and YONG LI, Tsinghua University, China

Developing smart cities is vital for ensuring sustainable development and improving human well-being. One critical aspect of building smart cities is designing intelligent methods to address various decision-making problems that arise in urban areas. As machine learning techniques continue to advance rapidly, a growing body of research has been focused on utilizing these methods to achieve intelligent urban decision making. In this survey, we conduct a systematic literature review on the application of machine learning methods in urban decision making, with a focus on planning, transportation, and healthcare. First, we provide a taxonomy based on typical applications of machine learning methods for urban decision making. We then present background knowledge on these tasks and the machine learning techniques that have been adopted to solve them. Next, we examine the challenges and advantages of applying machine learning in urban decision making, including issues related to urban complexity, urban heterogeneity and computational cost. Afterward and primarily, we elaborate on the existing machine learning methods that aim to solve urban decision making tasks in planning, transportation, and healthcare, highlighting their strengths and limitations. Finally, we discuss open problems and the future directions of applying machine learning to enable intelligent urban decision making, such as developing foundation models and combining reinforcement learning algorithms with human feedback. We hope this survey can help researchers in related fields understand the recent progress made in existing works, and inspire novel applications of machine learning in smart cities.

CCS Concepts: • **Computing methodologies** → **Machine learning**; **Planning and scheduling**; • **Applied computing** → **Transportation**.

Additional Key Words and Phrases: machine learning, urban planning, optimization, decision making

ACM Reference Format:
Yu Zheng, Qianyue Hao, Jingwei Wang, Changzheng Gao, Jinwei Chen, Depeng Jin, and Yong Li. 2024. A Survey of Machine Learning for Urban Decision Making: Applications in Planning, Transportation, and Healthcare. *ACM Comput. Surv.* 1, 1 (November 2024), 37 pages. https://doi.org/10.1145/3695986

1 INTRODUCTION

With rapid urbanization, cities now host more than half of the world's population and are centers of economic activity [261]. Thus, utilizing the power of advanced technologies to build smart cities is critical for achieving sustainable development and improving living standards. In particular, smart cities give rise to a variety of decision-making tasks that can have a significant impact on the development of cities. For example, intelligent planning of facilities in the city can make residents access various services in close proximity, significantly improving the efficiency of the city's

Authors' address: Yu Zheng, y-zheng19@mails.tsinghua.edu.cn; Qianyue Hao, hqy22@mails.tsinghua.edu.cn; Jingwei Wang, jingwei22@mails.tsinghua.edu.cn; Changzheng Gao, gcz20@mails.tsinghua.edu.cn; Jinwei Chen, cjw20@mails.tsinghua.edu.cn; Depeng Jin, jindp@tsinghua.edu.cn; Yong Li, liyong07@tsinghua.edu.cn, Tsinghua University, Beijing, China.

Manuscript submitted to ACM

operation. In addition, decision regarding vehicles and traffic lights can reduce traffic congestion and air pollution, lower carbon emissions, and reduce commuting times for residents. Furthermore, making informed decisions about the allocation of medical resources and the movement of people can help to control the spread of infectious diseases in cities and safeguard public health. In general, building intelligent models to solve urban decision-making tasks lies at the core of realizing the full potential of smart cities.

Researchers have proposed numerous solutions to address the long-standing urban decision-making tasks. In the past, due to computational limitations and a lack of data, traditional methods such as meta-heuristics, genetic algorithms, and mixed integer optimization were primarily used. However, cities in reality are much complicated systems, rendering the results of these methods far from optimal. In recent years, as data collection and storage capabilities have increased and computing power has leaped forward, machine learning methods, including deep learning (DL) and reinforcement learning (RL), have made remarkable strides. In particular, neural networks in DL methods can handle multivariate inputs and capture complex high-dimensional nonlinear relationships. When combined with RL, they enable value estimation and strategy search in huge action spaces, solving many decision-making problems that were previously thought to be intractable for machines. These technological advances have provided new tools for intelligent urban decision making, resulting in the state-of-the-art solutions in many tasks and producing a significant number of research articles. As smart city research is still emerging, it is necessary to conduct a review of articles that apply machine learning to solve urban decision-making tasks.

In this survey, we aim to provide a systematic literature review on existing approaches of machine learning for urban decision making. In fact, there are diverse decision-making scenarios in smart city, organically interconnected and influenced by each other, which together determine the dynamics of the city. We conduct a systematic literature review of over 160 research papers published in the past six years from mainstream journals and conferences. After thorough investigation of these papers, we identify three most relevant and typical use cases of machine learning in urban decision making, which influence cities at different time scales. **Firstly, on the long-term scale**, urban planning decisions such as land use and road layout essentially determine how residents use the city, incurring far-reaching effects on urban dynamics measured by years. **Secondly, on the medium-term scale**, urban healthcare decisions such as pandemic spreading control strategies shape the urban dynamics through periodic assessments and targeted interventions measured by days. **Thirdly, on the short-term scale**, urban transportation decisions like traffic light control and vehicle dispatching directly affects the mobility flow, changing urban dynamics in real time measured by seconds. Therefore, we propose a taxonomy of research topics based on the application tasks in smart cities, focusing on the above three major decision-making tasks with the most number of publications, namely planning, transportation, and healthcare. For other decision-making scenarios in the city, we can not cover everything in a single survey, but they are also promising research directions, and we refer readers to other scenario-specific surveys [2, 47, 57, 96, 179, 245].

For each decision-making task, we provide detailed introduction of typical approaches, and give the necessary preliminary background, including the problem formulation of each task, as well as the technical information of the adopted machine learning methods. Meanwhile, we analyze the challenges and advantages of applying machine learning methods in these tasks, and discuss open problems and future directions in intelligent urban decision making. Fig. 1 illustrates the structure of this survey. Ultimately, our goal is to inform policymakers, practitioners, and other stakeholders on how they can leverage machine learning to improve the quality of life for urban residents, enhance sustainability, and advance the state of the art in these domains. The contributions of this survey are as follows:

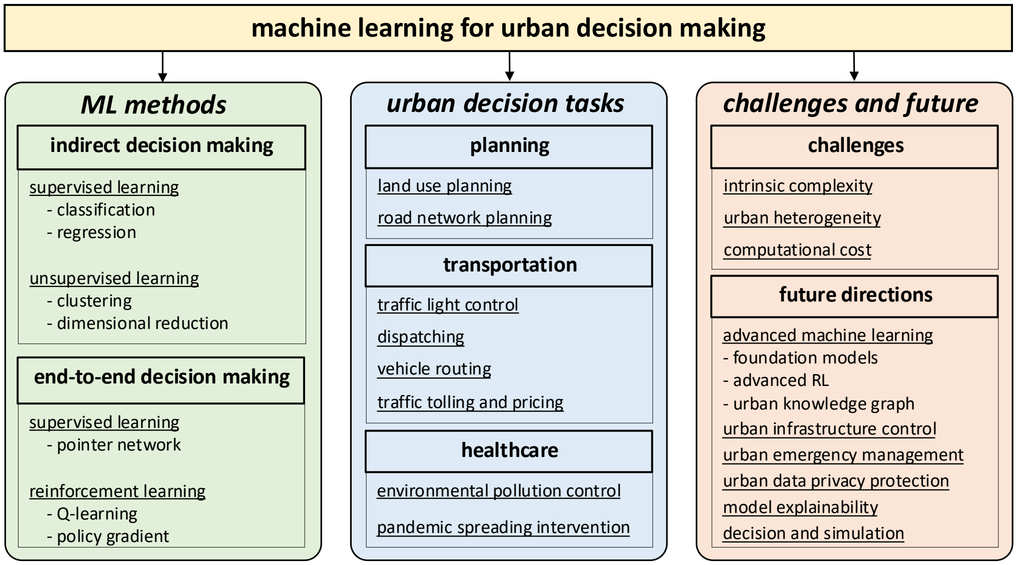


Fig. 1. Overall structure of this survey.

Paper	Problems	ML methods	Applications	Example Tasks
[39]	prediction analysis	DL SL	transportation healthcare environment public safety	transportation flow prediction medical imaging air quality prediction vehicle detection
[8]	prediction analysis	DL SL	home healthcare transportation surveillance environment	energy monitoring disease prediction vehicle mobility prediction fire detection garbage detection
[211]	clustering analysis	DL UL	urban sustainability urbanization and regional study built environment urban dynamics	flood mapping remote sensing function and morphology study urban behavior pattern study
ours	decision-making	RL DL SL & UL	planning transportation healthcare	road network design traffic light control medical resources allocation

Table 1. Comparison with related reviews in smart cities

- We propose a taxonomy of representative machine learning approaches in urban decision making, including planning, transportation, and healthcare, which covers the most relevant perspectives that influence the urban dynamics over long-term, short-term, and medium-term, respectively.
- We provide a comprehensive view of existing paradigms of machine learning techniques in urban decision making, summarizing its challenges and advantages. Besides systematically elaborating on the *status quo* of machine learning in urban decision-making, we provide guidance for future directions.
- Compared to the related surveys regarding **supervised learning (SL)** and **unsupervised learning (UL)** for urban prediction and analysis (see Table 1), we not only focus on the different decision tasks but also incorporate recent advances in machine learning, such as novel deep **reinforcement learning (RL)** algorithms.

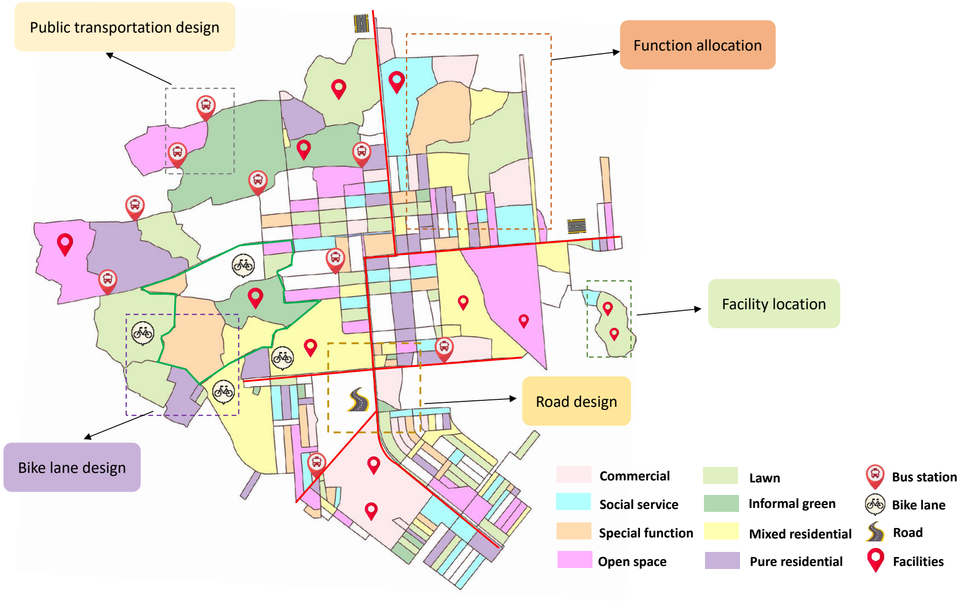


Fig. 2. An illustration of sub-problems in urban planning.

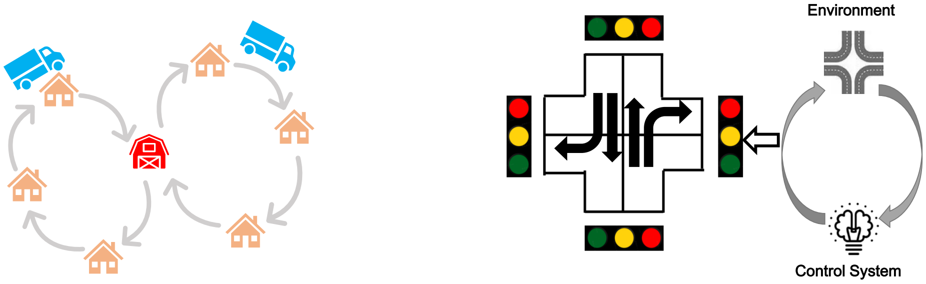
The paper is organized as follows: We first introduce the background of three urban decision-making tasks in Section 2. Next we provide necessary preliminaries of adopted machine learning methods in Section 3. After discussing the challenges of utilizing machine learning techniques in urban decision-making, we elaborate on existing representative methods in Section 4. Finally, we discuss several open problems and provide insights of future directions in Section 5 and conclude the survey in Section 6.

2 BACKGROUND

2.1 Decision in urban planning

Urban planning is a process of designing the spatial arrangement of cities to improve the quality of life for residents. As human activities and environmental factors significantly impact urban planning schemes, traditional approaches often require field surveys, which can be time-consuming and costly. Specifically, it is challenging for traditional optimization methods to deal with problems with large scales and numerous constraints, which are common in practical planning problems. To address these issues, researchers have increasingly turned to machine learning methods, due to their powerful ability to solve nonlinear problems. As demonstrated in Fig. 2, machine learning for urban planning can be broadly categorized into five sub-problems: function allocation, facility location, road (general lane) design, bike lane (exclusive lane) design, and public transportation design. The first two fall under the umbrella of land-use planning, while the remaining three concern road network planning.

Land use planning. Aiming to maximize the comprehensive benefits for urban residents by rationally grouping urban land, urban land use planning entails two critical issues: urban function allocation and facility location planning. Urban function allocation divides urban land into logical and contiguous functional areas, such as residential and commercial areas, while facility location planning selects and allocates land for public facilities, such as hospitals,



(a) Vehicle Routing Problem

(b) Traffic Signal Control

Fig. 3. An illustration of sub-problems in traffic optimization.

schools, and charging stations. Both sub-tasks significantly impact residents' daily commuting costs and living expenses, as unreasonable division schemes and facility locations can greatly reduce the accessibility of public services.

Road network planning. Urban road network planning aims to minimize the average travel cost for urban residents by expanding or reconstructing the traffic network, which can be categorized into road design, bike lane design, and public transportation design according to the planning content. Road design focuses on planning ordinary roads to ensure connectivity between regions, serving as the foundation of other planning efforts. Bike lane and public transportation design aim to reduce dependence on automobiles after establishing the road network, focusing on selecting cycling routes and locating public stops that meet the population's travel needs. Urban road network planning is a bi-level optimization problem that concerns policy-makers and traffic participants, where the decisions made by policy-makers at the top level affect the behavior of the participants at the lower level, and vice versa, rendering road network planning a challenging problem. Specifically, traffic participants generate traffic behaviour in the road network designed by policy-makers, and these traffic behaviours are also feedback for decision-makers to evaluate whether the road network is reasonable so that decision-makers can develop a road network of higher quality.

2.2 Decision in urban transportation

Cities are encountering a rising number of vehicles on the roads, posing significant challenges to urban transportation, such as traffic congestion and air pollution. The purpose of urban transportation decision is to mitigate these issues by effectively managing the traffic flow, such that cities can provide safe, reliable, and sustainable travel for their inhabitants. A large number of studies have emerged which leverage machine learning to address urban transportation decision, optimizing from various perspectives. For example, the flow of vehicles can be directly controlled by dynamically adjusting traffic lights. Meanwhile, the traffic flow can be indirectly controlled by modifying road tolls and pricing according to real-time traffic conditions. Here we concentrate on four primary tasks in urban transportation decision: traffic light control, vehicle routing problem, dispatching problem, and traffic tolling and pricing problem, which have been extensively researched in both academia and industry, with a broad range of applications in modern urban services, such as logistics and ride-hailing scenarios.

Vehicle Routing Problem (VRP). VRP is an extension of the Traveling Salesman Problem (TSP) which is also an NP-hard problem, aiming to design the shortest routing of vehicles given several destinations [24]. This problem is extensively applied in urban transportation and has great significance to reduce vehicle energy consumption. Almost all VRP problems are derived from the fundamental Capacitated Vehicle Routing Problem (CVRP) [159]. As shown in Fig. 3(a), in CVRP, before the vehicle eventually return to the depot, it needs to visit several customers with different demands. We can use $G = (V, E)$ to represent the graph formed by depot and customers. $V = \{v_0, v_1, ..., v_N\}$ is the node

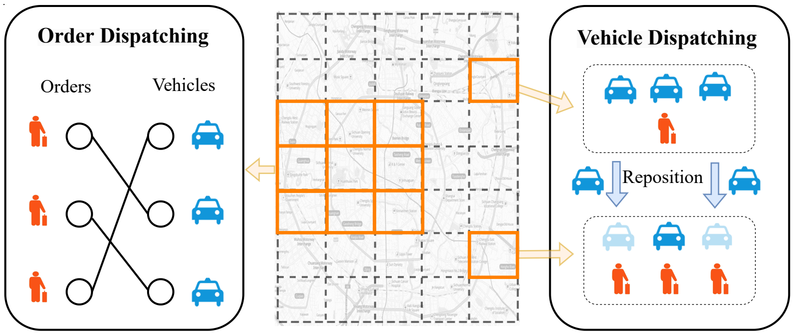


Fig. 4. An illustration of sub-problems in dispatching.

set representing the depot and customers, where v_0 represents depot, and E is the set of edges. Since the capacity of the vehicle is limited (represented as C), the vehicle may return to the depot multiple times during the course of its mission. We call the sequence of each time a vehicle leaves and returns to depot as a tour and the m-th tour can be denoted as $\pi_m = (v_0, v_{m,1}, ..., v_{m,t}, ...,)$ where $v_{m,t}$ denotes the t-th customer that vehicle visits in m-th tour. Then, the whole routing strategy of vehicle can be represented as $\pi = \{\pi_1, \pi_2, ..., \pi_m, ...\pi_K\}$. Given the above definitions, CVRP is to find a route strategy to minimize the total distance of all vehicles with all customer demands fulfilled. We denote the overall distance cost of the m-th tour of the vehicle as $D(\pi_m)$. CVRP can be formulated as belows:

$$\min \sum\nolimits_{m=1}^{K} D_(\pi_m), \tag{1}$$

$$s.t. \quad \pi_1 \cup \pi_2 \cup \cdots \cup \pi_m \cdots \cup \pi_K = V, \tag{2}$$

$$\pi_m \cap \pi_n = \{v_0\}, \forall m, n \leq K, m \neq n, \tag{3}$$

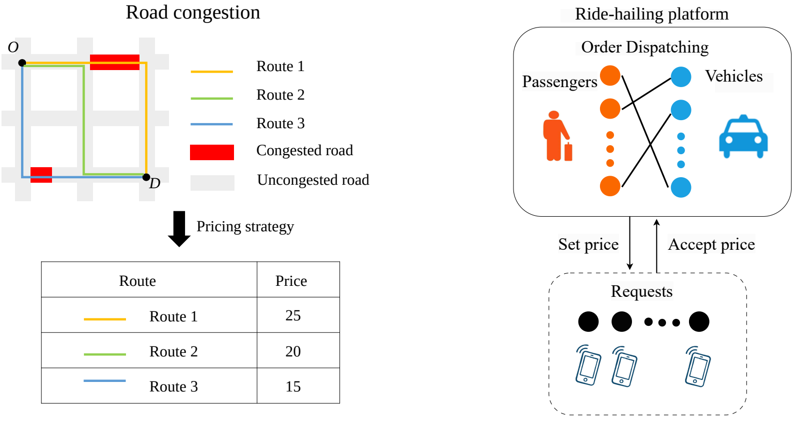
$$\sum\nolimits_{i=2}^{|\pi_m|} d_{m,i} \leq C, \forall m \leq K \tag{4}$$

where $d_{m,i}$ is the demand of customer $v_{m,i}$. Constraints (2) and (3) indicate that all customers must be visited and only visited once with demands satisfied, and constraint (4) indicates that a vehicle cannot carry more goods than its capacity.

Traffic Light Controlling. Fig. 3(b) depicts how traffic in the city can navigate through intersections controlled by traffic signals, which regulate vehicular and pedestrian movement. Green lights indicate vehicle and pedestrian movement, whereas red lights signal them to stop until the light turns green. The main objective of traffic signal control is to reduce vehicle wait times and queue lengths at intersections by modifying traffic light states, including signal phase and time intervals, based on various relevant factors [13].

Dispatching. As illustrated in Fig. 4, the dispatching task consists of order dispatching and vehicle repositioning, which can be modeled and optimized jointly, as they both have an impact on the distribution of vehicles and passengers.

Order Dispatching. As illustrated in Fig. 4 left, order dispatching involves matching orders with workers, where the definition of workers is context-dependent. For example, in the ride-hailing scenario, a worker refers to a driver, while in the express and food delivery scenarios, a worker denotes a rider. The task of order dispatching can be formulated mathematically using a dynamic bipartite graph $\mathbf{G}$, consisting two types of nodes, worker nodes $\mathbf{N_w}$ and order nodes $\mathbf{N_o}$, whose node information changes dynamically over time. The process of order dispatching is to connect edges between nodes of workers and orders. Notably, there are different restrictions on the edges in different scenarios. For instance, in the ride-hailing scenario, a driver typically serves only one order, so one worker node can only connect to one order node. Conversely, in the express and food delivery scenarios, one worker can receive multiple orders, so one worker node can be connected to multiple order nodes. There are usually three aspects of optimization objectives:



(a) traffic tolling

(b) pricing problem

Fig. 5. Illustrations of traffic tolling and pricing.

long-term income of the platform (*e.g.* total income of all workers), passenger experience (*e.g.* order response rate), and worker fairness (*e.g.* the lowest income of different workers).

Vehicle dispatching. As shown in Fig. 4 right, vehicle repositioning, also known as fleet management, re-positions resources (vehicles or bikes) to balance supply and demand. This task is generally combined with order dispatching tasks using rule-based methods, with the goal to maximize the total income of the platform. Although vehicle dispatching does not produce immediate rewards, the reward function is introduced through the price of finishing the task of order dispatching. As a result, the platform can obtain more income by achieving a balance between supply and demand.

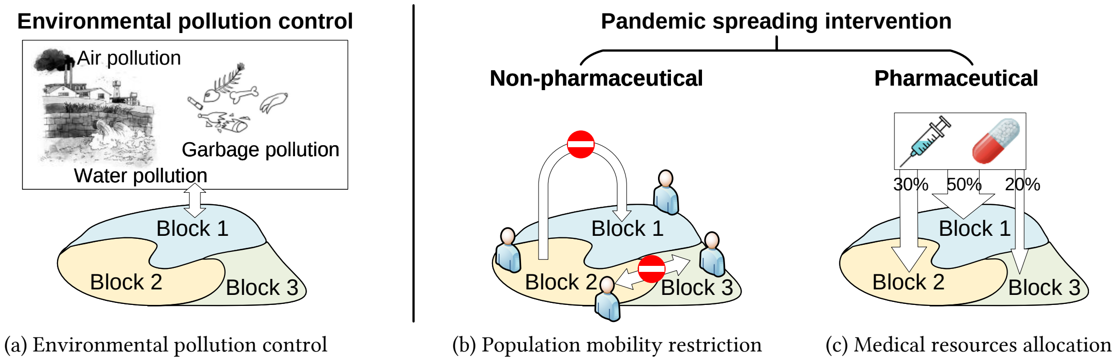
Traffic Tolling and Pricing. Dynamic tolling and pricing strategies based on the traffic situation can help alleviate traffic congestion and balance supply and demand distribution, which are critical for improving traffic efficiency.

Traffic Tolling. As shown in Fig. 5(a), traffic tolling involves charging travelers during specific time periods and in congested areas, in order to compensate for the economic losses caused by traffic congestion. Meanwhile, it can incentivize travelers to choose cheaper routes, thereby reducing traffic congestion. In essence, it aims to set a reasonable price based on the congestion condition of each road or region and guide people's travel intention.

Traffic Pricing. In ride-hailing and taxis services, dynamic pricing plays a crucial role in balancing supply and demand. As shown in Fig. 5(b), the ride-hailing platform sets the price based on the traffic scenario and travel distance given a travel request sent by a passenger, and drivers will enter the order matching list after accepting the price. During peak times, when demand exceeds supply, increasing prices can attract more drivers to *hot* areas and serve more passengers. Moreover, passengers may modify their travel demands based on the price changes. Closely linked to the supply and demand, it can be optimized together with order or vehicle dispatching.

2.3 Decision in urban healthcare

The generalized concept of healthcare refers to the science of preventing disease, prolonging life, and promoting health and efficiency through organized community efforts for the sanitation of the environment [198]. With the fast development of modern life and industry, the problem of environmental pollution is becoming increasingly serious, urgently calling for efficient measures to control the pollutant and maintain a clean living environment for humans. There exists extensive researches on environmental pollution control [89, 125, 160, 201], making efforts toward an earth which is free from pollution. On the other hand, the pandemic of COVID-19 has swept the globe in the past three years and caused billions of infections, warning us of the importance and necessity of continuous efforts in public health to



(a) Environmental pollution control (b) Population mobility restriction (c) Medical resources allocation

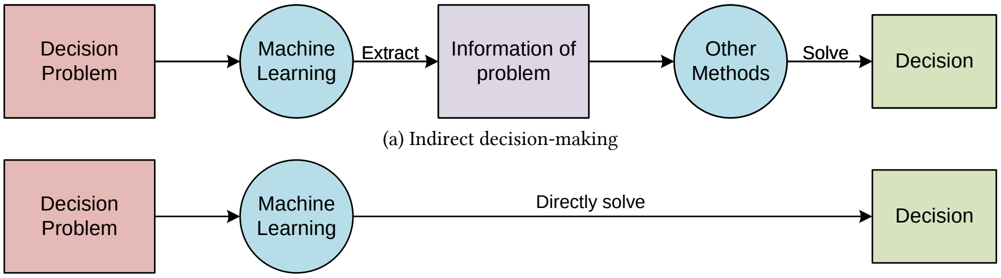
Fig. 6. An illustration of sub-problems in urban healthcare.

prevent the pandemic spreading and thus guarding humans' well-being. Plentiful researches on public health decisions for pandemic spreading intervention keep emerging during these years [67, 102, 141, 143, 191], trying to help the global fight against the COVID-19 pandemic. Therefore, in this survey, we focus on two of the most important items in urban healthcare, controlling the pollution and intervening the pandemic, as illustrated in Fig. 6.

Environmental pollution control. As shown in Fig. 6(a), the fast development of cities leads to environmental pollution from various sources, which profoundly threatens public health and people's well-being. First, air pollution is among the most common environmental pollutants, severely impacting everyone's daily life. Sources of air pollution include vehicle exhaust, industrial emissions, domestic emissions, etc., and the toxic substances in them can cause various diseases, including lung cancer. The control of air pollution includes manners from various aspects, such as monitoring [265] and forecasting [106, 188] the air qualify, optimizing transportation [54], and purifying indoor air [40, 69]. Second, water pollution is another environmental pollution mainly caused by industrial wastewater discharging, threatening people's drinking water safety. The main approaches to control water pollution include limiting wastewater discharge [196] and improving wastewater treatment efficiency [38, 241]. Besides, garbage pollution is also worthy of attention. Proper garbage recycling contributes to environmental protection, while improper garbage processing may lead to secondary pollution [146], causing even worse impacts on public health.

Pandemic spreading intervention. Due to the increasing population density and mobility, infectious diseases tend to spread faster in urban areas [181], where there are thousands of blocks and millions of people. The pandemic spreading intervention includes a non-pharmaceutical approach, i.e., population mobility restriction, and a pharmaceutical one, i.e., medical resources allocation. The decisions in these two sub-problems are all made based on the historical pandemic spreading situation and the intrinsic features of the urban area. The historical pandemic spreading situation includes the changing of the number of infections and deaths from the beginning of the pandemic (t_0) to the current time (t), and we denote it as $S[t_0, t]$. The intrinsic features describe the dynamics and structure of the urban area, including static features and dynamic ones. Static features do not change in a short time, such as population density distribution and population age structure, which we denote as F. Dynamic features frequently change over time, such as population mobility and contact, where we use $F[t_1, t_2]$ to denote the features from t_1 to t_2.

As shown in Fig. 6(b), population mobility restriction temporarily shutdowns or proportionally reduces the population flow between certain areas. Mathematically, when considering an urban area with n blocks and making decision at time t, the inputs are $S[t_0, t]$, F and $F[t_0, t]$, while the output is $\mathbf{M}_{n\times n}[t]$. The element $\mathbf{M}_{ij}[t] \in [0, 1]$ refers to the proportion of mobility restriction between block i and j, where 1 corresponds to no restriction, and 0 corresponds to



(a) Indirect decision-making

(b) End-to-end decision-making

Fig. 7. An illustration of two approaches of machine learning for decision-making.

complete shutdown. Since many infectious diseases, such as COVID-19, are mainly airborne, the mobility and contact of people is the main approach of the pandemic spreading. Therefore, reducing population mobility and contact is one of the most essential and efficient non-pharmaceutical approaches to intervene the pandemic spreading [7, 33, 49].

As shown in Fig. 6c, medical resources allocation is to allocate the limited number of medical resources in the urban area to maximize its utility and thus minimize the damage caused by the pandemic. Mathematically, when considering an urban area with n blocks and allocating m kinds of medical resources at time t, the inputs are $S[t_0, t]$, F and $F[t_0, t]$, while the output is $\mathbf{A}_{m\times n}[t]$. The element $\mathbf{A}_{kj}[t] \in [0, 1]$ refers to the proportion of the k-th kind of resource allocated to block j, and the elements satisfy the constraint of total available number, i.e., $\sum_j \mathbf{A}_{kj}[t] = 1$. When facing a sudden pandemic outbreak, critical medical resources are very likely to suffer severe shortage [139, 170]. Therefore, efficient allocation decisions on medical resources, such as vaccines [53, 138], ventilators [14, 153], hospital beds [52], play a significant role in intervening the pandemic spreading.

3 MACHINE LEARNING METHODS FOR DECISION

In this section, we briefly review the machine learning methods frequently applied in decision-making. Here, we mainly focus on methods with deep neural networks, i.e., deep learning, which has achieved huge success in urban decision making. We view the machine learning methods from the angle of function estimation, where the input x is processed by the network, which is denoted as F_θ with learnable parameters θ. The network is essentially a complex function, and we denote its output as $F_\theta(x)$. The training process in machine learning is actually estimating a given target value y with the function $F_\theta(x)$ by adjusting the learnable parameters θ, which can be mathematically expressed as:

$$\theta^* = \arg\min_\theta DIFF\{F_\theta(x), y\}, \tag{5}$$

where $DIFF$ is the loss function for measuring the difference between the output of the function and the target value. There are several kinds of frequently used loss functions, such as the mean square error (MSE) [5], the negative log-likelihood (NLL) [97], and the cross entropy (CE) [137]. The most common way of adjusting the learnable parameters θ is stochastic gradient descent [163], where some modified gradient optimizers are designed, such as the Nesterov momentum [145] and the Adam [88] to improve the efficiency and the stability of the learning process.

According to the training paradigm, machine learning methods can be roughly divided into 3 categories, supervised learning, unsupervised learning, and reinforcement learning. Typically, reinforcement learning is used for end-to-end decision-making, while the remaining two categories are used for extracting valuable information from the problem and thus assisting decision-making [16]. Fig. 7 illustrates the two approaches of machine learning for decision-making.

3.1 Machine learning for indirect decision-making

Supervised learning and unsupervised learning are typically used for extracting information from the raw problem and thus assisting the decision-making. Supervised learning is the most common training paradigm in machine learning, which typically solves two kinds of problems, i.e., classification problems and regression problems. The target values y in supervised learning are the exact values of ground truth in the corresponding problem. In the classification problems [92], the goal is to predict a class label according to each input x where y is the ground truth label of the class that x belongs to. In the regression problems [29], the goal is to predict a corresponding value according to each input x where y is the ground truth of the value that corresponds to x. If the ground truth is available in a certain problem, supervised learning is a good choice, utilizing the ground truth to adjust the parameters θ and thus minimizing the difference between the ground truth and the prediction of the network. On the other hand, unsupervised learning is another paradigm where no external signal is provided, e.g., the ground truth in supervised learning and the reward in reinforcement learning. Instead of minimizing the difference between the output of the function and the target value, unsupervised learning typically directly optimizes a given metric function constructed according to the intrinsic features of the input x. For example, clustering methods aim to maximize the similarity within each cluster and the dissimilarity among different clusters, such as the Kmeans [66] approach while some dimensional reduction methods maximize the variance of the data, like the principal component analysis (PCA) [231] approach.

This approach of using machine learning to assist decision-making is especially common in combinatorial optimization problems. For example, Parmentier et al. [93] use machine learning on mixed-integer linear programming to estimate whether applying a Dantzig-Wolf decomposition will be effective, and Zarpellon et al. [21] apply machine learning on mixed-integer quadratic programming to decide if linearizing the problem will solve faster.

3.2 Machine learning for end-to-end decision-making

As another kind of machine learning method, reinforcement learning (RL) is the specialized paradigm solving sequential decision-making problems, i.e., determining what to do (action) according to the changing outside situation (state) in each time step to maximize a specific target (reward). Therefore, reinforcement learning is widely applied for the end-to-end decision making, serving as the development and supplement of traditional decision methods in operations research and management science (OR & MS). The typical setting for reinforcement learning is the Markov Decision Processes (MDPs) defined as $\langle \mathcal{S}, \mathcal{A}, P, R, \gamma \rangle$, where $s_t \in \mathcal{S}$ denotes the state space, $a_t \in \mathcal{A}$ denotes the action space, $P : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$ denotes the state transition probability given the current state and action, and $r_t = R : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ denotes the one-step reward given the current state and action. In a sequential decision process with T steps, the long-term return at t_0 is calculated according to the discount factor $\gamma \in (0, 1)$ and the reward of each step as:

$$R_{t_0} = \sum_{t=t_0}^{T} \gamma^{t-t_0} r_t. \tag{6}$$

RL approaches can be roughly divided into three categories, value-based RL, policy-based RL, and their intersection called actor-critic. Specifically, value-based RL learns a value function that represents the expected return that the agent can achieve from a given state or action. For example, Q-Learning [222] is a basic value-based RL method learning the following Q function given a state-action pair:

$$Q(s, a) = \mathbb{E}[R_t | s = s_t, a = a_t] = \mathbb{E}_{s_{t+1}}[r_t + \gamma \mathbb{E}_{a_{t+1}}[Q(s_{t+1}, a_{t+1})]], \tag{7}$$

where the recursive form is Bellman Equation. Instead of the exact values of ground truth, the target values y in reinforcement learning is given greedily according to the one-step reward:

$$y = r_t + \gamma \max_{a'} Q'(s_{t+1}, a'), \tag{8}$$

telling the value of a certain action. Deep Q-network (DQN) [142] keeps the same mathematical essence as Q-learning but estimates the value function with a deep neural network, representing more complex environmental situations. Policy-based RL directly learns a policy mapping states to actions without explicitly estimating the value function:

$$\pi(a|s) = P(a|s, \theta), \tag{9}$$

where the parameters of the policy function θ are directly optimized towards higher return:

$$\Delta\theta = \alpha \nabla_\theta \log \pi_\theta \left(s_t, a_t\right) R_t, \tag{10}$$

Examples include policy gradient methods like REINFORCE [187] and Trust Region Policy Optimization (TRPO) [168]. Finally, actor-critic is an intersection between value-based and policy-based RL, where the agent learns a policy (the "actor") to select actions based on feedback from a value function (the "critic") that estimates the expected return of the actions. Examples of Actor-Critic methods include Deep Deterministic Policy Gradient (DDPG) [176] and proximal policy optimization (PPO) [169].

Tracing back decades of history, traditional optimization methods have provided plentiful solutions to various categories of problems, including linear optimization [18], convex optimization [23, 25], non-convex optimization [6, 45], discrete optimization [17, 48], etc., which provide advantages including high efficiency and sound mathematical theories. These methodologies require numerical representation of state $s_t \in \mathcal{S}$ and explicit mathematical expression of the action-reward function $r_t = R : \mathcal{S} \times \mathcal{A}$. However, decision problems in urban scenarios always incorporate non-numerical states, e.g., representing the road networks and land use with graphs in urban planning [259, 260], and implicit action-reward function, e.g., drawing pandemic intervention reward from action via differential equations in urban healthcare tasks [63, 65], limiting the application of traditional optimization methods. Thereby, RL develops from value-based [142, 222] to policy-based methods [187], and further actor-critic ones [169, 176] to compensate the gap, as mentioned above. Nevertheless, RL methods suffer shortcomings, including unstable training process [70] and lack of action explainability [70, 154]. This enlightens further investigations combining RL and traditional optimization methods [214], maximizing the advantages of both methods.

There also exist some applications of supervised learning in end-to-end decision-making. One of the most widely known works is the pointer networks [132, 202], which takes in a sequential input and gives a ranking of the input elements. The pointer networks are suitable for solving the traveling salesman problem (TSP), where the networks are trained via supervised learning, using the optimal solutions of given instances as the ground truths.

4 CHALLENGES, ADVANTAGES, AND EXISTING METHODS

The use of machine learning for intelligent urban decision making poses several key challenges in actual research and deployment as follows:

- **Intrinsic Complexity.** Urban decision-making involves navigating a complicated, multi-scale, and interconnected system. Specifically, cities are complex systems with numerous static elements like geographical blocks, road networks, buildings, *etc.*, and dynamic processes, such as human movements, traffic flows, and disease transmissions. Different elements in cities tend to have distinct scales, with rich and usually unknown connections and dependencies between them. For example, the traffic flow is closely related to the distribution of land functionalities in different areas, and the spread of diseases is also strongly correlated with the movements of urban residents.

- **Urban Heterogeneity.** Urban decision usually exhibits diverse problem setups, such as different forms of road networks and distributions of urban functionality in urban planning, as well as different order distributions and distinct modes of human mobility in urban transportation and healthcare. Such strong heterogeneity makes it challenging for a model to generalize across cities and scenarios with quite different attributes, particularly under data-scarce conditions where we do not have access to data of all cities to train the model.
- **Computational Cost.** Urban decision often optimizes multiple objectives in an enormous space with various constraints, making it very difficult to find optimal solutions. For example, facility location in urban planning and vehicle routing in urban transportation are both NP-hard problems (non-deterministic polynomial). Meanwhile, fleet management and order dispatching require solutions with reliable performance within a strict time constraint to deliver prompt services, while urban planning consider various targets including efficiency and cost.

We now elaborate on how machine learning methods effectively address the above challenges in urban planning, transportation, and healthcare.

4.1 Machine learning for decision in urban planning

4.1.1 Road network planning. The problem of road network planning can be categorized into three sub-tasks according to the planning contents: road design, bike lane design, and public transportation design. We summarize existing machine learning methods for the above three sub-tasks in Table 2.

Urban road design problem. As introduced previously in Section 2.1, urban road design a very complex bi-level optimization problem, involving both policy-makers in the upper level and traffic participants in the lower level. Actions taken in one level are generated considering the conditions of the other level, and in the meantime also influence the behaviors in the other level. For instance, policy-makers design road networks according to the traffic patterns of the participants, while the designed roads determine how participants take routes in the city. Similarly, traffic participants take travel choices considering the road network, and their mobility provides feedback to policy-makers. Therefore, it is necessary to consider the upper and lower levels of the problem comprehensively before solving this problem. One traditional research idea is to unify the decision variables of the two levels through approximation methods, transforming it into a single-level optimization problem, and then use mixed integer programming or meta-heuristic methods to solve it. However, the solution accuracy tend to be sub-optimal due to over-simplification. Therefore, another idea is to optimize the bi-level problems directly using machine learning methods. For example, the deep Bayesian Optimization method is introduced to the solution [43, 44]. Specifically, due to the geometric properties, road networks are naturally suitable for modelling using graph structures, so the road network design problem can be transformed into a graph optimization problem. Following this idea, they use the Frank-Wolfe algorithm [58] to optimize the problem's lower level and propose a deep Bayesian graph optimization algorithm to optimize the upper level. In addition, some researchers have also introduced machine learning methods such as the Monte Carlo method [12] and multiple linear regression [252] into traditional models to enhance their ability to solve the bi-level problem. Fang *et al.* [55] adopted generative adversarial network (GAN) which generates a road network with the generator and use the discriminator to evaluate the generated results, achieving state-of-the-art compared with traditional optimization algorithms.

Bike lanes planning problem. For this problem, most traditional methods rely heavily on experience, which analyzes the necessity of building bicycles through public surveys or geographical condistions, making it challenging to consider

Problems	Paper	Scenarios	Type	Methods	Year
Road design	[43]	Designing urban road network	end-to-end	Bayesian Optimization	2019
	[44]	Designing urban road network	end-to-end	Graph neural network, etc.	2020
	[12]	Designing urban road network	indirect	Monte Carlo method, PSO	2021
	[55]	Designing urban road network	end-to-end	GAN, etc.	2022
	[260]	Designing urban road network	end-to-end	Graph neural network, RL, etc.	2023
	[252]	Adjusting urban road network	indirect	multivariate linear regression	2023
Bike lane design	[11]	Expanding bike network	indirect	Hierarchical Clustering, etc.	2017
	[3]	Expanding bike network	indirect	Density-based clustering, etc.	2018
	[68]	Expanding bike network	indirect	Hierarchical Clustering, etc.	2019
	[148]	Expanding bike network	indirect	Louvain Algorithm, etc.	2020
	[31]	Expanding bike network	indirect	DB-SCAN, etc.	2022
Public transportation design	[157]	Designing bus network	indirect	Density-based clustering, etc	2018
	[129]	Designing bus network	indirect	K-medoids, etc	2019
	[228]	Expanding city metro network	end-to-end	Actor-critic (RL)	2020
	[230]	Designing bus network	indirect	Monte-Carlo search tree	2020
	[213]	Designing bus network	end-to-end	GAN, Metric learning	2022
	[183]	Designing city metro network	end-to-end	Graph neural network, RL	2024

Table 2. A summary of machine learning methods used for road network planning

the constraints such as budget constraints, construction convenience and cycle lane utilization in the planning method. Therefore, data-driven approaches for bicycle lane planning were proposed, mainly using the indirect decision-making approach [3, 31]. For example, [11, 68] proposed a greedy network expansion algorithm guided by a scoring function related to user coverage and travel length for bicycle lane planning based on bicycle sharing data. This method obtains the initial point of network expansion employing spatial clustering, obtains the candidate set of lanes based on the greedy principle, and obtains the final planning result through continuous algorithm iteration. Moreover, [148] improved the selection method of targeting lanes. After obtaining the initial point through the hierarchical clustering method, it uses the penetration theory to select the target lane, further improving the solution's accuracy.

Public transport lanes planning problem. The public transportation planning involves the location selection of bus and subway stations. For subway station, Wei *et al.* [228] used RL to expand the subway network in a grid manner, which iteratively selects station locations (grid cells) according to the current planning results. Su *et al.* [183] further leveraged graphs to provide a more accurate geo-spatial representation of urban regions, and proposed an RL approach based on GNN to select new metro stations to maximize the served passenger flow. It is worth noting that subway planning involves various constraints which are usually characterized by action masks in RL based approaches. As for bus stations, there are mainly two categories of methods. One is similar to the subway network expansion mentioned above, using the end-to-end scheme [213]; the other is similar to bike lane planning, introducing clustering methods in traditional planning models [129, 157, 230] and using the indirect scheme.

4.1.2 *Land use planning.*

As introduced previously, land use planning consists of two sub-problems, urban function allocation and facility location. The two sub-problems focus on different scales, where function allocation plans a more extensive region with multiple blocks, and facility location usually selects point-level locations for urban facilities. Existing machine learning methods for land use planning are illustrated in Table 3.

Function allocation problem. With semi-structured or unstructured problems involved, traditional methods are brutal to solve land use planning effectively, where most existing research uses simulation technology to assist planning. For example, Li *et al.* [101] proposed an agent-based embedded learning model for residential land growth simulation, which integrates the learning model, decision-making model, land-use conversion model, and urban land-use overall planning

Problems	Paper	Scenarios	Type	Methods	Year
Function allocation	[101]	Allocate residential area	indirect	ABM-learning, etc.	2020
	[117]	Allocate residential area	indirect	CA-Markov, etc.	2020
	[208]	Allocate functional area	end-to-end	Adversarial learning	2020
	[112]	Allocate rural area	indirect	BP-ANN, etc.	2021
	[207]	Allocate functional area	end-to-end	Adversarial learning	2023
	[259]	Allocate functional area	end-to-end	Reinforcement learning	2023
	[209]	Allocate functional area and POI	end-to-end	GAN	2023
	[156]	Building layout generation	end-to-end	Diffusion	2024
Facility location	[217]	Planning fire station	indirect	K-Means, etc.	2018
	[118]	Planning charging station	indirect	Particle swarm optimization	2019
	[243]	Planning warehouse location	indirect	Weighted K-Means, etc.	2019
	[254]	Planning Electric fence	indirect	DB-SCAN Algorithm, etc.	2019
	[200]	Planning dry port	indirect	Apriori algorithm, etc.	2020
	[237]	Planning collection and delivery points	indirect	Gradient boosting tree, Dynamic clustering, etc.	2021
	[37]	Planning bicycle stations	indirect	Gated Graph Neural Network	2021
	[77]	Planning charging station	indirect	K-medoids, etc.	2022
	[203]	Planning charging station	end-to-end	DQN (RL)	2022
	[215]	Commercial site selection	end-to-end	One-shot non-autoregressive neural networks	2023
	[182]	Planning public infrastructure	end-to-end	Reinforcement learning graph neural networks	2024

Table 3. A summary of machine learning methods used for land use planning

constraints. In addition to the agent-based learning-embedded model, Liu *et al.* [117] introduced the CA-Markov model to study the impact of policy changes on urban residential land. Regarding multi-functional area planning, Liao *et al.* [112] proposed BP-ANN based on historical data to predict the proportion of land allocation in various functional areas to improve the final planning effect of the functional regions. Recent advances in generative AI and reinforcement learning has brought new perspectives in function allocation, where end-to-end approaches [156, 207–209, 259] were utilized to directly plan the urban functional area in a shorter time, including land use [259] and building layout [156].

Facility location problem. Charging station planning is an important problem in urban facility location, which aims to minimize users' waiting time or maximize the operators' benefits with the laid charging piles . Traditional planning methods mainly follow three schemes [140]. The first is the node-based method, which directly selects nodes from the road network as charging stations. It is a typical NP-hard problem, and heuristic methods are commonly used to solve it. The second is path-based method that builds charging stations on the path with the enormous traffic flow to meet user needs. The third method is based on user behaviour, which considers the starting point, destination, travel distance, vehicle path and dwell time, and selects the best location to set up the charging station. However, traditional methods are limited by human experience, most of which are approximate solutions, challenging to meet the requirements in various scenarios. Therefore, more and more researchers have begun to use machine learning to solve the problem of charging facility planning [77, 118, 203]. For example, Wang *et al.* [215] proposed a differentiable optimal transport (OT) layer to establish a general machine learning solver with one-shot neural networks, which effectively accelerate the solution generation of facility location problems. Leonie von Wahl *et al.* [203] used the weighted sum of the income and the cost as the objective function to solve it using reinforcement learning. The income function represents the total amount of user demand that can be satisfied by this laying method, and the cost function is the travel time (i.e., the time to go to the charging station) and the time consumed by the user to meet the charging demand. The optimization goal is to minimize the charging fee of the user while maximizing the satisfaction of the user's needs. At each step,

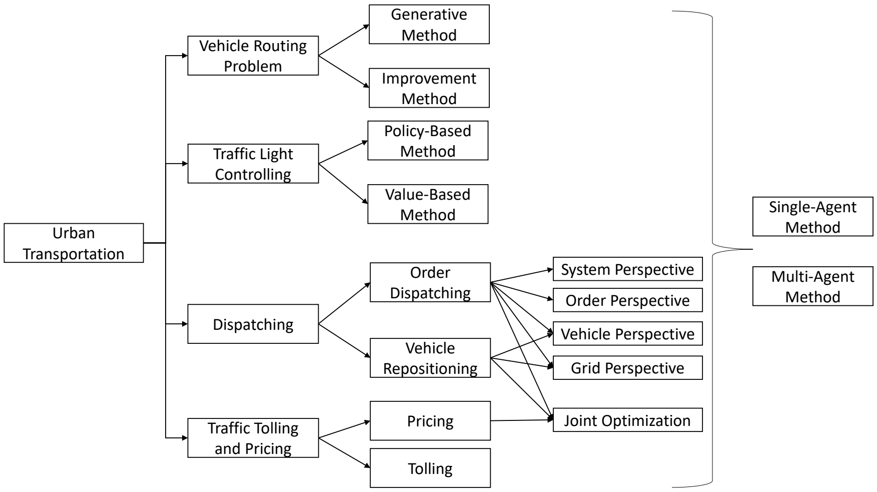


Fig. 8. The relationship between each sub-task and the methods used in urban transportation decision.

the reinforcement learning model will optimize the layout of charging stations from three aspects: adding charging stations, increasing the capacity of existing charging stations, and deleting existing charging stations until the number of charging stations reaches the upper limit or exceeds the total construction cost.

Besides charging station planning, machine learning for facility planning in recent years also considers logistics facilities, public safety facilities, and bike sharing stations [37, 182, 254]. For example, in logistics planning problems, unsupervised clustering algorithms are usually combined with mixed integer programming methods to improve the quality of the solution [200, 243]. In planning public safety facilities, the location of candidate facilities will be initially obtained through clustering algorithms based on population density, historical safety accident data, *etc.*, and then further optimized based on other empirical methods [217].

In summary, real-world urban planning applications often exhibit an enormous solution space, causing traditional searching approaches struggle in generating reliable solutions in an acceptable timeframe. In contrast, machine learning approaches significantly accelerate decision-making in urban planning by accurately predicting the value of different decisions, which facilitates efficient exploration of the solution space. Specifically, they substantially narrows down the search space, allowing high-quality solutions to be explored more frequently. This acceleration has been verified in real-world urban planning applications, such as urban functionality layout for communities. These tasks, which previously relied heavily on professional human designers, have been accelerated by over 3000 times using a deep reinforcement learning (DRL) approach that effectively addresses computational cost challenges [259], demonstrating the practical applicability of machine learning methods in urban planning.

4.2 Machine learning for decision in urban transportation

In this section, we introduce machine learning methods for solving Vehicle Routing Problem, traffic light controlling problem, dispatching, and traffic tolling and pricing, as shown in Fig. 8.

4.2.1 Vehicle Routing Problem. Considering the diverse and complex vehicle routing optimization scenarios in urban traffic, in addition to the standard Vehicle Routing Problem (VRP), existing research investigates VRP variants in

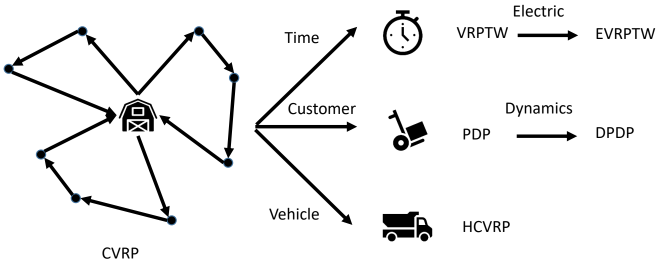
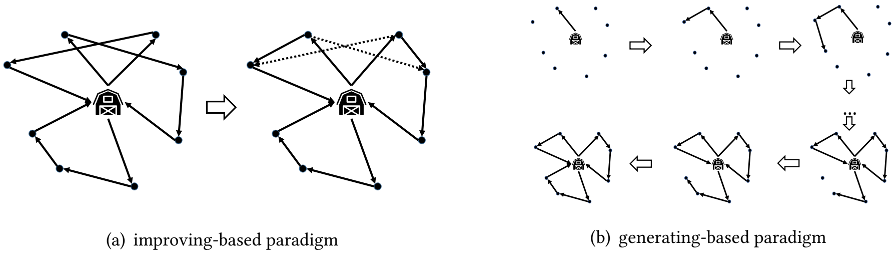


Fig. 9. Main variations of the Vehicle Routing Problem.


(a) improving-based paradigm

(b) generating-based paradigm

Fig. 10. Two VRP solving paradigms based on deep reinforcement learning.

practical urban traffic, where the main variations of VRP are shown in Fig. 9. For example, VRP with time windows (VRPTW) attach each routing service with a corresponding time window, VRP with pickup and deliveries (VRPPD) or pickup and delivery problem (PDP) features a set of transportation requests where customers or goods need to be moved from certain pickup locations to other delivery locations. We refer readers to two awesome surveys [28, 267] for more detailed introduction of a richer catalog of VRP variants featuring more diverse and complicated additional constraints and objectives. Existing machine learning methods to solve VRP problems are summarized in Table 4 regarding their learning paradigm, problem, solving paradigm, *etc.*

Machine learning methods for VRP and its extensions follow two paradigms: generating-based and improving-based, as illustrated in Fig. 10. Generation based methods generate the partial solution step by step until a complete solution is obtained. At each step of generating a partial solution, the vehicle selects the next accessible node from the unvisited nodes according to the constraints of the problem. Improving-based methods generate a complete solution initially and then destroys and reconstructs the complete solution using operators to improve the performance of the solution.

Following Pointer Network (Ptr-Net) [202] that solves TSP with the generation-based paradigm, several methods were proposed that utilized supervised learning to solve routing problems. Joshi *et al.* [85] utilized Graph Convolution Network (GCN) to output a heat map corresponding to the TSP tour. Sultana *et al.* [184] proposed a convolutional neural network combined with a Long Short-Term Memory for solving non-Euclidean TSP in a supervised manner. As deep reinforcement learning flourishes, researchers are applying it to solve VRP and its variants. Kool *et al.* [90] proposed a encoder-decoder framework based on the attention mechanism for combinatorial optimization problems such as CVRP, where the encoder and decoder calculate the embeddings of CVRP instances and progressively generate the solutions of CVRP respectively, trained via the REINFORCE algorithm. Due to its superior solution performance and inference time, Kool *et al.*'s model [90] has been improved for addressing VRPs in various ways [95, 236], and adapted to different problems, such as Vehicle Routing Problem with Time Windows (VRPTW) [250], Transit Network

Problem	Category	Paper	Year	Method	Solving Paradigm
TSP	DL	[202]	2015	RNN, Attention	Generation
	DL	[85]	2019	GCN	Generation
	RL	[95]	2021	Attention, REINFORCE	Generation
	RL	[84]	2023	Attention, REINFORCE	Generation
	RL	[152]	2023	REINFORCE, Hierarchical RL	Improvement
CVRP	DL	[71]	2021	VAE	Generation
	RL	[144]	2018	RNN, REINFORCE, A3C	Generation
	RL	[90]	2019	Attention, REINFORCE	Generation
	RL,DL	[51]	2020	Attention, GCN, REINFORCE	Generation
	RL	[236]	2021	Attention, REINFORCE	Generation
	RL	[20]	2022	Attention, REINFORCE, Knowledge Distilling	Generation
	RL	[263]	2023	REINFORCE, Meta Learning	Generation
	RL	[80]	2024	REINFORCE, Deep Ensemble	Generation
	RL	[127]	2020	MLP, REINFORCE	Improvement
	RL	[136]	2021	Attention, PPO	Improvement
	RL	[235]	2021	Attention, A2C	Improvement
	RL	[268]	2022	Attention, REINFORCE	Improvement
	RL	[81]	2023	REINFORCE, Neural Heiristic	Improvement
	RL	[133]	2024	REINFORCE, Neural Heiristic	Improvement
PDP	RL	[104]	2021	Attention, REINFORCE	Generation
	RL	[269]	2022	Attention, Cooperative A2C	Generation
	RL	[135]	2022	Attention, PPO	Improvement
Non-Euclidean TSP	DL	[184]	2022	GCN, LSTM	Generation
OVRP	RL	[78]	2019	Struct2Vec, Pre-Net, REINFORCE	Generation
VRPTW	RL	[250]	2020	Attention, REINFORCE	Generation
TNDFSP	RL	[46]	2020	Attention, REINFORCE	Generation
HCVRP	RL	[103]	2021	Attention, REINFORCE	Generation
EVRPTW	RL	[115]	2022	Struct2Vec, Attention, REINFORCE	Generation
DPDP	RL	[134]	2021	GIN, DQN, REINFORCE	Improvement

Table 4. A summary of machine learning methods used for Vehicle Routing Problem

Design and Frequency Setting Problem (TNDFSP) [46], heterogeneous CVRP (HCVRP) [103], pickup and delivery problem (PDP) [104, 269]. Moreover, Jin *et al.* [84] utilize multi-pointer Transformer with reversible residual networks to manage memory consumption efficiently which demonstrates effective scalability to large-scale TSP instances while maintaining competitive results on smaller ones. These methods provide a speed advantage without compromising the solution's quality compared to traditional methods.

In addition to the above generating-based RL methods, improving-based RL models were proposed. Lu *et al.* [127] proposed a Learn-to-Iterate (L2I) framework to solve CVRP. Besides enhancing the current solution with RL, this framework includes perturbation that prevents it from being trapped in a local optimal solution. Wu *et al.* [235] improved the existing solutions by selecting pairwise operators (such as 2-opt) with the attention mechanism, based on which Ma *et al.* [136] incorporated positional information of nodes to enhance the solution characterization using the attention mechanism. Pan *et al.* [152] employ hierarchical reinforcement learning to tackle large-scale TSPs, with an upper-level policy selecting a subset of nodes and a lower-level policy generating a tour, thereby eliminating the need for time-consuming search procedures.

4.2.2 Traffic Light Controlling Problem. Traditionally, researchers approach this problem by converting it into an optimization problem with assumptions. However, these assumptions might deviate from real-world scenarios, rendering the control scheme impractical, which leads to an increasing reliance on learning-based methods. Particularly,

Learning Paradigm	Paper	Year	Method	Simulator
Policy-based RL	[30]	2017	DDPG	Aimsun
	[162]	2019	REINFORCE	SUMO
	[42]	2019	A2C	SUMO
	[234]	2020	DDPG	SUMO
	[164]	2024	Hierarchical RL, MARL	SUMO
Value-based RL	[227]	2018	DQN	SUMO
	[258]	2019	DQN	SUMO
	[223]	2019	DQN	CityFlow
	[216]	2021	DQN	SUMO
	[247]	2020	DQN	CityFlow
	[34]	2020	DQN	CityFlow
	[224]	2019	Attention, VFA	CityFlow
	[220]	2020	Attention, DQN, RNN	SUMO
	[59]	2016	Q-learning	SUMO
	[199]	2016	DQN	SUMO
	[126]	2022	GAT, Meta Learning, DQN	CityFlow
	[122]	2023	GCN, QMIX	CitiFlow
	[128]	2024	GAT, DQN	SUMO

Table 5. A summary of machine learning methods used for Traffic lighting Controlling Problem

reinforcement learning can monitor traffic conditions and generate corresponding control strategies based on the feedback received from the environment. Additionally, it can effectively process high-dimensional data and produce control strategies of higher quality, avoiding the unrealistic assumptions of traditional approaches [226].

Table 5 illustrate existing RL methods for traffic light controlling problems. Several policy-based RL methods were proposed which directly optimize policy parameters. For example, Rizzo *et al.* [162] proposed a policy gradient method with a time baseline that effectively reduces policy gradient variance. Moreover, Chu *et al.* [42] proposed the Advantage Actor Critic algorithm (A2C)-based, scalable, and decentralized MARL algorithm for large-scale traffic light control. Ruan *et al.* [164] proposed to separate the collaborator selection as a second policy to be learned, concurrently being updated with the original signal-controlling policy. Furthermore, several studies have utilized value-based RL which estimates the value function to generate an policy implicitly. For instance, Wang *et al.* [216] developed a collaborative double-Q learning (Co-DQL) method for large-scale traffic light control, utilizing the mean-field approximation for better agent interactions. Lu *et al.* [128] enhanced decision-making by leveraging both experiential information within individual scenarios and generalizable information across different scenarios.

4.2.3 Dispatching. Dispatching optimization includes the problems of order dispatching, vehicle dispatching and the joint optimization of both. Most existing machine learning methods solving dispatching optimization utilize reinforcement learning, which can be categorized into four perspectives: system perspective, vehicle perspective, order perspective and grid perspective, of which the first two are mostly adopted. These methods can be also classified into single-agent setting and multi-agent setting. Table 6 summarize the existing methods.

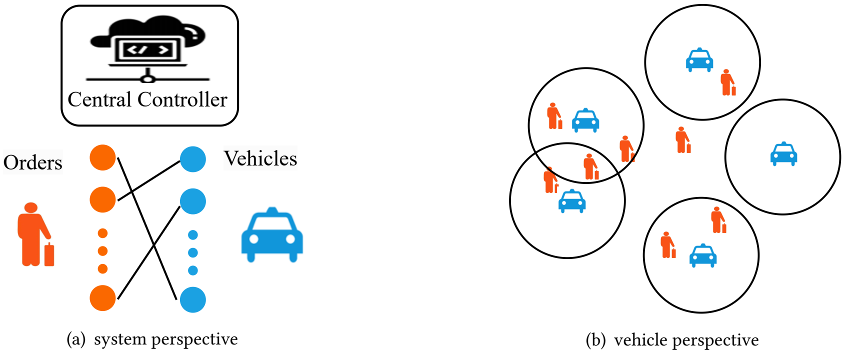
Order Dispatching. Traditional works address order dispatching problem through rule-based approaches in centralized or decentralized settings. In the centralized setting, Liao *et al.* [113] and Lee *et al.* [98] use the myopic algorithm to match vehicles with nearest orders. Zhang *et al.* [251] dispatch taxis to serve multiple orders based on combinatorial optimization within a short time window, which can improve global performance. These methods are difficult to be applied in the large-scale ride-hailing system due to the need to compute all available driver-order matches. In the

Problem	Paper	Scenarios	Method	Year	Agent Perspective
Order Dispatching	[108]	Express	DQN	2019	Multi-agent
	[109]	Express	DQN	2020	Multi-agent
	[221]	Ride-hailing	DQN	2018	Single-agent
	[238]	Ride-hailing	TD, Bipartite matching	2018	Single-agent
	[189]	Ride-hailing	DQN, Bipartite matching	2019	Single-agent
	[108]	Ride-hailing	Mean-field A2C	2019	Multi-agent
	[264]	Ride-hailing	DQN	2019	Multi-agent
	[173]	Ride-hailing	REINFORCE	2020	Single-agent
	[86]	Ride-hailing	QQN, A2C, PPO, ACER	2019	Multi-agent
	[155]	Ride-hailing	ACER	2021	Single-agent
	[100]	Ride-hailing	Bipartite matching	2019	-
	[174]	Ride-hailing	DQN, Bipartite matching	2021	Single-agent
	[193]	Ride-hailing	DQN, Bipartite matching	2021	Single-agent
	[219]	Ride-hailing	DQN, Bipartite matching, Federated learning	2022	Single-agent
	[62]	Ride-hailing	TD, Bipartite matching	2022	Single-agent
	[165]	Ride-hailing	DQN, Bipartite matching	2022	Single-agent
	[111]	Ride-hailing	DQN, Bipartite matching	2024	Single-agent
	[248]	Ride-hailing	CQL, offline RL, ensemble	2024	Single-agent
Vehicle Dispatching	[107]	Bike-sharing	DQN	2018	Single-agent
	[116]	Ride-hailing	DQN	2019	Multi-agent
	[175]	Ride-hailing	MFRL, Bayesian optimization	2020	Multi-agent
	[257]	Ride-hailing	DQN	2022	Single-agent
	[124]	Ride-hailing	DQN	2022	Single-agent
	[123]	Ride-hailing	DQN	2022	Single-agent
	[225]	Ride-hailing	DQN, Linear programming	2023	Single-agent
	[76]	Ride-hailing	Hierarchical RL, QMIX	2023	Multi-agent
Joint Optimization	[82]	Ride-hailing	DDPG, FeUdal Networks	2019	Multi-agent
	[60]	Ride-hailing	DQN	2020	Single-agent
	[190]	Ride-hailing	DQN, Bipartite matching	2021	Single-agent
	[110]	Ride-hailing	TD, DQN	2021	Multi-agent
	[186]	Ride-hailing	A2C	2022	Multi-agent
	[185]	Ride-hailing	A2C	2024	Multi-agent

Table 6. A summary of machine learning methods used for dispatching

decentralized setting, Seow *et al.* [171] propose a collaborative multi-agent taxi dispatching system, which concurrently matches multiple taxis with passengers in the same geographical area. However, this method requires multiple rounds of direct communication between vehicles, which limits it to a small area with small number of vehicles.

Reinforcement learning approaches are more popular in solving the order dispatching problem in recent years, which avoid complicated hand-crafted heuristics and features of rule-based approaches. Existing RL methods in ride-hailing can be categorized into two settings, single-agent setting and multi-agent setting. Single-agent methods optimize order dispatching from the perspective of the whole system, as shown in Fig. 11(a). A typical method is to combine RL with a bipartite graph and combinatorial optimization. For example, Xu *et al.* [238] learn the state value function in the table form from the historical real order dispatching data. With state value function and price of orders, they compute the advantage function as the weights of order-vehicle pairs in the bipartite graph. To improve the approximation and representation ability, Tang *et al.* [189] substitute the Cerebellar Value Network for tabular value function, which can capture the demand-supply dynamics from different geographical scales. Instead of online RL, Zhang *et al.* [248] proposed NondBREM, an offline RL method to learn policies from historical data to circumvent the high costs and safety concerns associated with online policy learning.



(a) system perspective

(b) vehicle perspective

Fig. 11. Two main perspectives for order dispatching.

On the other hand, multi-agent methods mainly optimize order dispatching from the perspective of each vehicle, as shown in Fig. 11(b). These methods usually model each vehicle as an agent and apply multi-agent reinforcement learning (MARL) to achieve coordination among vehicles. For instance, Li *et al.* [108] use mean-field RL method to simplify the interaction among vehicles by taking the average action as the complex interaction of neighboring vehicles. Extending from [108], Li *et al.* [109] propose a cooperative MARL algorithm by combining the markov decision processes of cooperation in one time slot and independent optimization in a long term.

There are also some methods in multi-agent setting from perspectives of grid and order. Jin *et al.* [82] model each grid as an agent, which means that each grid is a local controller for order dispatching task in its field. They divide the city map into hierarchical areas, where each grid is a unique worker and the grid with its neighbors is called a manager. They apply FeuDal Network to achieve cooperation among workers in the same manager.

Vehicle Dispatching. Reinforcement learning methods follow the vehicle perspective or grid perspective. [107, 116, 123, 124, 175, 257]. In the vehicle dispatching problem, the coordination among vehicles is important. Without coordination, if too many vehicles reposition to the areas where demand exceeds supply, these areas will turn to the opposite situation where supply exceeds demand. Lin *et al.* [116] consider vehicles in the same grid as homogeneous elements, so they design their method form grid perspective instead of vehicle perspective. They propose contextual DQN and contextual Actor Critic to realize coordination among vehicles. Specifically, they incorporate geographic context to avoid infeasible movements by removing the invalid areas such as river and mountain. Collaborative context is also incorporated to avoid conflict movements by eliminating actions from higher value areas to lower value areas. Huang *et al.* [76] proposed a multi-level controller framework where the leader controller sets goals for the follower controller to execute, and a MIX module is implemented to enhance algorithm stability by computing the total value of joint actions.

Joint Order Dispatching and Vehicle Dispatching. There are some methods [60, 82, 110, 186, 190] considering the two tasks jointly to achieve better balance between vehicle-order distribution and improve performance. The challenge of joint optimization comes from the heterogeneous action space of the two tasks. One is matching vehicles with orders while the other is repositioning vehicles to certain areas. To address this challenge, Jin *et al.* [82] treat candidate repositioning areas as fake orders with the same feature space of real orders but price zero. With homogeneous feature space of fake orders and real orders, they design action as ranking weight vector for homogeneous features. The ranking scores are used to select orders and repositioning areas. Tang *et al.* [190] overcome the challenge by designing shared value function for both tasks. For order dispatching, they use value function to compute advantage for each order-vehicle pair, which is served as the weight of each order-vehicle matching. Then they match orders with vehicles by maximizing the total weights of all order-vehicle matching. For vehicle dispatching, they treat value function as

Problem	Learning Paradigm	Paper	Year	Method
Dynamic Tolling	RL	[158]	2019	MARL, eGCN
	RL	[166]	2021	Q-learning
	RL	[83]	2021	A2C, DNN
	RL	[218]	2022	Attention, SAC
Pricing	RL	[233]	2016	Qlearning
	RL	[178]	2020	Qlearning
	RL	[61]	2020	DQN
	RL	[35]	2021	PPO
	RL	[75]	2022	SAC
	RL	[253]	2022	MARL
	RL	[99]	2023	SAC
	RL	[10]	2024	DQN
Pricing and dispatching	RL	[197]	2020	PPO
	RL	[119]	2022	MARL
	RL	[205]	2022	MARL
	RL	[36]	2019	bandit algorithm, TD learning

Table 7. A summary of machine learning methods used for traffic tolling and pricing

score for each repositioning area and take action by sampling from *softmax* of value functions. Sun *et al.* [186] address the heterogeneous action challenge by treating both order dispatching and vehicle dispatching as selecting destination for vehicles. Specifically, since price of order is highly correlated to distance between origin and destination, they omit the price feature and only consider destinations of different orders when taking actions.

4.2.4 Traffic Tolling and Pricing. Dynamic traffic tolling and pricing strategies based on the real time traffic situation and supply-demand information, can help to ease traffic congestion and balance the supply-demand distribution, which plays an important role in increasing traffic efficiency. Table 7 summarize existing methods.

Traffic Tolling. Existing static tolling is based on the analysis of the historical state of road congestion so as to establish the price scheme [240, 262], which is difficult to be applied in the complex and changeable scene. Thus, the study of dynamic tolling emerges which develops the tolling scheme based on real-time traffic conditions, as shown in Table 7. Wang *et al.* [218] proposed a reinforcement learning approach for dynamic traffic toll collection, which uses state-based attention mechanism to represent the congestion of each route and designs appropriate reward function to optimize the charging strategy. Jin *et al.* [83] considered the deadline of travelers to simulate the requirements of travelers, and automatically assigned the optimal toll value of each road based on deep reinforcement learning to meet the requirements of travelers.

Traffic pricing. Since the pricing strategy is closely related to order dispatching and vehicle dispatching problems, a good pricing strategy can help to balance the distribution between supply and demand, increase driver incomes and enhance travel satisfaction of passengers. Different from traditional methods that optimize pricing strategy with stochastic model, spatio-temporal analysis, equilibrium analysis, *etc* [130, 131, 150, 239], reinforcement learning methods dynamically adjust pricing strategies according to real-time traffic situations. On the one hand, the pricing strategy can be optimized separately, combined with ruled based order dispatching or vehicle repositioning solutions [35, 61, 75, 99, 178, 233, 253]. For example, Haliem *el al.* [61] design a distributed dynamic pricing method, where drivers are allowed to propose their price based on the trip. On the other hand, due to the close relation between pricing and dispatching, some works optimize pricing and dispatching jointly [36, 119, 197, 205]. For instance, to combine the model of pricing and order dispatching, Chen *et al.* [36] propose to integrate contextual bandit with TD learning to handle the two tasks respectively, trained in a mutually bootstrapping manner.

Action Space	Paper	Year	Pollution	Algorithm	Action	Optimization Target
Discrete	[72]	2020	Water	DQN	Open or close valves	Minimize the mass of contaminant
	[177]	2020	Water	Q-Learning	Share watersource or not	Maximize the groundwater quality
	[192]	2024	Water	DQN	Turn on or off the pumps	Minimize the flood inflow
	[40]	2018	Air	Q-Learning	Open or close windows	Minimize energy consumption and maximize air quality
	[69]	2019	Air	DQN	Set the inverter frequency	Minimize energy consumption and maximize air quality
	[106]	2021	Air	Q-Learning	Change the weight coefficient	Maximize the prediction accuracy
	[188]	2022	Air	Q-Learning	Change the weight coefficient	Maximize the prediction accuracy
	[172]	2023	Air	DQN	Air purifier perform mode	Minimize energy consumption and maximize air quality
	[244]	2023	Air	Sarsa	Change the weight coefficient	Maximize the prediction accuracy
	[167]	2022	Garbage	DQN	Adjust the air flaps	Minimize the emission from waste incineration
	[4]	2022	Garbage	DQN	Choose the category	Maximize the classification accuracy
Continuous	[22]	2022	Water	DDPG	Adjust the valves	Minimize the pollutant
	[38]	2021	Water	DDPG	Set dissolved oxygen	Minimize the negative impacts of wastewater treatment process
	[241]	2021	Water	Actor-Critic	Set internal recycle flow	Minimize the negative impacts of wastewater treatment process
	[79]	2023	Water	SAC	Set water quality parameters	Maximize the prediction accuracy

Table 8. A summary of methods used for environmental pollution control

In summary, among the previously introduced three major challenges, the intrinsic complexity is particularly prominent in real-world urban transportation decision due to its high-dimensional nature and the dynamically changing problem conditions, leading to limited applicability of solely data-driven methods. Nevertheless, most existing works were evaluated on simplified scenarios that deviate from the reality, preventing the use of existing machine learning-based approaches in real-world urban transportation decision applications. It is worthwhile to notice that there exists specific domain knowledge in traditional optimization solvers, which can effectively benefits decision-making in urban transportation. Specifically, advanced machine learning methods such as RL can be utilized in combination with optimization solvers, resulting in a hybrid solution generation that mixes data-driven and knowledge-driven approaches. The benefits of incorporating domain knowledge into machine learning approaches have been demonstrated in real-world urban transportation application, such as large-scale VRP containing thousands of customers. Notably, RL approaches combined with optimization solvers can effectively decompose the problem into multiple regional pieces and efficiently generate solutions, which have been deploted to an online logistic platform in Guangdong, China, showcasing the applicability in practice [268].

4.3 Machine learning for decision in urban healthcare

4.3.1 Environmental pollution control. Most existing works solving decision-making tasks in environmental pollution control use reinforcement learning-based methods. Since the emission and diffusion process of the pollutant is complex in large-scale urban contexts, it is difficult for human experts to propose optimal control strategies. Therefore, utilizing the capability of RL in solving complex sequential decision problems is a natural choice. Here, we mainly summarize control methods on three major kinds of environmental pollution, i.e., water pollution [22, 38, 72, 79, 177, 192, 241], air pollution [40, 69, 106, 172, 188, 244], and garbage pollution [4, 167], as illustrated in Table 8.

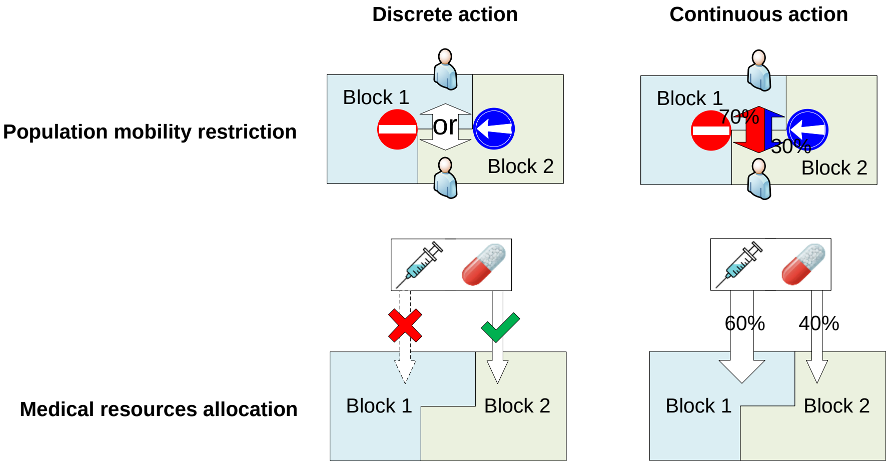


Fig. 12. An illustration of discrete and continuous action in the sub-problems of pandemic spreading intervention.

According to the properties of the output action, existing methods can be roughly divided into two categories. The first is discrete action space methods, selecting one action from a finite and discrete set of actions. Wang *et al.* [72] design a DQN-based method to determine whether to open or close the valves in the water system, minimizing the mass of contaminant in the drinking water. Malkawi *et al.* [40] employ the Q-Learning algorithm to decide whether to open or close the windows, maximizing indoor air quality and saving energy. Besides, Spinler *et al.* [167] hire the DQN method to control garbage incineration, reducing the air pollutant emission. And the second is continuous action space methods, in which the output is to determine the action as a continuous value. Similar to [72], Goodall *et al.* [22] design a method for controlling the valves in the water system to minimize the mass of contaminant, but they hire the DDPG algorithm to determine the percentage of valves' opening, which is a continuous value. Also, Wang *et al.* [38] and Si *et al.* [241] respectively adopt the DDPG and the Actor-Critic algorithm to control the wastewater treatment process, minimizing its negative impact on the environment.

4.3.2 Pandemic spreading intervention. We investigate two sub-problems of population mobility restriction and medical resources allocation. Existing methods can be similarly divided into discrete and continuous action space methods, whose differences are illustrated in Fig. 12.

Population mobility restriction. It is mainly a sequential decision problem, *i.e.* adjusting the restriction strength among urban blocks in real-time according to the pandemic spreading situation, minimizing the damage caused by the pandemic. Since the pandemic spreading situation changes fast, especially for cities with large populations and strong population mobility, it is almost impossible for human experts to make real-time decisions on a fine-grained level. Thus there leaves ample space for machine learning methods, and the existing works mostly use reinforcement learning-based methods since it is the reliable method for solving sequential decision problems.

Both discrete action space methods [1, 27, 87, 94, 147, 151, 204] and continuous action space approaches [32, 50, 114] have been proposed. Table 9 illustrates the commonality and differences among existing methods for population mobility restriction in terms of the algorithm, action, and optimization target. In terms of discrete action, Hitmi *et al.* [151] employ the Q-Learning algorithm to determine the restriction strength from 20 optional levels, reducing infections and minimizing the cost of restrictions. Büyüktahtakın *et al.* [27] use the DQN algorithm to choose the intervention and vaccination policy from 9 options with the aim of reducing infections and maintaining the social economy. Hamid

Action Space	Paper	Year	Algorithm	Action	Optimization Target
Discrete	[94]	2021	D3QN	Lockdown and travel restriction (3*3 levels)	Reduce infections, and accelerate the recovery
	[147]	2020	DDQN	Movement restriction (3 levels)	Reduce infections, and maintain the economy
	[204]	2021	model-based RL	Official intervention policy (3 levels)	Reduce infections, and maintain the economy
	[151]	2021	Q-Learning	Restriction strength (20 levels)	Reduce infections, and minimize the cost
	[27]	2023	DQN	Interventions and vaccination policy (9 levels)	Reduce infections, and maintain the economy
	[87]	2020	DQN	Whether to keep each node open or lock it down (binary)	Reduce infections, and minimize the cost
	[1]	2022	DQN	How the agent move to aviod crowd situation (5 ways)	Reduce infections
Continuous	[32]	2022	DDPG, PPO, TD3	The strength of intervention polices	Reduce infections, and maintain the economy
	[114]	2020	PPO	Probablity of keeping schools open or locking them down	Reduce infections
	[50]	2023	PPO	Mode of intervention polices	Reduce infections, and maintain the economy

Table 9. A summary of RL methods used for population mobility restriction

Action Space	Paper	Year	Resource	Algorithm	Action	Optimization Target
Discrete	[15]	2021	Vaccines	Actor-Critic	To whom the vaccines should be allocated	Reduce infections, and minimize the cost of resources
	[266]	2022	Vaccines	Actor-Critic	Level of allocation strategies (5 levels)	Reduce infections, and maintain the economy
	[14]	2021	Ventilators	Q-Learning	Policies of resources transfer (3 kinds)	Reduce the shortage of resources
	[206]	2020	Protective resources	Q-Learning	To whom the resources should be allocated	Reduce infections, and minimize the cost of resources
	[195]	2022	Vaccines	DDQN	To whom the vaccines should be allocated	Reduce infections
Continuous	[64]	2021	Surgical masks, hospital beds	DDPG	Percentage of resources allocated to each block	Reduce infections
	[63]	2022	Vaccines	PPO	Percentage of resources allocated to each block	Reduce infections
	[65]	2022	Surgical masks, hospital beds	DDPG	Percentage of resources allocated to each block	Reduce infections
	[232]	2021	Vaccines	PG	Probablity of individualistic or collectivist strategy	Reduce infections
	[9]	2020	Vaccines	Actor-Critic, DQN	Percentage of resources allocated to each block	Reduce infections, and minimize the cost of resources

Table 10. A summary of RL methods used for medical resources allocation

et al. [147] apply the DDQN algorithm to select an official intervention policy from 3 candidate policies to reduce infections and maintain the social economy. Hui *et al.* [94] design a method based on the D3QN algorithm to determine the strength of the intra-city lockdown and inter-city traveling restriction policies, each from 3 options. And Song *et al.* [204] use model-based RL to decide official restriction policy, selected from 3 optional levels. With respect to continuous action space methods, Nowé *et al.* [114] propose using the PPO algorithm to determine the probability of keeping schools open or locking them down, a continuous value varying from 0 to 1, and they manage to reduce the infections. Mousannif *et al.* [32] hire the DDPG, PPO, and TD3 algorithms to calculate the strength of mobility restriction policies, balancing the pandemic intervention and the social economy.

Medical resources allocation. Similar to the population mobility restriction decision, medical resource allocation is also a sequential decision problem. The goal is to allocate limited resources to each urban block at each time step according to the pandemic spreading situation and thus minimizing the damage caused by the pandemic. Therefore, reinforcement learning is commonly used in medical resource allocation problems. Existing methods respectively focus on various kinds of medical resources, such as vaccines [9, 15, 63, 195, 232, 266], personal protection equipment [64, 65, 206], ventilators [14], and hospital beds [64, 65], as illustrated in Table 10.

In the discrete category, Jones *et al.* [14] design a Q-Learning algorithm to determine the sharing policy of ventilators among cities, reducing the shortage of such kind of resources. Hanzo *et al.* [206], Falou *et al.* [195], and Jahanshahi *et al.* [15], respectively use the Q-Learning, DDQN, and actor-critic algorithm to determine to whom the resources should be allocated with priority, with the same aim of reducing infections and minimizing the cost of resources. In the continuous category, Li *et al.* [64, 65] propose using the DDPG algorithm to decide the percentage of surgical masks and hospital beds allocated to each urban block, aiming to reduce infections. Besides, Sethi *et al.* [9] and Li *et al.* [63] respectively employ actor-critic and PPO algorithm to determine the percentage of vaccines allocated to each block.

In summary, the task of urban healthcare decisions often necessitates multiple data sources, which are collected through various approaches and contain plentiful information about the urban decision environment. Such data sources exhibit great heterogeneity, presented in various modalities, including tensors, tables, graphs, texts, etc, bringing about the challenge of high urban information heterogeneity. Facing highly heterogeneous data, human experts and conventional decision methods lack the capability to thoroughly extract the hidden information from the data, causing them to struggle to comprehensively utilize the plentiful information to make optimized decisions. In contrast, advanced machine learning approaches can process multi-modal heterogeneous data with various network structures, such as MLP, CNN, GNN, and Transformers, extracting hidden information from them. Meanwhile, RL algorithms can automatically consider all available information, and output comprehensively optimized decisions. For example, one existing machine learning approach that utilizes self-supervised representation learning with MLP and GNN to combine heterogeneous data of population mobility, pandemic spreading situation, and demographic features, and then uses RL to solve the medical resources allocation problem, has shown supreme performance tested on various metropolis with millions of population [63].

5 FUTURE DIRECTIONS AND OPEN PROBLEMS

5.1 Advanced Machine Learning Techniques

The previously mentioned three major challenges of urban decision making—intrinsic complexity, urban heterogeneity, computational cost—can be better addressed by utilizing more advanced machine learning techniques. First, rich knowledge from diverse sources such as urban related texts and spatio-temporal data provides valuable insights into the complexity of cities, making the construction of urban knowledge graphs beneficial for data-driven decision approaches. Second, the emergent ability of foundation models indicate a promising direction for tackling urban heterogeneity by pretraining urban decision models with large-scale cross-city data. Third, while reinforcement learning approaches offer superior inference speed for generating urban decision solutions, the training process is often expensive, where advanced techniques can be leveraged to build sample efficient RL models and reduce training costs. We now elaborate on these three aspects with specific potential machine learning techniques for addressing the three challenges.

Urban Knowledge Graph. Cities are complex systems containing various types of knowledge, such as the functional connections between different areas. However, existing methods for urban decision making are mainly based on data and ignore the rich urban knowledge, which can lead to suboptimal results. A knowledge graph is a structured representation of knowledge that describes the relations between entities through edge connections between nodes, and has a wide range of applications such as question answering (QA). Urban knowledge graph (UrbanKG) [120, 210] combines knowledge graph and urban computing, which aggregates multi-source data such as area and POI, and can support different tasks. For example, UrbanKG is used to select locations for brands to open stores [121]. Considering the intrinsic complexity of cities, UrbanKG is a promising direction to support knowledge-driven urban decision-making.

Foundation Models. Currently, most of existing approaches train separate models for different cities and decision tasks, which can be time-consuming and resource-intensive. More importantly, different cities and tasks share commonalities, and what is learned in one city or task can be useful in other scenarios. Foundation models for decision making are recently proposed [161, 229], capable of addressing different decision tasks, including video games, continuous control, and TSP problems. In urban scenarios, universal prediction models [246] trained with large-scale cross-city data has been proposed and we believe urban decision foundation models can have far-reaching implications in future research.

Advanced Reinforcement Learning. Training RL models for urban decision can be expensive as it is risky to deploy bad policies to real cities. Additionally, building city simulators as the environment to collect online training data is often costly. In contrast, there exists large amount of offline data from the city's operations which can be utilized to train RL models in a pure data-driven way without the need for massive active interactions with the environment. Such training strategies, also known as **offline RL**, has become popular which combines the strengths of supervised learning and reinforcement learning, and we believe further exploration can significantly address the computational cost of training RL models for urban decision.

Reinforcement learning methods rely on quantified reward functions, which are not well-suited for tasks with non-quantifiable metrics. However, since the city represents a dense gathering of people, the actual feedback from citizens is the ultimate evaluation criterion for urban decision-making tasks, which is usually difficult to quantify. Recently, **reinforcement learning from human feedback (RLHF)** has significantly advanced the boundaries of what RL is capable of. For example, RL models can learn human preferences and generate policies that better match human preferences [41], or obey human instructions [149], or even design mechanisms that align with human values [91]. As for smart cities, RLHF makes it possible to learn from actual citizen feedback, and thus has the potential to realize human-centered urban decision making.

5.2 Other Urban Decision Issues

Control of Urban Infrastructure. With the development of machine learning, it is now possible to manage the urban infrastructures automatically, which can better meet the needs of urban development while consuming fewer resources compared to human operators. First, deep learning methods can be used for prediction and detection tasks to assist in infrastructure management. For example, Hu *et al.* [73] proposed a multi-scale convolutional networks to detect leakage in water networks. Second, reinforcement learning based approaches can be developed to directly control urban infrastructure. For instance, multi-agent reinforcement learning (MARL) is employed to regulate the voltage of power distribution networks [212]. The application of machine learning in urban infrastructure control is still in its early stage, and we recommend two related surveys [2, 57] for further details.

Urban Emergency Management. Developing machine learning methods for emergency management can significantly improve resource allocation efficiency and reduce disaster losses, which is a crucial future direction for smart cities. To achieve this, it is crucial to gather sufficient data on disasters, based on which machine learning methods offer promising solutions for emergency management. Initially, machine learning methods were used for prediction and simulation tasks such as forecasting disasters and estimating post-disaster resource requirements [74, 180]. With the rise of RL, more research has focused on decision-making tasks in emergency management. For example, Yang *et al.* [242] proposed a multi-agent reinforcement learning approach for volunteer scheduling to reduce response time and improve the efficiency of victim rescuing. Zhao *et al.* [256] developed a resource distribution framework for health crisis based on the DQN [142] model. We refer to two surveys [96, 179] for more details on urban emergency management.

Data Privacy. The massive data used in urban decision research is generated by people in cities and machines manipulated by people, such as mobility data of people and vehicles in urban decision making regarding healthcare and transportation. However, using such data requires caution to avoid compromising the privacy of urban residents, whose trajectories can be recovered from the data. Therefore, integrating privacy protection into data-driven models is an important research question in smart cities. Additionally, the data in the city may be distributed among different organizations, such as different companies providing logistics services and different bike-sharing platforms. It is also a challenge to manage the use of the data from different sources in an integrated manner to achieve maximum intelligence. In recent years, federated learning has been proposed that can train machine learning models using data from different platforms simultaneously without violating user privacy [105], which is a promising direction for future research and we believe it can significantly enhance the applicability of machine learning methods for urban decision-making.

Explainability. Although machine learning models significantly outperform traditional methods in urban decision-making tasks, they often function as black boxes with limited explainability, posing a critical issue in practical applications, as urban planning, transportation and healthcare typically involve multiple stakeholders. On the one hand, the actual deployment of the solution is often irreversible, necessitating adequate explanations to achieve consensus among stakeholders−−for instance, when constructing a planned road network. On the other hand, explainability helps us understand what the machine learning model captures, enabling iterative improvements by introducing inductive biases based on domain knowledge and verifying whether the model has learned them. In recent years, a series of studies on the explainability of machine learning models have been proposed [19, 26]. We believe that explainable machine learning is a promising future direction in urban decision-making, enhancing both the transparency and effectiveness of these advanced techniques.

5.3 Urban Decision and Simulation

In practical urban decision-making, we need to frequently evaluate the effects of different strategies to improve the model and finally obtain a solution for deployment. Only the obtained final solution is operated in the real world, while a large number of solutions in the training process must be evaluated virtually through simulation. Existing approaches tend to build a simplified simulator to simulate and evaluate the solutions given by the machine learning models. For example, in road network planning, a simplified model of traffic participants is constructed to evaluate different road plans [56]. However, the city is a complex system with non-trivial correlations between different factors, and a simplified simulator may make the model biased, leading to undesirable performance. Therefore, building a high-fidelity city simulator that models various kinds of dynamics in the city will be very helpful for urban decision research [249].

A high-precision urban simulator makes it possible to accurately evaluate the effects of different strategies, enabling better performance of the obtained solution when actually deployed. For instance, the simulator can be integrated into the environment of reinforcement learning to achieve successful simulation-to-real (sim2real) transfer [255].

The disparity in execution time between decision models and urban simulations presents a significant bottleneck in simulation-based urban decision-making. Decision models, particularly those using neural network inference in deep reinforcement learning, can deliver solutions within milliseconds. In contrast, urban simulations—due to the large number of simulated entities and intricate dynamics between them—are computationally expensive, often requiring several minutes, substantially slower than decision-making models. This discrepancy hampers the training process, as decision models are forced to wait for simulation results, leading to inefficiencies. Given that urban decision models typically necessitate millions of episodic interactions with the simulator, this delay can render model training prohibitively costly, if at all feasible.

Addressing this critical challenge requires future research to harmonize the disparate timescales of decision-making and simulation, creating an efficient, synchronized feedback loop without delay. One promising direction is the use of reward models [194], which approximate the outcomes of slow simulations using neural networks, thus bridging the gap between fast decision-making and slow simulation. We believe this intersection of urban decision-making and simulation holds considerable potential and advocate for increased attention to this area in future research.

6 CONCLUSION

Urban decision making can become substantially more effective and efficient with the help of machine learning. In this survey, we systematically reviewed existing approaches that use machine learning for urban decision making related to planning, transportation, and healthcare. Besides what has been widely studied, we discussed some open problems and future directions in smart cities, including recent advances in machine learning that have not yet been applied in smart cities, as well as other urban decision issues, such as the control of urban infrastructure and user privacy protection in data-driven urban decision models. This survey can be a valuable resource for researchers in related fields, aiding their understanding of the evolution of smart cities and fostering innovative developments in smart urban research.

REFERENCES

[1] Wejden Abdallah, Dalel Kanzari, Dorsaf Sallami, Kurosh Madani, and Khaled Ghedira. 2022. A deep reinforcement learning based decision-making approach for avoiding crowd situation within the case of Covid'19 pandemic. *Comput. Intell.* 38, 2 (2022), 416–437. https://doi.org/10.1111/coin.12516

[2] Tanveer Ahmad, Rafal Madonski, Dongdong Zhang, Chao Huang, and Asad Mujeeb. 2022. Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renewable and Sustainable Energy Reviews* 160 (2022), 112128.

[3] Meisam Akbarzadeh, Syed Sina Mohri, and Ehsan Yazdian. 2018. Designing bike networks using the concept of network clusters. *Applied network science* 3, 1 (2018), 1–21.

[4] Mesfer Al Duhayyim, Taiseer Abdalla Elfadil Eisa, Fahd N Al-Wesabi, Abdelzahir Abdelmaboud, Manar Ahmed Hamza, Abu Sarwar Zamani, Mohammed Rizwanullah, and Radwa Marzouk. 2022. Deep reinforcement learning enabled smart city recycling waste object classification. *Comput. Mater. Contin* 71 (2022), 5699–5715.

[5] David M Allen. 1971. Mean square error of prediction as a criterion for selecting variables. *Technometrics* 13, 3 (1971), 469–475.

[6] Zeyuan Allen-Zhu and Elad Hazan. 2016. Variance reduction for faster non-convex optimization. In *International conference on machine learning*. PMLR, 699–707.

[7] Nikolaos Askitas, Konstantinos Tatsiramos, and Bertrand Verheyden. 2021. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. *Scientific reports* 11, 1 (2021), 1–13.

[8] Safa Ben Atitallah, Maha Driss, Wadii Boulila, and Henda Ben Ghézala. 2020. Leveraging Deep Learning and IoT big data analytics to support the smart cities development: Review and future directions. *Computer Science Review* 38 (2020), 100303.

[9] Raghav Awasthi, Keerat Kaur Guliani, Arshita Bhatt, Mehrab Singh Gill, Aditya Nagori, Ponnurangam Kumaraguru, and Tavpritesh Sethi. 2020. VacSIM: Learning Effective Strategies for COVID-19 Vaccine Distribution using Reinforcement Learning. *CoRR* abs/2009.06602 (2020).

[10] Sangjun Bae, Balázs Kulcsár, and Sébastien Gros. 2024. Personalized dynamic pricing policy for electric vehicles: Reinforcement learning approach. *Transportation Research Part C: Emerging Technologies* 161 (2024), 104540.

[11] Jie Bao, Tianfu He, Sijie Ruan, Yanhua Li, and Yu Zheng. 2017. Planning bike lanes based on sharing-bikes' trajectories. In *KDD*. 1377–1386.

[12] Amir Hossein Barahimi, Alireza Eydi, and Abdolah Aghaie. 2021. Multi-modal urban transit network design considering reliability: multi-objective bi-level optimization. *Reliability Engineering & System Safety* 216 (2021), 107922.

[13] Ana LC Bazzan. 2009. Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. *Autonomous Agents and Multi-Agent Systems* 18 (2009), 342–375.

[14] Bryan P. Bednarski, Akash Deep Singh, and William M. Jones. 2021. On collaborative reinforcement learning to optimize the redistribution of critical medical supplies throughout the COVID-19 pandemic. *J. Am. Medical Informatics Assoc.* 28, 4 (2021), 874–878. https://doi.org/10.1093/jamia/ocaa324

[15] Alireza Beigi, Amin Yousefpour, Amirreza Yasami, JF Gómez-Aguilar, Stelios Bekiros, and Hadi Jahanshahi. 2021. Application of reinforcement learning for effective vaccination strategies of coronavirus disease 2019 (COVID-19). *The European Physical Journal Plus* 136, 5 (2021), 1–22.

[16] Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. 2021. Machine learning for combinatorial optimization: a methodological tour d'horizon. *European Journal of Operational Research* 290, 2 (2021), 405–421.

[17] Dimitris Bertsimas and Melvyn Sim. 2003. Robust discrete optimization and network flows. *Mathematical programming* 98, 1 (2003), 49–71.

[18] Dimitris Bertsimas and John N Tsitsiklis. 1997. *Introduction to linear optimization*. Vol. 6. Athena Scientific Belmont, MA.

[19] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura, and Peter Eckersley. 2020. Explainable machine learning in deployment. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 648–657.

[20] Jieyi Bi, Yining Ma, Jiahai Wang, Zhiguang Cao, Jinbiao Chen, Yuan Sun, and Yeow Meng Chee. 2022. Learning generalizable models for vehicle routing problems via knowledge distillation. *Advances in Neural Information Processing Systems* 35 (2022), 31226–31238.

[21] Pierre Bonami, Andrea Lodi, and Giulia Zarpellon. 2018. Learning a classification of mixed-integer quadratic programming problems. In *CPAIOR*. Springer, 595–604.

[22] Benjamin D Bowes, Cheng Wang, Mehmet B Ercan, Teresa B Culver, Peter A Beling, and Jonathan L Goodall. 2022. Reinforcement learning-based real-time control of coastal urban stormwater systems to mitigate flooding and improve water quality. *Environmental Science: Water Research & Technology* 8, 10 (2022), 2065–2086.

[23] Stephen P Boyd and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

[24] Kris Braekers, Katrien Ramaekers, and Inneke Van Nieuwenhuyse. 2016. The vehicle routing problem: State of the art classification and review. *Computers & industrial engineering* 99 (2016), 300–313.

[25] Sébastien Bubeck et al. 2015. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning* 8, 3-4 (2015), 231–357.

[26] Nadia Burkart and Marco F Huber. 2021. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research* 70 (2021), 245–317.

[27] Sabah Bushaj, Xuecheng Yin, Arjeta Beqiri, Donald Andrews, and İ Esra Büyüktahtakın. 2022. A simulation-deep reinforcement learning (SiRL) approach for epidemic control optimization. *Annals of Operations Research* (2022), 1–33.

[28] Jose Caceres-Cruz, Pol Arias, Daniel Guimarans, Daniel Riera, and Angel A Juan. 2014. Rich vehicle routing problem: Survey. *ACM Computing Surveys (CSUR)* 47, 2 (2014), 1–28.

[29] Rich Caruana and Alexandru Niculescu-Mizil. 2006. An empirical comparison of supervised learning algorithms. In *ICML*. 161–168.

[30] Noe Casas. 2017. Deep deterministic policy gradient for urban traffic light control. *arXiv preprint arXiv:1703.09035* (2017).

[31] Marisdea Castiglione, Rosita De Vincentis, Marialisa Nigro, and Vittorio Rega. 2022. Bike network design: an approach based on micro-mobility geo-referenced data. *Transportation research procedia* 62 (2022), 51–58.

[32] Mohamed-Amine Chadi and Hajar Mousannif. 2022. A Reinforcement Learning Based Decision Support Tool for Epidemic Control: Validation Study for COVID-19. *Appl. Artif. Intell.* 36, 1 (2022). https://doi.org/10.1080/08839514.2022.2031821

[33] Eduarda TC Chagas, Pedro H Barros, Isadora Cardoso-Pereira, Igor V Ponte, Pablo Ximenes, Flávio Figueiredo, Fabricio Murai, Ana Paula Couto da Silva, Jussara M Almeida, Antonio AF Loureiro, et al. 2021. Effects of population mobility on the COVID-19 spread in Brazil. *PloS one* 16, 12 (2021).

[34] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. 2020. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 3414–3421.

[35] Chuqiao Chen, Fugen Yao, Dong Mo, Jiangtao Zhu, and Xiqun Michael Chen. 2021. Spatial-temporal pricing for ride-sourcing platform with reinforcement learning. *Transportation Research Part C: Emerging Technologies* 130 (2021), 103272.

[36] Haipeng Chen, Yan Jiao, Zhiwei Qin, Xiaocheng Tang, Hao Li, Bo An, Hongtu Zhu, and Jieping Ye. 2019. InBEDE: Integrating contextual bandit with TD learning for joint pricing and dispatch of ride-hailing platforms. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 61–70.

[37] Jianguo Chen, Kenli Li, Keqin Li, Philip S Yu, and Zeng Zeng. 2021. Dynamic planning of bicycle stations in dockless public bicycle-sharing system using gated graph neural network. *ACM Transactions on Intelligent Systems and Technology (TIST)* 12, 2 (2021), 1–22.

[38] Kehua Chen, Hongcheng Wang, Borja Valverde-Pérez, Siyuan Zhai, Luca Vezzaro, and Aijie Wang. 2021. Optimal control towards sustainable wastewater treatment plants based on multi-agent reinforcement learning. *Chemosphere* 279 (2021), 130498.

[39] Qi Chen, Wei Wang, Fangyu Wu, Suparna De, Ruili Wang, Bailing Zhang, and Xin Huang. 2019. A survey on an emerging area: Deep learning for smart city data. *IEEE Transactions on Emerging Topics in Computational Intelligence* 3, 5 (2019), 392–410.

[40] Yujiao Chen, Leslie K Norford, Holly W Samuelson, and Ali Malkawi. 2018. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings* 169 (2018), 195–205.

[41] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems* 30 (2017).

[42] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. 2019. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems* 21, 3 (2019), 1086–1095.

[43] Jiaxu Cui, Bo Yang, and Xia Hu. 2019. Deep Bayesian optimization on attributed graphs. In *AAAI*, Vol. 33. 1377–1384.

[44] Jiaxu Cui, Bo Yang, Bingyi Sun, Xia Hu, and Jiming Liu. 2020. Scalable and Parallel Deep Bayesian Optimization on Attributed Graphs. *IEEE Transactions on Neural Networks and Learning Systems* (2020).

[45] Marina Danilova, Pavel Dvurechensky, Alexander Gasnikov, Eduard Gorbunov, Sergey Guminov, Dmitry Kamzolov, and Innokentiy Shibaev. 2022. Recent theoretical advances in non-convex optimization. In *High-Dimensional Optimization and Probability: With a View Towards Data Science*. Springer, 79–163.

[46] Ahmed Darwish, Momen Khalil, and Karim Badawi. 2020. Optimising Public Bus Transit Networks Using Deep Reinforcement Learning. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 1–7.

[47] Djamel Djenouri, Roufaida Laidi, Youcef Djenouri, and Ilangko Balasingham. 2019. Machine learning for smart building applications: Review and taxonomy. *ACM Computing Surveys (CSUR)* 52, 2 (2019), 1–36.

[48] Marco Dorigo, Gianni Di Caro, and Luca M Gambardella. 1999. Ant algorithms for discrete optimization. *Artificial life* 5, 2 (1999), 137–172.

[49] Thomas M Drake, Annemarie B Docherty, Thomas G Weiser, Steven Yule, Aziz Sheikh, and Ewen M Harrison. 2020. The effects of physical distancing on population mobility during the COVID-19 pandemic in the UK. *The Lancet Digital Health* 2, 8 (2020), e385–e387.

[50] Xinqi Du, Hechang Chen, Bo Yang, Cheng Long, and Songwei Zhao. 2023. HRL4EC: Hierarchical reinforcement learning for multi-mode epidemic control. *Information Sciences* 640 (2023), 119065.

[51] Lu Duan, Yang Zhan, Haoyuan Hu, Yu Gong, Jiangwen Wei, Xiaodong Zhang, and Yinghui Xu. 2020. Efficiently solving the practical vehicle routing problem: A novel joint learning approach. In *KDD*. 3054–3063.

[52] BRP e Oliveira, JA De Vasconcelos, JFF Almeida, and LR Pinto. 2020. A simulation-optimisation approach for hospital beds allocation. *International Journal of Medical Informatics* 141 (2020), 104174.

[53] Ezekiel J Emanuel, Govind Persad, Adam Kern, Allen Buchanan, Cécile Fabre, Daniel Halliday, Joseph Heath, Lisa Herzog, RJ Leland, Ephrem T Lemango, et al. 2020. An ethical framework for global vaccine allocation. *Science* 369, 6509 (2020), 1309–1312.

[54] Alexandre Fabregat, Anton Vernet, Marc Vernet, Lluís Vázquez, and Josep A Ferré. 2022. Using Machine Learning to estimate the impact of different modes of transport and traffic restriction strategies on urban air quality. *Urban Climate* 45 (2022), 101284.

[55] Zhou Fang, Jiaxin Qi, Lubin Fan, Jianqiang Huang, Ying Jin, and Tianren Yang. 2022. A framework for human-computer interactive street network design based on a multi-stage deep learning approach. *Computers, Environment and Urban Systems* 96 (2022), 101853.

[56] Reza Zanjirani Farahani, Elnaz Miandoabchi, Wai Yuen Szeto, and Hannaneh Rashidi. 2013. A review of urban transportation network design problems. *European journal of operational research* 229, 2 (2013), 281–302.

[57] Guangtao Fu, Yiwen Jin, Siao Sun, Zhiguo Yuan, and David Butler. 2022. The role of deep learning in urban water management: A critical review. *Water Research* (2022), 118973.

[58] Masao Fukushima. 1984. A modified Frank-Wolfe algorithm for solving the traffic assignment problem. *Transportation Research Part B: Methodological* 18, 2 (1984), 169–177.

[59] Wade Genders and Saiedeh Razavi. 2016. Using a deep reinforcement learning agent for traffic signal control. *arXiv preprint arXiv:1611.01142* (2016).

[60] Ge Guo and Yangguang Xu. 2020. A deep reinforcement learning approach to ride-sharing vehicle dispatching in autonomous mobility-on-demand systems. *IEEE Intelligent Transportation Systems Magazine* 14, 1 (2020), 128–140.

[61] Marina Haliem, Ganapathy Mani, Vaneet Aggarwal, and Bharat Bhargava. 2020. A distributed model-free ride-sharing algorithm with pricing using deep reinforcement learning. In *Proceedings of the 4th ACM Computer Science in Cars Symposium*. 1–10.

[62] Benjamin Han, Hyungjun Lee, and Sébastien Martin. 2022. Real-Time Rideshare Driver Supply Values Using Online Reinforcement Learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2968–2976.

[63] Qianyue Hao, Wenzhen Huang, Fengli Xu, Kun Tang, and Yong Li. 2022. Reinforcement Learning Enhances the Experts: Large-scale COVID-19 Vaccine Allocation with Multi-factor Contact Network. In *KDD*. ACM, 4684–4694.

[64] Qianyue Hao, Fengli Xu, Lin Chen, Pan Hui, and Yong Li. 2021. Hierarchical Reinforcement Learning for Scarce Medical Resource Allocation with Imperfect Information. In *KDD*. ACM, 2955–2963.

[65] Qianyue Hao, Fengli Xu, Lin Chen, Pan Hui, and Yong Li. 2022. Hierarchical Multi-agent Model for Reinforced Medical Resource Allocation with Imperfect Information. *ACM Transactions on Intelligent Systems and Technology* 14, 1 (2022), 1–27.

[66] John A Hartigan, Manchek A Wong, et al. 1979. A k-means clustering algorithm. *Applied statistics* 28, 1 (1979), 100–108.

[67] Johannes Haushofer and C Jessica E Metcalf. 2020. Which interventions work best in a pandemic? *Science* 368, 6495 (2020), 1063–1065.

[68] Tianfu He, Jie Bao, Sijie Ruan, Ruiyuan Li, Yanhua Li, Hui He, and Yu Zheng. 2019. Interactive bike lane planning using sharing bikes' trajectories. *IEEE Transactions on Knowledge and Data Engineering* 32, 8 (2019), 1529–1542.

[69] SungKu Heo, KiJeon Nam, Jorge Loy-Benitez, Qian Li, SeungChul Lee, and ChangKyoo Yoo. 2019. A deep reinforcement learning-based autonomous ventilation control system for smart indoor air quality management in a subway station. *Energy and Buildings* 202 (2019), 109440.

[70] Alexandre Heuillet, Fabien Couthouis, and Natalia Díaz-Rodríguez. 2021. Explainability in deep reinforcement learning. *Knowledge-Based Systems* 214 (2021), 106685.

[71] André Hottung, Bhanu Bhandari, and Kevin Tierney. 2020. Learning a latent search space for routing problems using variational autoencoders. In *International Conference on Learning Representations*.

[72] Chengyu Hu, Junyi Cai, Deze Zeng, Xuesong Yan, Wenyin Gong, and Ling Wang. 2020. Deep reinforcement learning based valve scheduling for pollution isolation in water distribution network. *Math. Biosci. Eng* 17 (2020), 105–121.

[73] Xuan Hu, Yongming Han, Bin Yu, Zhiqiang Geng, and Jinzhen Fan. 2021. Novel leakage detection and water loss management of urban water supply network using multiscale neural networks. *Journal of Cleaner Production* 278 (2021), 123611.

[74] Di Huang, Shuaian Wang, and Zhiyuan Liu. 2021. A systematic review of prediction methods for emergency management. *International Journal of Disaster Risk Reduction* 62 (2021), 102412.

[75] Jianbin Huang, Longji Huang, Meijuan Liu, He Li, Qinglin Tan, Xiaoke Ma, Jiangtao Cui, and De-Shuang Huang. 2022. Deep reinforcement learning-based trajectory pricing on ride-hailing platforms. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 3 (2022), 1–19.

[76] Xiaohui Huang, Jiahao Ling, Xiaofei Yang, Xiong Zhang, and Kaiming Yang. 2023. Multi-Agent Mix Hierarchical Deep Reinforcement Learning for Large-Scale Fleet Management. *IEEE Transactions on Intelligent Transportation Systems* (2023).

[77] Ying-Chao Hung and George Michailidis. 2022. A Novel Data-Driven Approach for Solving the Electric Vehicle Charging Station Location-Routing Problem. *IEEE Transactions on Intelligent Transportation Systems* 23, 12 (2022), 23858–23868.

[78] JQ James, Wen Yu, and Jiatao Gu. 2019. Online vehicle routing with neural combinatorial optimization and deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems* 20, 10 (2019), 3806–3817.

[79] Minhyuk Jeung, Jiyi Jang, Kwangsik Yoon, and Sang-Soo Baek. 2023. Data assimilation for urban stormwater and water quality simulations using deep reinforcement learning. *Journal of Hydrology* 624 (2023), 129973.

[80] Yuan Jiang, Zhiguang Cao, Yaoxin Wu, Wen Song, and Jie Zhang. 2024. Ensemble-based deep reinforcement learning for vehicle routing problems under distribution shift. *Advances in Neural Information Processing Systems* 36 (2024).

[81] Yuan Jiang, Zhiguang Cao, Yaoxin Wu, and Jie Zhang. 2023. Multi-view graph contrastive learning for solving vehicle routing problems. In *Uncertainty in Artificial Intelligence*. PMLR, 984–994.

[82] Jiarui Jin, Ming Zhou, Weinan Zhang, Minne Li, Zilong Guo, Zhiwei Qin, Yan Jiao, Xiaocheng Tang, Chenxi Wang, Jun Wang, et al. 2019. Coride: joint order dispatching and fleet management for multi-scale ride-hailing platforms. In *CIKM*. 1983–1992.

[83] Jiahui Jin, Xiaoxuan Zhu, Biwei Wu, Jinghui Zhang, and Yuxiang Wang. 2021. A dynamic and deadline-oriented road pricing mechanism for urban traffic management. *Tsinghua Science and Technology* 27, 1 (2021), 91–102.

[84] Yan Jin, Yuandong Ding, Xuanhao Pan, Kun He, Li Zhao, Tao Qin, Lei Song, and Jiang Bian. 2023. Pointerformer: Deep reinforced multi-pointer transformer for the traveling salesman problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 8132–8140.

[85] Chaitanya K Joshi, Thomas Laurent, and Xavier Bresson. 2019. An efficient graph convolutional network technique for the travelling salesman problem. *arXiv preprint arXiv:1906.01227* (2019).

[86] Jintao Ke, Feng Xiao, Hai Yang, and Jieping Ye. 2020. Learning to delay in ride-sourcing systems: a multi-agent deep reinforcement learning framework. *IEEE Transactions on Knowledge and Data Engineering* 34, 5 (2020), 2280–2292.

[87] Harshad Khadilkar, Tanuja Ganu, and D Seetharam. 2020. Optimising lockdown policies for epidemic control using reinforcement learning: An AI-driven control approach compatible with existing disease and network models. *Transactions of the Indian National Academy of Engineering* (2020), 1–4.

[88] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

[89] Catherine Kling and Jonathan Rubin. 1997. Bankable permits for the control of environmental pollution. *Journal of Public Economics* 64, 1 (1997).

[90] Wouter Kool, Herke van Hoof, and Max Welling. 2018. Attention, Learn to Solve Routing Problems!. In *ICLR*.

[91] Raphael Koster, Jan Balaguer, Andrea Tacchetti, Ari Weinstein, Tina Zhu, Oliver Hauser, Duncan Williams, Lucy Campbell-Gillingham, Phoebe Thacker, Matthew Botvinick, et al. 2022. Human-centred mechanism design with Democratic AI. *Nature Human Behaviour* 6, 10 (2022), 1398–1407.

[92] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* 160, 1 (2007), 3–24.

[93] Markus Kruber, Marco E Lübbecke, and Axel Parmentier. 2017. Learning when to use a decomposition. In *Integration of AI and OR Techniques in Constraint Programming: 14th International Conference, CPAIOR 2017, Padua, Italy, June 5-8, 2017, Proceedings 14*. Springer, 202–210.

[94] Gloria Hyunjung Kwak, Lowell Ling, and Pan Hui. 2021. Deep reinforcement learning approaches for global public health strategies for COVID-19 pandemic. *PloS one* 16, 5 (2021), e0251550.

[95] Yeong-Dae Kwon, Jinho Choo, Byoungjip Kim, Iljoo Yoon, Youngjune Gwon, and Seungjai Min. 2020. Pomo: Policy optimization with multiple optima for reinforcement learning. *Advances in Neural Information Processing Systems* 33 (2020), 21188–21198.

[96] Christos Kyrkou, Panayiotis Kolios, Theocharis Theocharides, and Marios Polycarpou. 2022. Machine learning for emergency management: A survey and future outlook. *Proc. IEEE* 111, 1 (2022), 19–41.

[97] Luis Alfonso Lastras-Montaño. 2019. Information Theoretic lower bounds on negative log likelihood. In *ICLR*.

[98] Der-Horng Lee, Hao Wang, Ruey Long Cheu, and Siew Hoon Teo. 2004. Taxi dispatch system based on current demands and real-time traffic conditions. *Transportation Research Record* 1882, 1 (2004), 193–200.

[99] Zengxiang Lei and Satish V Ukkusuri. 2023. Scalable reinforcement learning approaches for dynamic pricing in ride-hailing systems. *Transportation Research Part B: Methodological* 178 (2023), 102848.

[100] Nixie S Lesmana, Xuan Zhang, and Xiaohui Bei. 2019. Balancing efficiency and fairness in on-demand ridesourcing. *NeurIPS* 32 (2019).

[101] Feixue Li, Zhifeng Li, Honghua Chen, Zhenjie Chen, and Manchun Li. 2020. An agent-based learning-embedded model (ABM-learning) for urban land use planning: A case study of residential land growth simulation in Shenzhen, China. *Land Use Policy* 95 (2020), 104620.

[102] Hao Li, Luqi Wang, Mengxi Zhang, Yihan Lu, and Weibing Wang. 2022. Effects of vaccination and non-pharmaceutical interventions and their lag times on the COVID-19 pandemic: Comparison of eight countries. *PLoS neglected tropical diseases* 16, 1 (2022), e0010101.

[103] Jingwen Li, Yining Ma, Ruize Gao, Zhiguang Cao, Andrew Lim, Wen Song, and Jie Zhang. 2021. Deep reinforcement learning for solving the heterogeneous capacitated vehicle routing problem. *IEEE Transactions on Cybernetics* 52, 12 (2021), 13572–13585.

[104] Jingwen Li, Liang Xin, Zhiguang Cao, Andrew Lim, Wen Song, and Jie Zhang. 2021. Heterogeneous attentions for solving pickup and delivery problem via deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems* 23, 3 (2021), 2306–2315.

[105] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. 2020. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine* 37, 3 (2020), 50–60.

[106] Yanfei Li, Zheyu Liu, and Hui Liu. 2021. A novel ensemble reinforcement learning gated unit model for daily PM2. 5 forecasting. *Air Quality, Atmosphere & Health* 14 (2021), 443–453.

[107] Yexin Li, Yu Zheng, and Qiang Yang. 2018. Dynamic Bike Reposition: A Spatio-Temporal Reinforcement Learning Approach. In *Kdd*. 1724–1733.

[108] Yexin Li, Yu Zheng, and Qiang Yang. 2019. Efficient and effective express via contextual cooperative reinforcement learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 510–519.

[109] Yexin Li, Yu Zheng, and Qiang Yang. 2020. Cooperative multi-agent reinforcement learning in express system. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 805–814.

[110] Enming Liang, Kexin Wen, William HK Lam, Agachai Sumalee, and Renxin Zhong. 2021. An integrated reinforcement learning and centralized programming approach for online taxi dispatching. *IEEE Transactions on Neural Networks and Learning Systems* 33, 9 (2021), 4742–4756.

[111] Yuan Liang. 2024. Fairness-Aware Dynamic Ride-Hailing Matching Based on Reinforcement Learning. *Electronics* 13, 4 (2024), 775.

[112] Guitang Liao, Peng He, Xuesong Gao, Zhengyu Lin, Chengyi Huang, Wei Zhou, Ouping Deng, Chenghua Xu, and Liangji Deng. 2022. Land use optimization of rural production–living–ecological space at different scales based on the BP–ANN and CLUE–S models. *Ecological Indicators* 137 (2022), 108710.

[113] Ziqi Liao. 2003. Real-time taxi dispatching using global positioning systems. *Commun. ACM* 46, 5 (2003), 81–83.

[114] Pieter J. K. Libin, Arno Moonens, Timothy Verstraeten, Fabian Perez-Sanjines, Niel Hens, Philippe Lemey, and Ann Nowé. 2020. Deep Reinforcement Learning for Large-Scale Epidemic Control. In *ECML PKDD (Lecture Notes in Computer Science, Vol. 12461)*. Springer, 155–170.

[115] Bo Lin, Bissan Ghaddar, and Jatin Nathwani. 2021. Deep reinforcement learning for the electric vehicle routing problem with time windows. *IEEE Transactions on Intelligent Transportation Systems* 23, 8 (2021), 11528–11538.

[116] Kaixiang Lin, Renyu Zhao, Zhe Xu, and Jiayu Zhou. 2018. Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1774–1783.

[117] Fang Liu and Weilun Sun. 2020. Urban Residential Area Sprawl Simulation of Metropolitan "Suburbanization" Trend in Beijing. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 4938–4942.

[118] Qi Liu, Jiahao Liu, Weiwei Le, Zhaoxia Guo, and Zhenggang He. 2019. Data-driven intelligent location of public charging stations for electric vehicles. *Journal of cleaner production* 232 (2019), 531–541.

[119] Tianjiao Liu, Qiang Wang, Wenqi Zhang, and Chen Xu. 2022. CoRLNF: Joint Spatio-Temporal Pricing and Fleet Management for Ride-Hailing Platforms. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 395–401.

[120] Yu Liu, Jingtao Ding, and Yong Li. 2022. Developing knowledge graph based system for urban computing. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs*. 3–7.

[121] Yu Liu, Jingtao Ding, and Yong Li. 2023. KnowSite: Leveraging Urban Knowledge Graph for Site Selection. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*. 1–12.

[122] Yilin Liu, Guiyang Luo, Quan Yuan, Jinglin Li, Lei Jin, Bo Chen, and Rui Pan. 2023. GPLight: Grouped Multi-agent Reinforcement Learning for Large-scale Traffic Signal Control.. In *IJCAI*. 199–207.

[123] Yang Liu, Fanyou Wu, Cheng Lyu, Shen Li, Jieping Ye, and Xiaobo Qu. 2022. Deep dispatching: A deep reinforcement learning approach for vehicle dispatching on online ride-hailing platform. *Transportation Research Part E: Logistics and Transportation Review* 161 (2022), 102694.

[124] Zhidan Liu, Jiangzhou Li, and Kaishun Wu. 2020. Context-aware taxi dispatching at city-scale using deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems* 23, 3 (2020), 1996–2009.

[125] Caicheng Long, Zixin Jiang, Jingfang Shangguan, Taiping Qing, Peng Zhang, and Bo Feng. 2021. Applications of carbon dots in environmental pollution control: A review. *Chemical Engineering Journal* 406 (2021), 126848.

[126] Yican Lou, Jia Wu, and Yunchuan Ran. 2022. Meta-reinforcement learning for multiple traffic signals control. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4264–4268.

[127] Hao Lu, Xingwen Zhang, and Shuang Yang. 2019. A learning-based iterative method for solving vehicle routing problems. In *ICLR*.

[128] Jiaming Lu, Jingqing Ruan, Haoyuan Jiang, Ziyue Li, Hangyu Mao, and Rui Zhao. 2024. DuaLight: Enhancing Traffic Signal Control by Leveraging Scenario-Specific and Scenario-Shared Knowledge. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. 1283–1291.

[129] Yan Lyu, Chi-Yin Chow, Victor CS Lee, Joseph KY Ng, Yanhua Li, and Jia Zeng. 2019. CB-Planner: A bus line planning framework for customized bus systems. *Transportation Research Part C: Emerging Technologies* 101 (2019), 233–253.

[130] Hongyao Ma, Fei Fang, and David C Parkes. 2020. Spatio-temporal pricing for ridesharing platforms. *ACM SIGecom Exchanges* 18, 2 (2020), 53–57.

[131] Hongyao Ma, Fei Fang, and David C Parkes. 2022. Spatio-temporal pricing for ridesharing platforms. *Operations Research* 70, 2 (2022), 1025–1041.

[132] Qiang Ma, Suwen Ge, Danyang He, Darshan Thaker, and Iddo Drori. 2020. Combinatorial Optimization by Graph Pointer Networks and Hierarchical Reinforcement Learning. In *AAAI Workshop on Deep Learning on Graphs: Methodologies and Applications*.

[133] Yining Ma, Zhiguang Cao, and Yeow Meng Chee. 2024. Learning to search feasible and infeasible regions of routing problems with flexible neural k-opt. *Advances in Neural Information Processing Systems* 36 (2024).

[134] Yi Ma, Xiaotian Hao, Jianye Hao, Jiawen Lu, Xing Liu, Tong Xialiang, Mingxuan Yuan, Zhigang Li, Jie Tang, and Zhaopeng Meng. 2021. A hierarchical reinforcement learning based optimization framework for large-scale dynamic pickup and delivery problems. *Advances in Neural Information Processing Systems* 34 (2021), 23609–23620.

[135] Yining MA, Jingwen LI, Zhiguang CAO, Wen SONG, Hongliang GUO, Yuejiao GONG, and Meng Chee CHEE. 2022. Efficient neural neighborhood search for pickup and delivery problems.(2022). In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence Vienna, Austria*. 23–29.

[136] Yining Ma, Jingwen Li, Zhiguang Cao, Wen Song, Le Zhang, Zhenghua Chen, and Jing Tang. 2021. Learning to iteratively solve routing problems with dual-aspect collaborative transformer. *Advances in Neural Information Processing Systems* 34 (2021), 11096–11107.

[137] Shie Mannor, Dori Peleg, and Reuven Rubinstein. 2005. The cross entropy method for classification. In *Proceedings of the 22nd international conference on Machine learning*. 561–568.

[138] Laura Matrajt, Julia Eaton, Tiffany Leung, Dobromir Dimitrov, Joshua T Schiffer, David A Swan, and Holly Janes. 2021. Optimizing vaccine allocation for COVID-19 vaccines shows the potential role of single-dose vaccination. *Nature communications* 12, 1 (2021), 3449.

[139] Devon E McMahon, Gregory A Peters, Louise C Ivers, and Esther E Freeman. 2020. Global resource shortages during COVID-19: Bad news for low-income countries. *PLoS neglected tropical diseases* 14, 7 (2020), e0008412.

[140] Marc-Olivier Metais, O Jouini, Yannick Perez, Jaâfar Berrada, and Emilia Suomalainen. 2022. Too much or not enough? Planning electric vehicle charging infrastructure: A review of modeling options. *Renewable and Sustainable Energy Reviews* 153 (2022), 111719.

[141] George J Milne, Joel K Kelso, Heath A Kelly, Simon T Huband, and Jodie McVernon. 2008. A small community model for the transmission of infectious diseases: comparison of school closure as an intervention in individual-based models of an influenza pandemic. *PloS one* 3, 12 (2008).

[142] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *nature* 518, 7540 (2015), 529–533.

[143] Nguyen Hai Nam, Phan Thi My Tien, Le Van Truong, Toka Aziz El-Ramly, Pham Gia Anh, Nguyen Thi Hien, El Marabea Mahmoud, Mennatullah Mohamed Eltaras, Sarah Abd Elaziz Khader, Mohammed Salah Desokey, et al. 2021. Early centralized isolation strategy for all confirmed cases of COVID-19 remains a core intervention to disrupt the pandemic spreading significantly. *PloS one* 16, 7 (2021), e0254012.

[144] Mohammadreza Nazari, Afshin Oroojlooy, Lawrence Snyder, and Martin Takác. 2018. Reinforcement learning for solving the vehicle routing problem. *Advances in neural information processing systems* 31 (2018).

[145] Yurii Nesterov. 1983. A method for unconstrained convex minimization problem with the rate of convergence O (1/kˆ 2). In *Doklady an ussr*, Vol. 269. 543–547.

[146] Yuan Min Ni and Lei Li. 2014. Garbage Incineration and Intelligent Fusion Strategy of Secondary Pollution Control. In *Advanced Materials Research*, Vol. 853. Trans Tech Publ, 323–328.

[147] Abu Quwsar Ohi, MF Mridha, Muhammad Mostafa Monowar, Md Hamid, et al. 2020. Exploring optimal control of epidemic spread using reinforcement learning. *Scientific reports* 10, 1 (2020), 1–19.

[148] Luis E Olmos, Maria Sol Tadeo, Dimitris Vlachogiannis, Fahad Alhasoun, Xavier Espinet Alegre, Catalina Ochoa, Felipe Targa, and Marta C González. 2020. A data science framework for planning the growth of bicycle infrastructures. *Transportation research part C: emerging technologies* 115 (2020), 102640.

[149] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35 (2022), 27730–27744.

[150] Erhun Özkan and Amy R Ward. 2020. Dynamic matching for real-time ride sharing. *Stochastic Systems* 10, 1 (2020), 29–70.

[151] Regina Padmanabhan, Nader Meskin, Tamer Khattab, Mujahed Shraim, and Mohammed Al-Hitmi. 2021. Reinforcement learning-based decision support system for COVID-19. *Biomedical Signal Processing and Control* 68 (2021), 102676.

[152] Xuanhao Pan, Yan Jin, Yuandong Ding, Mingxiao Feng, Li Zhao, Lei Song, and Jiang Bian. 2023. H-tsp: Hierarchically solving the large-scale traveling salesman problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 9345–9353.

[153] Gina M Piscitello, Esha M Kapania, William D Miller, Juan C Rojas, Mark Siegler, and William F Parker. 2020. Variation in ventilator allocation guidelines by US state during the coronavirus disease 2019 pandemic: a systematic review. *JAMA network open* 3, 6 (2020), e2012606–e2012606.

[154] Erika Puiutta and Eric MSP Veith. 2020. Explainable reinforcement learning: A survey. In *International cross-domain conference for machine learning and knowledge extraction*. Springer, 77–95.

[155] Guoyang Qin, Qi Luo, Yafeng Yin, Jian Sun, and Jieping Ye. 2021. Optimizing matching time intervals for ride-hailing services using reinforcement learning. *Transportation Research Part C: Emerging Technologies* 129 (2021), 103239.

[156] Yiming Qin, Nanxuan Zhao, Bin Sheng, and Rynson WH Lau. 2024. Text2City: One-Stage Text-Driven Urban Layout Regeneration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 4578–4586.

[157] Guo Qiu, Rui Song, Shiwei He, Wangtu Xu, and Min Jiang. 2018. Clustering passenger trip data for the potential passenger investigation and line design of customized commuter bus. *IEEE Transactions on Intelligent Transportation Systems* 20, 9 (2018), 3351–3360.

[158] Wei Qiu, Haipeng Chen, and Bo An. 2019. Dynamic Electronic Toll Collection via Multi-Agent Deep Reinforcement Learning with Edge-Based Graph Convolutional Networks. In *IJCAI*. 4568–4574.

[159] Ted K Ralphs, Leonid Kopman, William R Pulleyblank, and Leslie E Trotter. 2003. On the capacitated vehicle routing problem. *Mathematical programming* 94 (2003), 343–359.

[160] CS Rao. 2007. *Environmental pollution control engineering*. New Age International.

[161] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gómez Colmenarejo, Alexander Novikov, Gabriel Barth-maron, Mai Giménez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley Edwards, Nicolas Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar, and Nando de Freitas. 2022. A Generalist Agent. *Transactions on Machine Learning Research* (2022). https://openreview.net/forum?id=1ikK0kHjvj Featured Certification, Outstanding Certification.

[162] Stefano Giovanni Rizzo, Giovanna Vantini, and Sanjay Chawla. 2019. Time critic policy gradient methods for traffic signal control in complex and congested scenarios. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 1654–1664.

[163] Herbert Robbins and Sutton Monro. 1951. A stochastic approximation method. *The annals of mathematical statistics* (1951), 400–407.

[164] Jingqing Ruan, Ziyue Li, Hua Wei, Haoyuan Jiang, Jiaming Lu, Xuantang Xiong, Hangyu Mao, and Rui Zhao. 2024. CoSLight: Co-optimizing Collaborator Selection and Decision-making to Enhance Traffic Signal Control. In *KDD*.

[165] Soheil Sadeghi Eshkevari, Xiaocheng Tang, Zhiwei Qin, Jinhan Mei, Cheng Zhang, Qianying Meng, and Jia Xu. 2022. Reinforcement Learning in the Wild: Scalable RL Dispatching Algorithm Deployed in Ridehailing Marketplace. In *KDD*. 3838–3848.

[166] Kimihiro Sato, Toru Seo, and Takashi Fuse. 2021. A reinforcement learning-based dynamic congestion pricing method for the morning commute problems. *Transportation Research Procedia* 52 (2021), 347–355.

[167] Martin Schlappa, Jonas Hegemann, and Stefan Spinler. 2022. Optimizing Control of Waste Incineration Plants Using Reinforcement Learning and Digital Twins. *IEEE Transactions on Engineering Management* (2022).

[168] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. 2015. Trust region policy optimization. In *International conference on machine learning*. PMLR, 1889–1897.

[169] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).

[170] Brendon Sen-Crowe, Mason Sutherland, Mark McKenney, and Adel Elkbuli. 2021. A closer look into global hospital beds capacity and resource shortages during the COVID-19 pandemic. *Journal of Surgical Research* 260 (2021), 56–63.

[171] Kiam Tian Seow, Nam Hai Dang, and Der-Horng Lee. 2009. A collaborative multiagent taxi-dispatch system. *IEEE Transactions on Automation science and engineering* 7, 3 (2009), 607–616.

[172] Wenzhe Shang, Junjie Liu, Congcong Wang, Jiayu Li, and Xilei Dai. 2023. Developing smart air purifier control strategies for better IAQ and energy efficiency using reinforcement learning. *Building and Environment* 242 (2023), 110556.

[173] Wei Shen, Xiaonan He, Chuheng Zhang, Qiang Ni, Wanchun Dou, and Yan Wang. 2020. Auxiliary-task based deep reinforcement learning for participant selection problem in mobile crowdsourcing. In *CIKM*. 1355–1364.

[174] Dingyuan Shi, Yongxin Tong, Zimu Zhou, Bingchen Song, Weifeng Lv, and Qiang Yang. 2021. Learning to assign: Towards fair task assignment in large-scale ride hailing. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 3549–3557.

[175] Zhenyu Shou and Xuan Di. 2020. Reward design for driver repositioning using multi-agent reinforcement learning. *Transportation research part C: emerging technologies* 119 (2020), 102738.

[176] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. 2014. Deterministic policy gradient algorithms. In *International conference on machine learning*. PMLR, 387–395.

[177] Mohammad Javad Emami Skardi, Reza Kerachian, and Ali Abdolhay. 2020. Water and treated wastewater allocation in urban areas considering social attachments. *Journal of Hydrology* 585 (2020), 124757.

[178] Jaein Song, Yun Ji Cho, Min Hee Kang, and Kee Yeon Hwang. 2020. An application of reinforced learning-based dynamic pricing for improvement of ridesharing platform service in Seoul. *Electronics* 9, 11 (2020), 1818.

[179] Xuan Song, Haoran Zhang, Rajendra Akerkar, Huawei Huang, Song Guo, Lei Zhong, Yusheng Ji, Andreas L Opdahl, Hemant Purohit, André Skupin, et al. 2020. Big data and emergency management: concepts, methodologies, and applications. *IEEE Transactions on Big Data* 8, 2 (2020).

[180] Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, and Ryosuke Shibasaki. 2014. Intelligent system for urban emergency management during large-scale disaster. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 28.

[181] Andrew J Stier, Marc G Berman, and Luís MA Bettencourt. 2021. Early pandemic COVID-19 case growth rates increase with city size. *npj Urban Sustainability* 1, 1 (2021), 31.

[182] Hongyuan Su, Yu Zheng, Jingtao Ding, Depeng Jin, and Yong Li. 2019. Large-scale Urban Facility Location Selection with Knowledge-informed Reinforcement Learning. *arXiv preprint arXiv:2409.01588* (2019).

[183] Hongyuan Su, Yu Zheng, Jingtao Ding, Depeng Jin, and Yong Li. 2024. MetroGNN: Metro Network Expansion with Reinforcement Learning. In *Companion Proceedings of the ACM on Web Conference 2024*. 650–653.

[184] Nasrin Sultana, Jeffrey Chan, Tabinda Sarwar, and AK Qin. 2022. Learning to optimise general TSP instances. *International Journal of Machine Learning and Cybernetics* 13, 8 (2022), 2213–2228.

[185] Jiahui Sun, Haiming Jin, Zhaoxing Yang, and Lu Su. 2024. Optimizing Long-Term Efficiency and Fairness in Ride-Hailing under Budget Constraint via Joint Order Dispatching and Driver Repositioning. *IEEE Transactions on Knowledge and Data Engineering* (2024).

[186] Jiahui Sun, Haiming Jin, Zhaoxing Yang, Lu Su, and Xinbing Wang. 2022. Optimizing Long-Term Efficiency and Fairness in Ride-Hailing via Joint Order Dispatching and Driver Repositioning. In *KDD*. 3950–3960.

[187] Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. 1999. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems* 12 (1999).

[188] Jing Tan, Hui Liu, Yanfei Li, Shi Yin, and Chengqing Yu. 2022. A new ensemble spatio-temporal PM2. 5 prediction method based on graph attention recursive networks and reinforcement learning. *Chaos, Solitons & Fractals* 162 (2022), 112405.

[189] Xiaocheng Tang, Zhiwei Qin, Fan Zhang, Zhaodong Wang, Zhe Xu, Yintai Ma, Hongtu Zhu, and Jieping Ye. 2019. A deep value-network based approach for multi-driver order dispatching. In *KDD*. 1780–1790.

[190] Xiaocheng Tang, Fan Zhang, Zhiwei Qin, Yansheng Wang, Dingyuan Shi, Bingchen Song, Yongxin Tong, Hongtu Zhu, and Jieping Ye. 2021. Value function is all you need: A unified learning framework for ride hailing platforms. In *KDD*. 3605–3615.

[191] Liang Tian, Xuefei Li, Fei Qi, Qian-Yuan Tang, Viola Tang, Jiang Liu, Zhiyuan Li, Xingye Cheng, Xuanxuan Li, Yingchen Shi, et al. 2021. Harnessing peak transmission around symptom onset for non-pharmaceutical intervention and containment of the COVID-19 pandemic. *Nature communications* 12, 1 (2021), 1147.

[192] Wenchong Tian, Guangtao Fu, Kunlun Xin, Zhiyu Zhang, and Zhenliang Liao. 2024. Improving the interpretability of deep reinforcement learning in urban drainage system operation. *Water Research* 249 (2024), 120912.

[193] Yongxin Tong, Dingyuan Shi, Yi Xu, Weifeng Lv, Zhiwei Qin, and Xiaocheng Tang. 2021. Combinatorial optimization meets reinforcement learning: Effective taxi order dispatching at large-scale. *IEEE Transactions on Knowledge and Data Engineering* 35, 10 (2021), 9812–9823.

[194] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

[195] Fouad Trad and Salah El Falou. 2022. Towards using deep reinforcement learning for better COVID-19 vaccine distribution strategies. In *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*. IEEE, 7–12.

[196] Neşe Tüfekci, Nüket Sivri, and İsmail Toroz. 2007. Pollutants of textile industry wastewater and assessment of its discharge limits by water quality standards. *Turkish Journal of Fisheries and Aquatic Sciences* 7, 2 (2007).

[197] Berkay Turan, Ramtin Pedarsani, and Mahnoosh Alizadeh. 2020. Dynamic pricing and fleet management for electric autonomous mobility on demand systems. *Transportation Research Part C: Emerging Technologies* 121 (2020), 102829.

[198] Bernard Turnock. 2012. *Public health*. Jones & Bartlett Publishers.

[199] Elise Van der Pol and Frans A Oliehoek. 2016. Coordinated deep reinforcement learners for traffic light control. *Proceedings of learning, inference and control of multi-agent systems (at NIPS 2016)* 8 (2016), 21–38.

[200] Truong Van Nguyen, Jie Zhang, Li Zhou, Meng Meng, and Yong He. 2020. A data-driven optimization of large-scale dry port location using the hybrid approach of data mining and complex network theory. *Transportation Research Part E: Logistics and Transportation Review* 134 (2020).

[201] P Aarne Vesilind, J Jeffrey Peirce, and Ruth F Weiner. 2013. *Environmental pollution and control*. Elsevier.

[202] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. *Advances in neural information processing systems* 28 (2015).

[203] Leonie von Wahl, Nicolas Tempelmeier, Ashutosh Sao, and Elena Demidova. 2022. Reinforcement Learning-based Placement of Charging Stations in Urban Road Networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3992–4000.

[204] Runzhe Wan, Xinyu Zhang, and Rui Song. 2021. Multi-Objective Model-based Reinforcement Learning for Infectious Disease Control. In *KDD*, Feida Zhu, Beng Chin Ooi, and Chunyan Miao (Eds.). ACM, 1634–1644.

[205] Arthur Wang and Berkay Turan. 2022. Multi-Agent Renforcement Learning for Dynamic Pricing and Fleet Management in Autonomous Mobility-On-Demand Systems. International Foundation for Telemetering.

[206] Bowen Wang, Yanjing Sun, Trung Q. Duong, Long Dinh Nguyen, and Lajos Hanzo. 2020. Risk-Aware Identification of Highly Suspected COVID-19 Cases in Social IoT: A Joint Graph Theory and Reinforcement Learning Approach. *IEEE Access* 8 (2020), 115655–115661.

[207] Dongjie Wang, Yanjie Fu, Kunpeng Liu, Fanglan Chen, Pengyang Wang, and Chang-Tien Lu. 2023. Automated urban planning for reimagining city configuration via adversarial learning: quantification, generation, and evaluation. *ACM Transactions on Spatial Algorithms and Systems* 9, 1 (2023).

[208] Dongjie Wang, Yanjie Fu, Pengyang Wang, Bo Huang, and Chang-Tien Lu. 2020. Reimagining city configuration: Automated urban planning via adversarial learning. In *Proceedings of the 28th international conference on advances in geographic information systems*. 497–506.

[209] Dongjie Wang, Lingfei Wu, Denghui Zhang, Jingbo Zhou, Leilei Sun, and Yanjie Fu. 2023. Human-instructed deep hierarchical generative learning for automated urban planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 4660–4667.

[210] Huandong Wang, Qiaohong Yu, Yu Liu, Depeng Jin, and Yong Li. 2021. Spatio-temporal urban knowledge graph enabled mobility prediction. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 4 (2021), 1–24.

[211] Jing Wang and Filip Biljecki. 2022. Unsupervised machine learning in urban studies: A systematic review of applications. *Cities* 129 (2022), 103925.

[212] Jianhong Wang, Wangkun Xu, Yunjie Gu, Wenbin Song, and Tim C Green. 2021. Multi-agent reinforcement learning for active voltage control on power distribution networks. *Advances in Neural Information Processing Systems* 34 (2021), 3271–3284.

[213] Jiguang Wang, Yilun Zhang, Xinjie Xing, Yuanzhu Zhan, Wai Kin Victor Chan, and Sunil Tiwari. 2022. A data-driven system for cooperative-bus route planning based on generative adversarial network and metric learning. *Annals of Operations Research* (2022), 1–27.

[214] Runzhong Wang, Zhigang Hua, Gan Liu, Jiayi Zhang, Junchi Yan, Feng Qi, Shuang Yang, Jun Zhou, and Xiaokang Yang. 2021. A bi-level framework for learning to solve combinatorial optimization on graphs. *Advances in Neural Information Processing Systems* 34 (2021), 21453–21466.

[215] Runzhong Wang, Li Shen, Yiting Chen, Xiaokang Yang, Dacheng Tao, and Junchi Yan. 2023. Towards one-shot neural combinatorial solvers: Theoretical and empirical notes on the cardinality-constrained case. In *The Eleventh International Conference on Learning Representations*.

[216] Xiaoqiang Wang, Liangjun Ke, Zhimin Qiao, and Xinghua Chai. 2020. Large-scale traffic signal control using a novel multiagent reinforcement learning. *IEEE transactions on cybernetics* 51, 1 (2020), 174–187.

[217] Yunqian Wang. 2018. Optimization on fire station location selection for fire emergency vehicles using K-means algorithm. In *2018 3rd International Conference on Advances in Materials, Mechatronics and Civil Engineering (ICAMMCE 2018)*. Atlantis Press, 323–333.

[218] Yiheng Wang, Hexi Jin, and Guanjie Zheng. 2022. CTRL: Cooperative Traffic Tolling via Reinforcement Learning. In *CIKM*. 3545–3554.

[219] Yansheng Wang, Yongxin Tong, Zimu Zhou, Ziyao Ren, Yi Xu, Guobin Wu, and Weifeng Lv. 2022. Fed-LTD: Towards Cross-Platform Ride Hailing via Federated Learning to Dispatch. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4079–4089.

[220] Yanan Wang, Tong Xu, Xin Niu, Chang Tan, Enhong Chen, and Hui Xiong. 2020. STMARL: A spatio-temporal multi-agent reinforcement learning approach for cooperative traffic light control. *IEEE Transactions on Mobile Computing* 21, 6 (2020), 2228–2242.

[221] Zhaodong Wang, Zhiwei Qin, Xiaocheng Tang, Jieping Ye, and Hongtu Zhu. 2018. Deep reinforcement learning with knowledge transfer for online rides order dispatching. In *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 617–626.

[222] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8, 3 (1992), 279–292.

[223] Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, Kai Xu, and Zhenhui Li. 2019. Presslight: Learning max pressure control to coordinate traffic signals in arterial network. In *KDD*. 1290–1298.

[224] Hua Wei, Nan Xu, Huichu Zhang, Guanjie Zheng, Xinshi Zang, Chacha Chen, Weinan Zhang, Yanmin Zhu, Kai Xu, and Zhenhui Li. 2019. Colight: Learning network-level cooperation for traffic signal control. In *CIKM*. 1913–1922.

[225] Honghao Wei, Zixian Yang, Xin Liu, Zhiwei Qin, Xiaocheng Tang, and Lei Ying. 2023. A Reinforcement Learning and Prediction-Based Lookahead Policy for Vehicle Repositioning in Online Ride-Hailing Systems. *IEEE Transactions on Intelligent Transportation Systems* (2023).

[226] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. 2021. Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. *ACM SIGKDD Explorations Newsletter* 22, 2 (2021), 12–18.

[227] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. 2018. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *KDD*. 2496–2505.

[228] Yu Wei, Minjia Mao, Xi Zhao, Jianhua Zou, and Ping An. 2020. City metro network expansion with reinforcement learning. In *KDD*. 2646–2656.

[229] Ying Wen, Ziyu Wan, Ming Zhou, Shufang Hou, Zhe Cao, Chenyang Le, Jingxiao Chen, Zheng Tian, Weinan Zhang, and Jun Wang. 2023. On Realization of Intelligent Decision Making in the Real World: A Foundation Decision Model Perspective. *CAAI Artificial Intelligence Research* 2 (2023).

[230] Di Weng, Chengbo Zheng, Zikun Deng, Mingze Ma, Jie Bao, Yu Zheng, Mingliang Xu, and Yingcai Wu. 2020. Towards better bus networks: A visual analytics approach. *IEEE transactions on visualization and computer graphics* 27, 2 (2020), 817–827.

[231] Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems* 2, 1-3 (1987).

[232] Chaohao Wu, Tong Qiao, Hongjun Qiu, Benyun Shi, and Qing Bao. 2021. Individualism or Collectivism: A Reinforcement Learning Mechanism for Vaccination Decisions. *Inf.* 12, 2 (2021), 66. https://doi.org/10.3390/info12020066

[233] Tony Wu, Anthony D Joseph, and Stuart J Russell. 2016. Automated pricing agents in the on-demand economy. *University of California at Berkeley: Berkeley, CA, USA* (2016).

[234] Tong Wu, Pan Zhou, Kai Liu, Yali Yuan, Xiumin Wang, Huawei Huang, and Dapeng Oliver Wu. 2020. Multi-agent deep reinforcement learning for urban traffic light control in vehicular networks. *IEEE Transactions on Vehicular Technology* 69, 8 (2020), 8243–8256.

[235] Yaoxin Wu, Wen Song, Zhiguang Cao, Jie Zhang, and Andrew Lim. 2021. Learning improvement heuristics for solving routing problems. *IEEE transactions on neural networks and learning systems* 33, 9 (2021), 5057–5069.

[236] Liang Xin, Wen Song, Zhiguang Cao, and Jie Zhang. 2021. Multi-decoder attention model with embedding glimpse for solving vehicle routing problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 12042–12049.

[237] Xianhao Xu, Yaohan Shen, Wanying Amanda Chen, Yeming Gong, and Hongwei Wang. 2021. Data-driven decision and analytics of collection and delivery point location problems for online retailers. *Omega* 100 (2021), 102280.

[238] Zhe Xu, Zhixin Li, Qingwen Guan, Dingshui Zhang, Qiang Li, Junxiao Nan, Chunyang Liu, Wei Bian, and Jieping Ye. 2018. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. In *KDD*. 905–913.

[239] Chiwei Yan, Helin Zhu, Nikita Korolko, and Dawn Woodard. 2020. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)* 67, 8 (2020), 705–724.

[240] Hai Yang and Xiaoning Zhang. 2003. Optimal toll design in second-best link-based congestion pricing. *Transportation Research Record* 1857, 1 (2003), 85–92.

[241] Qinmin Yang, Weiwei Cao, Wenchao Meng, and Jennie Si. 2021. Reinforcement-learning-based tracking control of waste water treatment process under realistic system conditions and control performance requirements. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52, 8 (2021).

[242] Zhou Yang, Long Nguyen, Jiazhen Zhu, Zhenhe Pan, Jia Li, and Fang Jin. 2020. Coordinating disaster emergency response with heuristic reinforcement learning. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 565–572.

[243] Meng You, Yiyong Xiao, Siyue Zhang, Pei Yang, and Shenghan Zhou. 2019. Optimal mathematical programming for the warehouse location problem with Euclidean distance linearization. *Computers & Industrial Engineering* 136 (2019), 70–79.

[244] Chengqing Yu, Guangxi Yan, Kaiyi Ruan, Xinwei Liu, Chengming Yu, and Xiwei Mi. 2023. An ensemble convolutional reinforcement learning gate network for metro station PM2. 5 forecasting. *Stochastic Environmental Research and Risk Assessment* (2023), 1–16.

[245] Liang Yu, Shuqi Qin, Meng Zhang, Chao Shen, Tao Jiang, and Xiaohong Guan. 2021. A review of deep reinforcement learning for smart building energy management. *IEEE Internet of Things Journal* 8, 15 (2021), 12046–12063.

[246] Yuan Yuan, Jingtao Ding, Jie Feng, Depeng Jin, and Yong Li. 2024. UniST: A Prompt-Empowered Universal Model for Urban Spatio-Temporal Prediction. In *KDD*.

[247] Xinshi Zang, Huaxiu Yao, Guanjie Zheng, Nan Xu, Kai Xu, and Zhenhui Li. 2020. Metalight: Value-based meta-reinforcement learning for traffic signal control. In *AAAI*, Vol. 34.

[248] Hongbo Zhang, Guang Wang, Xu Wang, Zhengyang Zhou, Chen Zhang, Zheng Dong, and Yang Wang. 2024. NondBREM: Nondeterministic Offline Reinforcement Learning for Large-Scale Order Dispatching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 401–409.

[249] Jun Zhang, Depeng Jin, and Yong Li. 2022. Mirage: an efficient and extensible city simulation framework (systems paper). In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*. 1–4.

[250] Ke Zhang, Fang He, Zhengchao Zhang, Xi Lin, and Meng Li. 2020. Multi-vehicle routing problems with soft time windows: A multi-agent reinforcement learning approach. *Transportation Research Part C: Emerging Technologies* 121 (2020), 102861.

[251] Lingyu Zhang, Tao Hu, Yue Min, Guobin Wu, Junying Zhang, Pengcheng Feng, Pinghua Gong, and Jieping Ye. 2017. A taxi order dispatch model based on combinatorial optimization. In *KDD*. 2151–2159.

[252] Qiang Zhang, Shi Qiang Liu, and Andrea D'Ariano. 2023. Bi-objective bi-level optimization for integrating lane-level closure and reversal in redesigning transportation networks. *Operational Research* 23, 2 (2023), 23.

[253] Weijia Zhang, Hao Liu, Jindong Han, Yong Ge, and Hui Xiong. 2022. Multi-agent graph convolutional reinforcement learning for dynamic electric vehicle charging pricing. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 2471–2481.

[254] Yongping Zhang, Diao Lin, and Zhifu Mi. 2019. Electric fence planning for dockless bike-sharing services. *Journal of cleaner production* 206 (2019).

[255] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. 2020. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 737–744.

[256] Xianli Zhao and Guixin Wang. 2022. Deep Q networks-based optimization of emergency resource scheduling for urban public health events. *Neural Computing and Applications* (2022), 1–10.

[257] Bolong Zheng, Lingfeng Ming, Qi Hu, Zhipeng Lü, Guanfeng Liu, and Xiaofang Zhou. 2022. Supply-demand-aware deep reinforcement learning for dynamic fleet management. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 3 (2022), 1–19.

[258] Guanjie Zheng, Yuanhao Xiong, Xinshi Zang, Jie Feng, Hua Wei, Huichu Zhang, Yong Li, Kai Xu, and Zhenhui Li. 2019. Learning phase competition for traffic signal control. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1963–1972.

[259] Yu Zheng, Yuming Lin, Liang Zhao, Tinghai Wu, Depeng Jin, and Yong Li. 2023. Spatial planning of urban communities via deep reinforcement learning. *Nature Computational Science* 3, 9 (2023), 748–762.

[260] Yu Zheng, Hongyuan Su, Jingtao Ding, Depeng Jin, and Yong Li. 2023. Road planning for slums via deep reinforcement learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5695–5706.

[261] Zhu Zhongming, Lu Linong, Yao Xiaona, Liu Wei, et al. 2020. World Cities Report 2020: The value of sustainable urbanization. (2020).

[262] Bojian Zhou, Michiel Bliemer, Hai Yang, and Jie He. 2015. A trial-and-error congestion pricing scheme for networks with elastic demand and link capacity constraints. *Transportation Research Part B: Methodological* 72 (2015), 77–92.

[263] Jianan Zhou, Yaoxin Wu, Wen Song, Zhiguang Cao, and Jie Zhang. 2023. Towards omni-generalizable neural methods for vehicle routing problems. In *International Conference on Machine Learning*. PMLR, 42769–42789.

[264] Ming Zhou, Jiarui Jin, Weinan Zhang, Zhiwei Qin, Yan Jiao, Chenxi Wang, Guobin Wu, Yong Yu, and Jieping Ye. 2019. Multi-agent reinforcement learning for order-dispatching via order-vehicle distribution matching. In *CIKM*. 2645–2653.

[265] Zhengqiu Zhu, Bin Chen, Yong Zhao, and Yatai Ji. 2021. Multi-sensing paradigm based urban air quality monitoring and hazardous gas source analyzing: a review. *Journal of Safety Science and Resilience* 2, 3 (2021), 131–145.

[266] Kai Zong and Cuicui Luo. 2022. Reinforcement learning based framework for COVID-19 resource allocation. *Comput. Ind. Eng.* 167 (2022), 107960.

[267] Zefang Zong, Tao Feng, Tong Xia, Depeng Jin, and Yong Li. 2021. Deep Reinforcement Learning for Demand Driven Services in Logistics and Transportation Systems: A Survey. *arXiv preprint arXiv:2108.04462* (2021).

[268] Zefang Zong, Hansen Wang, Jingwei Wang, Meng Zheng, and Yong Li. 2022. Rbg: Hierarchically solving large-scale routing problems in logistic systems via reinforcement learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4648–4658.

[269] Zefang Zong, Meng Zheng, Yong Li, and Depeng Jin. 2022. Mapdp: Cooperative multi-agent reinforcement learning to solve pickup and delivery problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 9980–9988.