Computers & Industrial Engineering 128 (2019) 851–860

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

An efficient integration of the genetic algorithm and the reinforcement learning for optimal deployment of the wireless charging electric tram system

Young Dae Ko

Department of Hotel and Tourism Management, College of Hospitality and Tourism, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea

ARTICLE INFO

Keywords:
Mathematical model
Wireless charging
Electric tram
Optimization
Genetic algorithm
Reinforcement learning
Integration level

ABSTRACT

To solve the two main issues such as long charging down time and expensive battery price of the electric vehicle systems, lots of researchers have studied about advanced charging and battery technology. In same time, the new electric vehicle system which applying the wireless charging technology is also regarded as alternative means of transportation because it can reduce both the long charging down time and the battery capacity. It is called as the wireless charging electric vehicle system and it can be supplied the electricity wirelessly from the wireless charging infrastructure buried under the road. In the similar reason, the wireless charging electric tram system is received lots of interests as the alternative mass transportation in urban and suburb area. To derive the optimal decision-making elements such as the required minimum battery capacity and the allocation of the wireless charging infrastructure, which have a trade-off relationship, a mathematical model is developed. In addition, the genetic algorithm integrated with the reinforcement learning are proposed to generate an optimal solution. By assuming the several integration situations of them, an efficient integration level of the genetic algorithm and the reinforcement learning are examined.

1. Introduction

Nowadays, the electric vehicle system is regarded as an alternative means of transportation because it can reduce the environmental pollution comparing the conventional internal combustion engine vehicle system. In particular, the electric vehicle system can improve the air pollution which is one of the main factors threaten human life. In addition, because the air pollution in Asia is becoming more serious as the demand for heating is increasing due to rapid urbanization in China and India, many countries are making efforts to develop and introduce electric vehicle systems as a way to reduce the air pollution. Recently, the Seoul Metropolitan Government of Korea newly regulates the "Free Mass Transportation Law" to reduce the level of the air pollution by inducing the self-driving people to mass transportation when the level of the air pollution exceeding certain limit. In addition, the electric vehicles are still expensive because of their expensive battery prices. Therefore, lots of countries are paying a certain subsidy for the purchase of the electric vehicles to increase the market share of them under these circumstances (Sierzchula, Bakker, Maat & van Wee, 2014).

In case of means of mass transportation, the electric tram system is considered as an alternative way to achieve the reduction of air pollution (Geels, 2011). However, the electric tram system is not a new technology. Lots of traditional cities have been operating the conventional electric tram system until now from after the first application of it in eighteen or nineteen centuries. At that time, there were no advanced technology for storage of the electricity, it should be provided continuously through the overhead power supply line to operate the electric trams. Therefore, if the electric tram runs from the intersection as many routes, the overhead power supply lines can be installed as overlapping in the air. It can cause several problems in the aesthetical aspect, and can be the reason of safety accidents by the electricity issues such as short-circuit. That is, though the conventional electric tram has some advantage in the aspect of the reduction of air pollution, it is gradually fade out and is replaced the other means of transportation. As a result, the overhead power supply lines tend to install in underground and the electric tram system changed as the bus system with internal combustion engine and the subway system with power supply line in last several decades (Moreno et al., 2015).

Recently, to cope with the increasing air pollution, the electric vehicle system is getting highlighted as eco-friendly means of transportation again and among them, the electric tram system gets lots of interests because it can be utilized as a public mass transportation system and easy to install in urban and suburb area. However, because the several problems caused by the overhead power supply line can not

E-mail address: youngdae.ko@sejong.ac.kr.

https://doi.org/10.1016/j.cie.2018.10.045

Available online 24 October 2018

solved with the conventional way, the battery-type electric tram system is also regarded as one of the alternatives. Then, the overhead power supply line can be discarded by applying the battery-type electric tram system, but still, the battery itself is very environmentally harmful material. Therefore, the practitioners start to have an interest about the wireless charging technology for the electric tram system to reduce the required battery capacity. As a result, nowadays, the wireless charging electric tram system is regarded as alternative means of mass transportation in urban and suburb area.

The wireless charging electric tram system is consisted of the electric tram with both relatively small battery and wireless pick-up device, and the wireless charging infrastructure such as inductive cable for transferring the electricity wirelessly based on the electromagnetic induction phenomenon and inverter for providing the electricity to inductive cable from external power sources (Ko & Jang, 2013). In case of the battery-type electric tram system, it required relatively big battery and it takes lots of time to charge after the operation. However, the wireless charging electric tram system can reduce the battery capacity because it can be supplied the electricity wirelessly when it is operated on the wireless charging infrastructure even though it is moving. Therefore, it can also reduce the charging time after the operation and the environmentally harmful battery capacity. This system is already success in commercialization by the several companies such as Bombardier and so on and Fig. 1 is about “Primove” developed by Bombardier.

In the wireless charging electric tram system, the required battery capacity and the allocation of the wireless charging infrastructure are expressed as the trade-off relationship. That is, when the wireless charging infrastructure is installed under the all railway, then the wireless charging electric tram do not require to equip the battery because it can be supplied the electricity wirelessly always from the wireless charging infrastructure. In this case, there is needed huge investment only at allocating the wireless charging infrastructure like a conventional electric tram system. On the contrary, when the wireless charging infrastructure is not installed at all, then the wireless charging electric tram requires the huge size battery for operating without the wireless charging like a battery-type electric tram. Therefore, to minimize the total investment cost, the optimal decision-making technique should be applied to derive an optimal battery capacity and optimal allocation of the wireless charging infrastructure (Ko, Jang & Lee, 2015).

In this study, a mathematical model based optimization is applied to reflect the features of the various kinds of the wireless charging electric tram system and to generate an optimal required battery capacity and allocation of the wireless charging infrastructure. And then, the optimal solution would be derived by adopting the appropriate solution generation algorithm. Fundamentally, the result from CPLEX can be initially obtained and considered as global optimal solution. In addition, the genetic algorithm would be developed to derive an optimal or a near optimal solution. Moreover, the reinforcement learning would be integrated at the genetic algorithm with the stage of (i) initial solution generation, (ii) crossover and/or mutation, and (iii) both of them. The results will be compared and evaluated each other in terms of accuracy. The reinforcement learning is an area of machine learning. A method inspired by behavioral psychology, in which an agent defined in an environment recognizes the current state and selects a behavior or sequence of actions that maximizes compensation among the selectable behaviors. It can significantly improve the quality of solution derived by genetic algorithm when it is integrated at crossover and/or mutation processes of genetic algorithm (Liu and Zeng, 2009).

This paper is organized as follows. In Section 2, it is presented about related previous studies as the literature review. The description of the optimal design for the wireless charging electric tram and the mathematical model for it to minimize total investment cost are proposed and developed in Section 3. The introduction about the reinforcement learning and the genetic algorithm integrated the reinforcement learning are described in Section 4 while the numerical examples are presented in Section 5. Finally, findings and insights from this study are announced as conclusions in Section 6.

2. Literature review

The wireless charging technology is started to apply for various kinds of the electric devices due to its conveniences. It is based on the electro-magnetic induction phenomena and the electricity can transfer wirelessly from its resources to other electric circuit (Hui, 2013). This wireless charging technology also can be applied the electric vehicle system both stationary and dynamic charge cases. That is, the wireless charging electric vehicle can receive the electricity wirelessly when it is parked at the wireless charging area without connecting the wire like conventional electric vehicle system. In addition, the wireless charging electric vehicle can be provided the electricity wirelessly when it is moving on the road where the wireless charging infrastructure is buried. The former is often called stationary wireless charging, and the latter is called dynamic wireless charging. In general, the wireless charging electric vehicle system can be applied both of them (Lukic & Pantic, 2013). In this section, the several important research with managerial aspects of the wireless charging electric vehicle system is mainly discussed because though there are lots of previous studied with technical aspects, but it is not the ours main concern (Fisher, Farley,

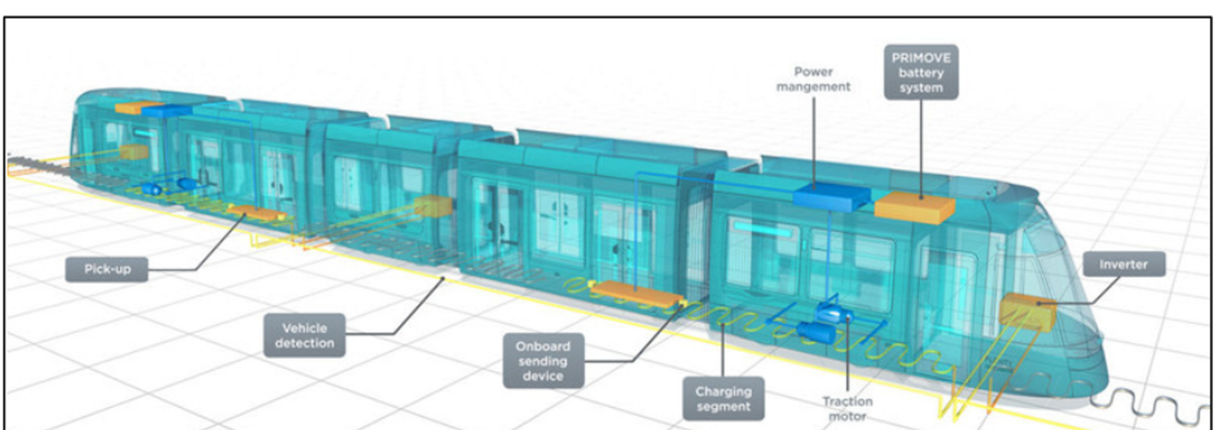

Fig. 1. “Primove Tram” from Bombardier.
Source: primove.bombardier.com

Gao, Bai & Tse, 2014). Within our knowledge, Ko and Jang (2013) introduced firstly about the optimal system design of the wireless charging electric vehicle system. They tried to derive the minimum required battery capacity equipped at the wireless charging electric vehicle and the location and length of the wireless charging infrastructure while pursuing to minimize the total investment cost. They developed a mathematical model considering the features of the wireless charging electric vehicle system and proposed appropriate meta-heuristic algorithm, particle swarm optimization (PSO) to generate an optimal or near-optimal solution. This research is extended by reflecting the real circumstances such as the nonlinear cost structure of the wireless charging infrastructure, the features of the closed environment, the characteristics of the battery life according to the charging level variation and so on (Jang, Jeong & Ko, 2015; Jeong, Jang & Kum, 2014; Ko et al., 2015). In addition, Hwang, Jang, Ko, and Lee (2017) performed the research about an optimal system design of the wireless charging electric vehicle system with complex multiple routes. The generalization of the multiple routes is hard to apply because the allocation of the wireless charging infrastructure at the complex multiple routes such as intersection is so complex. For that, the sophisticated mathematical model was developed, and appropriate solution procedure also proposed to generate an optimal battery capacity and the location and length of the wireless charging infrastructure. Liu and Song (2017) conducted similar research. Overall approach was very similar with that of Hwang et al. (2017), but they dealt with the uncertainty of energy consumption and travel time by using the robust optimization. Moreover, Sarker et al. (2016) proposed a system to balance the State of Charge (called BSoC) among the electric vehicles when a large number of electric vehicles passing a charging lane in same time. Also, they introduced the fog computing center which collecting the information from electric vehicles and schedules for the power distribution. Recently, Vaz, Nandi, and Koylu (2016) studied about the operation profile of the wireless charging electric vehicle. They examined the two approaches to reach the desired speed using either a single acceleration value or multiple acceleration values. They developed a multi-objective optimization problem (MOOP) with minimization of acceleration duration while minimization of consumed energy. To solve the MOOP, they suggested the multi-objective genetic algorithms (MOGAs) and obtained the results (Pareto-optimal fronts) using suitable performance metrics. Recently, Ko and Jang (2018) suggested the genetic algorithm to derive an optimal operation profile at every time unit for the wireless charging electric tram system. They validated the proposed mathematical algorithm and the developed genetic algorithm through the numerical example considering the candidate site of Pangyo, Gyeonggi Province, Korea.

In this study, the genetic algorithm integrated with the reinforcement learning are proposed and examined to derive an optimal deployment design of the wireless charging electric tram system. There are several research applying the reinforcement learning to the genetic algorithm. In all studies, applying reinforcement learning to a common genetic algorithm resulted in a better solution in reasonable time. Mikami and Kakazu (1994) addressed the traffic planning problem which decides the optimal traffic signal control to maximize the total traffic volume. Each agent (signal) decided its control plan through the reinforcement learning algorithm individually to find an optimal local solution. After that, global traffic signal control was optimized with genetic algorithm to maximize the total profit of signal based on the individual control plan of each agent. Zhang and Dietterich (1995) developed a heuristic algorithm that applied reinforcement learning to a job-shop scheduling. Crites and Barto (1996) addressed the elevator dispatching problem. They proposed a heuristic algorithm with reinforcement learning and the experimental results were compared with other existing heuristics. Mabu, Hirasawa, and Hu (2007) suggested genetic network programming (GNP) and its extended algorithm GNP with reinforcement learning (GNPRL). Tile-world problem which is well-known test problems of agents was used to check and compare the performance of the GNP and GNPRL. The simulation results showed that GNPRL can create effective graph structure and get better results in dynamic environments. Chen, Mabu, Shimada, and Hirasawa (2009) addressed enhanced stock trading model using a genetic network programming (GNP) with Sarsa Learning which is one of the well-known reinforcement learning method. Sarsa obtained the stock price information from the importance index (IMX) and candlestick charts and then selected appropriate action; buying or selling. To validate the proposed algorithm, simulation was performed with 16 selected brands and the results was shown to be extremely well compared with other traditional stock trading methods. Liu and Zeng (2009) dealt with the traveling salesman problem (TSP). TSP is one of the typical NP-hard problems, so they developed reinforcement mutation genetic algorithm (RMGA) which is improved general genetic algorithm with reinforcement mutation to solve the TSP problem. Numerical examples of different sizes were tested and compared to validate the proposed algorithm in terms of running time and result quality. The RMGA obtained the almost optimal solution for every experimental test in reasonable time. Buzdalova and Buzdalov (2012) adopted reinforcement learning to increase efficiency of evolutionary algorithm by choosing auxiliary fitness function. The proposed model was validated with Royal Roads problems and it was shown to be efficient. There were many other studies which adopted reinforcement learning to a well-known existing solution algorithm or a heuristic algorithm. Recently, Zhang and Maringer (2016) proposed GA-RRL(Genetic Algorithm – Recurrent Reinforcement Learning) system for maximizing the trading profits. They validated the proposed model with the data of 180 S&P stocks. Samma, Lim, and Saleh (2016) developed RLMPSO(Reinforcement Learning-based Memetic Particle Swarm Optimization) which is combining Reinforcement Learning and Particle Swarm Optimization. Each particle can do 5 defined operations (exploration, convergence, high-jump, low-jump, and fine-tuning) based on the reinforcement learning algorithm. They proved the usefulness of the proposed algorithm by testing various problems. Alipour, Razavi, Derakhshi, and Balafar (2017) introduced hybrid algorithm with genetic algorithm and multi-agent reinforcement learning(MARL) to solve the TSP. Initial population of genetic algorithm was produced by the best solution of MARL. The 34 test problems ranging from 29 to 7397 cities were solved and the results were compared with the other algorithms. The proposed algorithm showed better performance in terms of quality of solution and the speed of computation. Cao, Lin, Zhou, and Huang (2018) dealt with the scheduling problem of semiconductor final testing. They suggested a cuckoo search algorithm with reinforcement learning for minimizing the makespan for the scheduling problem. Reinforcement Learning algorithm was used to obtain the appropriated parameter values to be used in the cuckoo search algorithm. The effectiveness of proposed algorithm was proved by comparison with existing methods.

Therefore, the main contribution of this study is examined the genetic algorithm integrated with the reinforcement learning of different levels for the optimal deployment design of the wireless charging electric tram system. That is, the reinforcement learning will be integrated at the genetic algorithm with the stage of (i) initial solution generation, (ii) crossover and/or mutation, and (iii) both of them. In case of the genetic algorithm, it is designed according to the features of problem nature of each study, Thus, it is hard to compare the performance of genetic algorithm integrated with reinforcement learning proposed in this study with the others from previous study, Therefore, in this study, the accuracy of the results according to the several integration stages of the reinforcement learning at the genetic algorithm is compared with the result from the CPLEX, a well-known commercial solver.

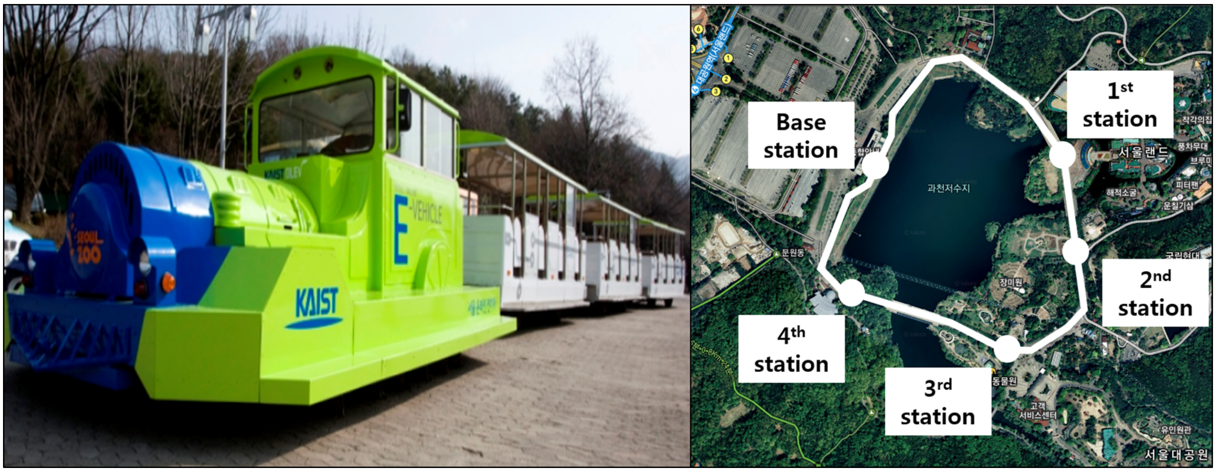

Fig. 2. The wireless charging electric tram and its operation route at Seoul National Grand Park.

3. Model development

3.1. Problem description

In this study, again, the main purpose is to validate and compare the results of the genetic algorithm integrated with the reinforcement learning for the optimal deployment design of the wireless charging electric tram system. Therefore, the relatively simple problem situation would be adopted from the research of Ko et al. (2015) for concentrating of the main purpose of this study. However, the decision variables such as the optimal battery capacity and the location of the wireless charging infrastructure can be derived a little bit differently because the regenerative braking is considered in this study.

There is the Seoul National Grand Park in Korea and it consists of an amusement park, zoo, art museum and so on around the Gwacheon Lake. People who want to use these facilities should park their vehicles at predetermined parking lots and go to their destinations by using the shuttle tram operated around those facilities (see Fig. 2). There are the Base station in huge parking lots, the 1st station in an amusement park, the 2nd station in an art museum, the 3rd station in a zoo, and the 4th station in a plaza. The shuttle tram starts its operation from the base station in a clockwise direction and it stops for a certain time to pick up and drop passengers at each station. Initially, it was operated by an internal combustion engine tram, but since 2009, a wireless charging electric tram has been in operation to reduce the emission of environmental pollution materials.

Therefore, it is required to decide the optimal battery capacity and the location of the wireless charging infrastructure at overall route to minimize the total investment cost because the battery capacity and the wireless charging infrastructure are two main elements of total investment cost. However, in here, the battery capacity and the wireless charging infrastructure have trade-off relationship. That is, when the battery capacity is decided relatively huge, then the length of total wireless charging infrastructure is determined relatively small and vice versa. For that, it is assumed that the overall route can be divided as I number of separate parts, which is called as segments, to decide the location of the wireless charging infrastructure at the segment level. It is further assumed that the cost structure of the battery and the wireless charging infrastructure are known and the operation profile such as an acceleration and a maximum operation velocity of the wireless charging electric tram are also known. Therefore, the battery consumption by the operation, and the battery charging by both the wireless charging and the regenerative braking are known parameters at every i^{th} segment.

3.2. Notations

To develop a mathematical model by reflecting the features of the proposed problem situation, several notations are defined as follows.

Index	
i	Set of segments, overall route is divided by I number of segments (i = 1, 2, 3, …, I)
Parameters	
c_{tram}	Unit purchasing cost of the wireless charging electric tram without battery [$]
$c_{battery}$	Unit purchasing cost of battery per kWh [$/kWh]
$c_{inverter}$	Unit purchasing cost of the inverter [$]
c_{cable}	Unit purchasing cost of the inductive cable per meter [$/meter]
N	The number of the wireless charging electric tram [unit]
$l(i)$	Length of i^{th} segment [meter]
L	Length of overall route [meter]
α_{max}	Maximum utilization ratio of battery, $0 \le \alpha \le 1$
α_{min}	Minimum utilization ratio of battery, $0 \le \beta \le 1$
I_{max}	Maximum utilization level of battery [kWh]
I_{min}	Minimum utilization level of battery [kWh]
$I(i)$	Battery charging level after passing i^{th} segment [kWh]
$d(i)$	An amount of consumed electricity at i^{th} segment [kWh]
$s(i)$	An amount of supplied electricity by the wireless charging at i^{th} segment [kWh]
$r(i)$	An amount of supplied electricity by the regenerative braking at i^{th} segment [kWh]
Decision variables	
$x(i)$	Binary decision variable which it has value of 1 when the inductive cable is allocated at i^{th} segment, otherwise, it has value of 0
$y(i)$	Binary decision variable which it has value of 1 when the inverter is allocated at i^{th} segment, otherwise, it has value of 0
$q_{battery}$	Minimum required battery capacity for the wireless charging electric tram [kWh]

3.3. Mathematical model

A mathematical model consists of both the objective function and the constraints. In this study, the objective function is to minimize the total investment cost.

$$\text{Minimize} \quad (c_{tram} + c_{battery}\cdot q_{battery})\cdot N + c_{inverter}\cdot \sum_{i=1}^{I} y(i) + c_{cable}\cdot \sum_{i=1}^{I} \{l(i)\cdot x(i)\} \tag{1}$$

The objective function is prepared with the sum of the unit purchasing cost of the wireless charging electric tram with battery multiplies purchasing quantity, the overall inverter cost, and the overall inductive cable cost. Please note that one separate inductive cable is required the one inverter regardless its length.

Subject to $$I_{max} = \alpha_{max}\cdot q_{battery} \tag{2}$$

$$I_{min} = \alpha_{min}\cdot q_{battery} \tag{3}$$

$$I_{min} \leqslant I(i) \leqslant I_{max}, \quad \forall\ i \tag{4}$$

$$I(0) = I_{max} \tag{5}$$

$$I(i-1) - d(i) + r(i) + s(i)\cdot x(i) \geqslant I(i), \quad \forall\ i \tag{6}$$

$$x(0) = 0 \tag{7}$$

$$y(i) \geqslant x(i) - x(i-1), \quad \forall\ i \tag{8}$$

$$q_{battery} \geqslant 0 \tag{9}$$

$$x(i),\ \ y(i) \in \{0,\ \ 1\}, \quad \forall\ i \tag{10}$$

Eq. (2) is about the maximum utilization level of the battery equipped at the wireless charging electric tram while Eq. (3) is for the minimum utilization level of it. Eq. (4) describes that the potential variable range of the battery charging level during the operation of the wireless charging electric tram while Eq. (5) represents the initial condition of the battery charging level of it. Eq. (6) depicts the variation of the battery charging level passing after the i^{th} segment considering the battery consumption and the battery charging by both the wireless charging and the regenerative braking. Eq. (7) means the initial condition for whether the inductive cable is located or not while Eq. (8) presents the condition for the required inverter according to the allocation of the inductive cable at i^{th} and $(i-1)^{th}$ segments. Eq. (9) is for the nonnegativity of the battery capacity and Eq. (10) declares the 0–1 binary integer decision variables for the inverter and the inductive cable.

4. Solution procedure

In this study, the genetic algorithm integrated with the reinforcement learning are proposed to generate an optimal solution for the design elements of the wireless charging electric tram system. Fundamentally, the exact result is derived through the CPLEX, a commercial optimal solution generation software and it is known that it can generate an optimal solution for the linear and integer programming. And then, several integration types of reinforcement learning at conventional genetic algorithm are introduced and applied to solve the given problem and finally the results will be compared and analyzed at Section 5.

4.1. Concept of the reinforcement learning

The reinforcement learning (RL) is an area of machine learning where agents learn how to behave in an environment and then perform an action among selectable behaviors that maximizes rewards (van Otterlo & Wiering, 2012). Reinforcement learning consists of three components (Kaelbling, Littman & Mooreet, 1996).

- a discrete set of environment states, S
- a discrete set of agent actions, A
- a set of scalar reinforcement signals; typically {0, 1}, or the real numbers.

At each time t, the agents have their own states s_t and possible actions $\underline{A}(s_t)$. The agent performs an action among the candidate actions and receives a new state s_{t+1} and reward r_{t+1} from the environment. Based on this interaction, the RL agent develops an optimal policy that maximizes the cumulative rewards.

Liu and Zeng (2009) solved the TSP problem by applying reinforcement learning to the genetic algorithm. They updated the cumulative reward between current city t and next city q, RQ(t, q), according to the following function where α is the learning rate, γ is the discount rate ($0 < \alpha, \gamma < 1$) and $J(t)$ denotes the unvisited cities when the current city is t.

$$RQ(t,q) = (1-\alpha)\cdot RQ(r,q) + \alpha\,[r(t,q) + \gamma \max_{z\in J(t)} RQ(q,z)] \tag{11}$$

RL is generally effective in problems where it is analytically unavailable or difficult to obtain an optimal solution. Many applications with reinforcement learning have been studied (Doya, 2000); chess game (Lai, 2015), backgammon (Tesauro, 1995), elevator dispatching problem (Crites & Barto, 1996, 1998), stock trading (Lee, 2001; Tan, Quek & Cheng, 2011), robot control (Matarić, 1997), etc.

An efficient solution for the design of the wireless charging electric tram system can be derived by applying only the reinforcement learning. However, it is not guarantee an optimal solution while the CPLEX can always generates an optimal solution mathematically using the simplex method. The results from both the CPLEX and the reinforcement learning will be compared in the numerical examples section.

4.2. Genetic algorithm

The genetic algorithm is one of the evolutionary algorithm introduced by Holland (1975). Generally, it is known that the genetic algorithm can generate an optimal or a near-optimal solution and it is also efficient meta-heuristic algorithm for deriving the good design of the wireless charging electric tram system (Jang et al., 2015; Ko et al., 2015). In this study, the basic approach of the genetic algorithm is similar from the previous study of Ko et al. (2015). However, the reinforcement learning will be applied either the initial solution generation step and/or the crossover/mutation step for more powerful application of the genetic algorithm.

4.2.1. Design of chromosome

The chromosome is designed with simple way in this study because of the feature of the wireless charging infrastructure. As explained at Section 3, there are I number of segments in overall route and the allocation of the inductive cable is decided according to each segment level. Therefore, there are I number of genes while each gene has a value of 0–1 binary integer whether the inductive cable is installed (=1) or not (=0) and one more gene for the value of the minimum required battery capacity. Then, the value of the minimum required battery capacity can be calculated by the Eqs. (2)–(6). Finally, the design of the chromosome is depicted in Fig. 3.

4.2.2. Initial solution generation

In this study, initial population, set of initial solution, is generated with the concept of random. That is, there are I number genes and the

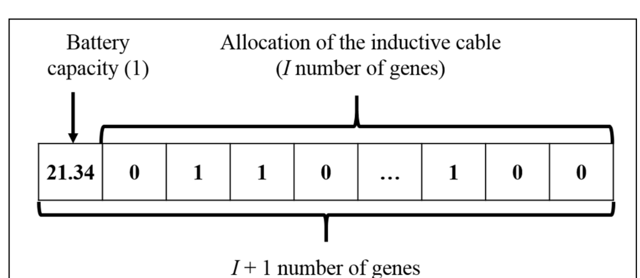

Fig. 3. The design of the chromosome.

value of each gene is decided as 0 or 1 (=0–1 binary integer) with random number generation. Then, the value of first gene which representing the minimum required battery capacity can be calculated with Eqs. (2)–(6).

4.2.3. Fitness function

The objective of the proposed mathematical model is to minimize the total investment cost. Therefore, the fitness function is regarded as the total investment cost. It consists of the battery purchasing cost, the inverter cost and the inductive cable cost like Eq. (1). Note that the tram purchasing cost is not important because it is constant regardless the values of decision variables.

4.2.4. Selection, crossover and mutation

For the selection, the roulette wheel method is adopted. The selection probability of each chromosome is decided with the fitness function value of it. When the fitness function value of certain chromosome is relatively high, then it has low section probability and vice versa because the objective of the proposed mathematical model is to minimize the total investment cost.

In case of the crossover, the two-point crossover is applied. When the randomly generated value is under the predetermined crossover ratio, then genes which belonging to the randomly decided two-point are exchanged between the selected two chromosomes. After the crossover, again, when the randomly generated value is under the predetermined mutation ratio, then genes which belonging to the randomly decided two-point of the selected chromosome are randomly reassigned as values of 0 or 1.

4.2.5. Termination

There are two kinds of generally applied conditions for terminating the genetic algorithm process. Firstly, when the ratio of same chromosomes exceeds a certain rate, then the process is finished. Secondly, when the creation of a child generation overs certain times, then the process is ended. In this study, because an optimal solution is already known, the termination condition is set as difference less than 0.1% comparing an optimal solution or 10,000 number of population generation.

4.3. Reinforcement learning for integrated genetic algorithm

The efficiency of the genetic algorithm tends to decrease when the length of a chromosome is extended. That means, it takes lots of time to generate an efficient solution because the genetic algorithm relies on the random characteristics when it derives the best solution. Sometimes, it is failed to find an optimal or near-optimal solution due to the long length of a chromosome. Therefore, to improve the quality of the chromosome, reinforcement learning could be integrated with the conventional genetic algorithm. In this study, the reinforcement learning would be integrated at the genetic algorithm with the stage of (i) initial solution generation, (ii) crossover and/or mutation, and (iii) both of them.

4.3.1. Reinforcement learning process

Basically, the reinforcement learning pursues to maximize the total reward. However, it considers not only the reward of current status but also the reward of future behavior and it is the main different with conventional several heuristics. The degree of consideration about the reward of future behavior is a key element of the reinforcement learning and if it considers more about the future behavior, the accuracy and the computation time tend to increase in the same time. The overall reinforcement learning process for the optimal design of the wireless charging electric tram is follow.

Step 1

Firstly, calculate the reward of current status using Eq. (1). In this study, the objective of proposed mathematical model is to minimize the total investment cost. Therefore, the reward can be expressed as the total investment cost of current status.

Step 2

Update the reward using Eq. (12) where S^T is a set of segments where the wireless charging infrastructure is installed and $R(S^{T-1}, s)$ is a reward when s^{th} segment is newly included at set of S^{T-1}. In same manner, $R(S)$ is a reward of S^T and $R(S^{T-1}, s, s')$ is a reward when both s^{th} and s'^{th} segments are newly included at set of S^{T-1}. Note that α and γ are the system parameters.

$$\begin{aligned} R(S^T) &= \underset{s}{Max}\ \mathrm{R}(S^T, s) \\ &= \underset{s}{Max}(1-\alpha)\cdot R(S^{T-1}) + \alpha\,[R(S^{T-1}, s) + \gamma \max_{s' \in S - S^{T-1} - \{s\}} R(S^{T-1}, s, s')], \\ & s \in S^0 - S^{T-1} \end{aligned} \tag{12}$$

Step 3

Step 2 is repeated until there is no more improvement at the reward. Then, the status at that time can be regarded as the optimal or near-optimal solution.

4.3.2. Application the reinforcement learning for genetic algorithm

4.3.2.1. Initial solution generation. Generally, the initial population of a genetic algorithm is prepared with the concept of random. It is because that it is helpful for avoiding the generated solution that falling into local optimum. However, it can not be guaranteed the qualities of the solutions of randomly generated initial population. Therefore, to improve the qualities of the solutions of initial population, some number of chromosomes of initial population is replaced with the solution from the reinforcement learning presented in Section 4.3.1.

4.3.2.2. Crossover and mutation. After the crossover, two selected chromosomes can be performed the mutation when the randomly generated number is under the predetermined mutation ratio. In here, mutation is also executed with the concept of random. Therefore, it is introduced second mutation with second mutation ratio. When the randomly generated number is under the predetermined second mutation ratio, then exchange is occurred with the solution from the reinforcement learning and the selected chromosome between the genes of randomly generated two locations as shown at Fig. 4. Because the part of selected chromosome is replaced as good solution by reinforcement learning, the required battery capacity has a possibility to improve more small value while it is calculated with Eqs. (2)–(6) based on the information about the allocation of the wireless charging infrastructure.

5. Numerical examples

To evaluate the qualities of solution from reinforcement learning itself and solutions from integrated genetic algorithm with reinforcement learning, the numerical example is designed and performed.

5.1. Parameters

Fundamental system parameters are set similarly with the study of Ko et al. (2015). There are five stations as shown in Fig. 2 and the location of each station from base station is 650 m (1st station), 1150 m

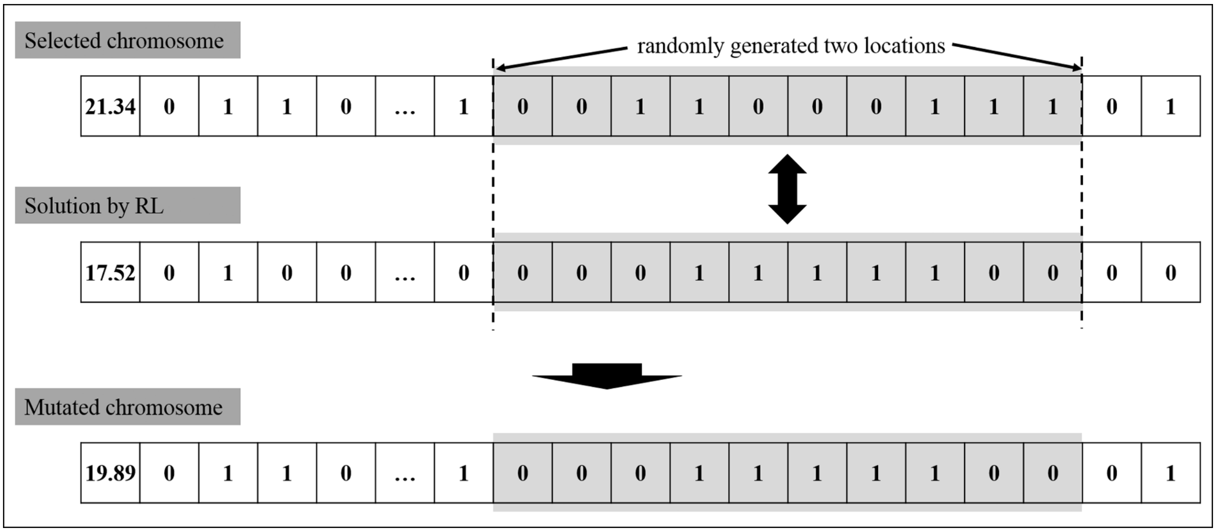

Fig. 4. Second Mutation with the solution by reinforcement learning.

Table 1
System parameters.

Parameter	Description	Value
c_{tram}	Unit tram cost without battery [$]	0
$c_{battery}$	Unit battery cost per kWh [$/kWh]	400
$c_{inverter}$	Unit inverter cost [$]	5000
c_{cable}	Unit inductive cable cost per meter [$/meter]	60
N	The number of the wireless charging electric tram [unit]	10
α_{max}	Maximum utilization ratio of battery, $0 \le \alpha \le 1$	0.8
α_{min}	Minimum utilization ratio of battery, $0 \le \beta \le 1$	0.2

(2nd station), 1800 m (3rd station), and 2150 m (4th station) while the overall route is 2800 m. Each segment is decided as equally 25 m, therefore there are 112 number of segments at overall route. The information about the battery consumption, and the battery charging by both the wireless charging (if the wireless charging infrastructure is installed) and the regenerative braking is presented as Appendix A. In addition, the parameters for numerical experiments modified from the commercial wireless charging electric tram system are derived as Table 1.

5.2. Results

First of all, the quality of solution by the reinforcement learning itself is compared with the optimal solution by the CPLEX while the number of tram in the system ranges from 5 to 20 and the results is presented as Table 2.

The solution by reinforcement learning derives an optimal solution when the number of tram in the system is five. At that time, the total investment cost, required battery capacity, no. of inverter and location of inductive cable are generated exactly same with the optimal solution of CPLEX. However, when the number of tram in the system increases, then the qualities of solution by reinforcement learning tend to decrease. It is because that when the number of tram in the system increases, then the wireless charging electric tram system tries to adopt more wireless charging infrastructure while keeping on small battery capacity. It can be confirmed by using the Gap while it is calculated by the (solution by RL – solution by CPLEX)/solution by CPLEX × 100. Note that computation times in both cases are less than one second.

The solutions by reinforcement learning have some differences comparing to the optimal solution by CPLEX even though the differences are not relatively large. It is because the proposed problem situation dealt with relatively simple circumstance for the deployment design of the wireless charging electric tram system. It is expected that if the problem situation is more complex, then the Gap between the solutions from CPLEX and reinforcement learning only will be larger.

Because it is impossible to derive an optimal solution by reinforcement learning only, in this study, integrated genetic algorithm with reinforcement learning is suggested. In addition, the reinforcement learning would be integrated at the genetic algorithm with the stage of (i) initial solution generation, (ii) crossover and/or mutation,

Table 2
Results by CPLEX and reinforcement learning only.

	CPLEX				Reinforcement learning			
	$N = 5$	$N = 10$	$N = 15$	$N = 20$	$N = 5$	$N = 10$	$N = 15$	$N = 20$
Total investment cost [$]	73,318	103,891	128,803	152,404	73,318	104,561	129,087	155,541
Required battery capacity [kWh]	20.66	12.60	11.80	11.80	20.66	13.14	12.60	12.57
No. of inverter [unit]	4	5	5	5	4	5	5	5
Location of inductive cable [meter]	625–675	0–75	0–100	0–100	625–675	0–75	0–75	0–100
	1125–1175	600–725	575–700	600–700	1125–1175	575–700	575–700	575–700
	1775–1825	1100–1225	1100–1250	1075–1250	1775–1825	1100–1225	1100–1225	1100–1225
	2125–2175	1750–1825	1775–1825	1775–1825	2125–2175	1750–1825	1750–1825	1750–1825
	–	2125–2200	2125–2250	2125–2250	–	2125–2175	2125–2200	2125–2200
Total length of inductive cable [meter]	200	475	550	550	200	450	475	500
Gap [%]	–	–	–	–	0	0.64	0.22	2.06

Table 3
Results by several versions of integrated genetic algorithm with reinforcement learning.

	CPLEX	RL only	GA only	GA + RL Case (i)	GA + RL Case (ii)	GA + RL Case (iii)
Total investment cost [$]	103,891	104,561	106,230	104,056	106,065	103,891
Required battery capacity [kWh]	12.60	13.14	15.06	13.01	14.64	12.60
No. of inverter [unit]	5	5	5	5	5	5
Location of inductive cable [meter]	0–75	0–75	0–25	0–50	0–50	0–75
	600–725	575–700	575–675	575–700	625–725	575–700
	1100–1225	1100–1225	1100–1175	1100–1225	1100–1225	1100–1225
	1750–1825	1750–1825	1750–1825	1750–1825	1775–1825	1750–1825
	2125–2200	2125–2175	2100–2175	2125–2200	2125–2175	2125–2175
Total length of inductive cable [meter]	475	450	350	450	375	475
Gap [%]	–	0.64	2.25	0.16	2.09	0

and (iii) both of them. It is assumed that there is 10 number of the tram in the system. The results are depicted in Table 3.

In case of only general genetic algorithm is applied, the value of the objective function has most large number. The gap which calculated by the (solution by GA – solution by CPLEX)/solution by CPLEX × 100 is 2.25%. It is reasonable result because it is known that the well-made genetic algorithm has a gap within 3% comparing an optimal solution. In this study, the reinforcement learning is adopted to reduce that gap at either initial solution generation step and/or crossover/mutation step of general genetic algorithm. All of three cases are derived smaller gap rather than that of only genetic algorithm is applied. They are 0.16%, 2.09%, and 0% for when the reinforcement learning is adopted at (i) initial solution generation, (ii) crossover and/or mutation, and (iii) both of them, respectively. Especially, when the reinforcement learning is adopted at both initial solution generation step and crossover/mutation step, the integrated genetic algorithm with the reinforcement learning can derive an optimal solution. Therefore, it is confirmed that the integrated genetic algorithm with the reinforcement learning is an efficient and powerful combination. In addition, the results show that the reinforcement learning which adopted at initial solution generation step is more efficient rather than when it adopted at crossover/mutation step. It seems that when the reinforcement learning is adopted at initial solution generation, it can influence more directly to the population while when it is adopted at crossover/mutation step, it can affect with probabilistic way. Moreover, when the reinforcement learning is adopted at initial solution generation step, the gap is smaller than that of when only the reinforcement learning is applied to solve the problem. Therefore, it is observed that the integrated genetic algorithm with the reinforcement learning is more efficient rather than when the reinforcement learning is only applied. Note that all of the cases mentioned at Table 3 have less than one second in terms of computation time. Therefore, it is expected that if the reinforcement learning is allowed to use more computation time, it seems that it can derive almost similar value comparing an optimal solution.

6. Conclusions

The wireless charging electric tram is an innovative means of transportation. It can reduce the battery capacity by adopting the wireless charging technology, but it is required the wireless charging infrastructure in same time. In this study, the mathematical model is proposed for the optimal deployment design of the wireless charging electric tram system with the objective to minimize total investment cost. That is, it is developed to derive the minimum required battery capacity which equipped at the wireless charging electric tram and the location and length of the wireless charging infrastructure. Both the battery and the wireless charging infrastructure are key elements about the total investment cost and they have the trade-off relationship.

Because the proposed mathematical model is integer programming, the CPLEX can generate an optimal solution. However, it is a commercial software, integrated genetic algorithm with reinforcement learning is suggested as the solution procedure. Firstly, the reinforcement learning scheme is introduced for deployment design of the wireless charging electric tram system. It can derive an optimal or a near-optimal solution within very short time, but it cannot guarantee an optimal solution when the problem situation tends to be complicated. In addition, general genetic algorithm is presented for deployment design of the wireless charging electric tram system. The termination condition of a genetic algorithm is set as the difference less than 0.1% comparing an optimal value or 10,000 number of population generation

For suggesting more efficient solution procedure, integrated genetic algorithm with reinforcement learning is developed. That is, the reinforcement learning integrated at the genetic algorithm with the stage of (i) initial solution generation, (ii) crossover and/or mutation, and (iii) both of them. It is conformed that the integrated genetic algorithm with the reinforcement learning is more efficient and powerful rather than the general genetic algorithm. When the reinforcement learning is adopted at both initial solution generation step and crossover/mutation step, it can derive an optimal solution. In addition, the reinforcement learning which adopted at initial solution generation step (Case i) is more efficient rather than when it adopted at crossover/mutation step (Case ii) because the reinforcement learning can directly influence in Case i while it can affect probabilistic way in Case ii.

The result of this study will be useful to the practitioner who want to apply the wireless charging electric tram system at certain candidate site. The practitioner is hard to know about where to allocate the wireless charging infrastructure among the overall route because optimal allocation of them is also related to either the operation profile and the battery capacity of wireless charging electric tram. That is, there is no exact rule to design of optimal deployment of the wireless charging electric tram without the algorithm proposed in this study. Then the total investment cost can be excessive, and it can slow the broad application of the wireless charging electric tram system.

As a further study, developed integrated genetic algorithm will be applied more complicated situation such as intersection of the route and multiple routes while assuming that the wireless charging electric tram are operated at those circumstances. In addition, the other kinds of meta-heuristic algorithms such as particle swarm optimization or mimetic algorithm can be integrated with the reinforcement learning and examine the efficiency of the solutions to solve the deployment design issues of the wireless charging electric tram system.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. 2017R1C1B5017672).

Appendix A

The energy supply and demand at each segment (kWh)

Segment no.	Battery consumption	Wireless charging	Regenerative braking	Segment no.	Battery consumption	Wireless charging	Regenerative braking
1	1.8096	4.1190	0.0000	57	0.1919	0.1389	0.0000
2	0.8664	0.3254	0.0000	58	0.1919	0.1389	0.0000
3	0.8519	0.2497	0.0000	59	0.1919	0.1389	0.0000
4	0.8465	0.2105	0.0000	60	0.1919	0.1389	0.0000
5	0.8446	0.1855	0.0000	61	0.1919	0.1389	0.0000
6	0.8445	0.1677	0.0000	62	0.1919	0.1389	0.0000
7	0.8455	0.1542	0.0000	63	0.1919	0.1389	0.0000
8	0.8471	0.1435	0.0000	64	0.1919	0.1389	0.0000
9	0.1919	0.1389	0.0000	65	0.0000	0.1435	0.3567
10	0.1919	0.1389	0.0000	66	0.0000	0.1542	0.3621
11	0.1919	0.1389	0.0000	67	0.0000	0.1677	0.3683
12	0.1919	0.1389	0.0000	68	0.0000	0.1855	0.3755
13	0.1919	0.1389	0.0000	69	0.0000	0.2105	0.3846
14	0.1919	0.1389	0.0000	70	0.0000	0.2497	0.3972
15	0.1919	0.1389	0.0000	71	0.0000	0.3254	0.4189
16	0.1919	0.1389	0.0000	72	0.0000	0.7857	0.5373
17	0.1919	0.1389	0.0000	73	1.8096	4.1190	0.0000
18	0.1919	0.1389	0.0000	74	0.8664	0.3254	0.0000
19	0.0000	0.1435	0.3567	75	0.8519	0.2497	0.0000
20	0.0000	0.1542	0.3621	76	0.8465	0.2105	0.0000
21	0.0000	0.1677	0.3683	77	0.8446	0.1855	0.0000
22	0.0000	0.1855	0.3755	78	0.8445	0.1677	0.0000
23	0.0000	0.2105	0.3846	79	0.8455	0.1542	0.0000
24	0.0000	0.2497	0.3972	80	0.0000	0.1542	0.3621
25	0.0000	0.3254	0.4189	81	0.0000	0.1677	0.3683
26	0.0000	0.7857	0.5373	82	0.0000	0.1855	0.3755
27	1.8096	4.1190	0.0000	83	0.0000	0.2105	0.3846
28	0.8664	0.3254	0.0000	84	0.0000	0.2497	0.3972
29	0.8519	0.2497	0.0000	85	0.0000	0.3254	0.4189
30	0.8465	0.2105	0.0000	86	0.0000	0.7857	0.5373
31	0.8446	0.1855	0.0000	87	1.8096	4.1190	0.0000
32	0.8445	0.1677	0.0000	88	0.8664	0.3254	0.0000
33	0.8455	0.1542	0.0000	89	0.8519	0.2497	0.0000
34	0.8471	0.1435	0.0000	90	0.8465	0.2105	0.0000
35	0.1919	0.1389	0.0000	91	0.8446	0.1855	0.0000
36	0.1919	0.1389	0.0000	92	0.8445	0.1677	0.0000
37	0.1919	0.1389	0.0000	93	0.8455	0.1542	0.0000
38	0.1919	0.1389	0.0000	94	0.8471	0.1435	0.0000
39	0.0000	0.1435	0.3567	95	0.1919	0.1389	0.0000
40	0.0000	0.1542	0.3621	96	0.1919	0.1389	0.0000
41	0.0000	0.1677	0.3683	97	0.1919	0.1389	0.0000
42	0.0000	0.1855	0.3755	98	0.1919	0.1389	0.0000
43	0.0000	0.2105	0.3846	99	0.1919	0.1389	0.0000
44	0.0000	0.2497	0.3972	100	0.1919	0.1389	0.0000
45	0.0000	0.3254	0.4189	101	0.1919	0.1389	0.0000
46	0.0000	0.7857	0.5373	102	0.1919	0.1389	0.0000
47	1.8096	4.1190	0.0000	103	0.1919	0.1389	0.0000
48	0.8664	0.3254	0.0000	104	0.1919	0.1389	0.0000
49	0.8519	0.2497	0.0000	105	0.0000	0.1435	0.3567
50	0.8465	0.2105	0.0000	106	0.0000	0.1542	0.3621
51	0.8446	0.1855	0.0000	107	0.0000	0.1677	0.3683
52	0.8445	0.1677	0.0000	108	0.0000	0.1855	0.3755
53	0.8455	0.1542	0.0000	109	0.0000	0.2105	0.3846
54	0.8471	0.1435	0.0000	110	0.0000	0.2497	0.3972
55	0.1919	0.1389	0.0000	111	0.0000	0.3254	0.4189
56	0.1919	0.1389	0.0000	112	0.0000	0.7857	0.5373

References

Alipour, M. M., Razavi, S. N., Derakhshi, M. R. F., & Balafar, M. A. (2017). A hybrid algorithm using a genetic algorithm and multiagent reinforcement learning heuristic to solve the traveling salesman problem. *Neural Computing and Applications,* 1–17.

Buzdalova, A., & Buzdalov, M. (2012). Increasing efficiency of evolutionary algorithms by choosing between auxiliary fitness functions with reinforcement learning. *Machine learning and applications (ICMLA), 2012 11th international conference on* (pp. 150–155). IEEE.

Cao, Z., Lin, C., Zhou, M., & Huang, R. (2018). Scheduling semiconductor testing facility by using cuckoo search algorithm with reinforcement learning and surrogate modeling. *IEEE Transactions on Automation Science and Engineering, 99*, 1–13.

Chen, Y., Mabu, S., Shimada, K., & Hirasawa, K. (2009). A genetic network programming with learning approach for enhanced stock trading model. *Expert Systems with Applications, 36*(10), 12537–12546.

Crites, R. H., & Barto, A. G. (1996). Improving elevator performance using reinforcement learning. In Advances in neural information processing systems (pp. 1017–1023).

Crites, R. H., & Barto, A. G. (1998). Elevator group control using multiple reinforcement learning agents. *Machine learning, 33*(2–3), 235–262.

Doya, K. (2000). Reinforcement learning in continuous time and space. *Neural computation, 12*(1), 219–245.

Fisher, T. M., Farley, K. B., Gao, Y., Bai, H., & Tse, Z. T. H. (2014). Electric vehicle wireless charging technology: A state-of-the-art review of magnetic coupling systems. *Wireless Power Transfer, 1*(2), 87–96.

Geels, F. (2011). The role of cities in technological transitions. *Cities and low carbon transitions,* 3e28.

Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with application to biology, control and artificial intelligence*.

Hui, S. Y. (2013). Planar wireless charging technology for portable electronic products and Qi. *Proceedings of the IEEE, 101*(6), 1290–1301.

Hwang, I., Jang, Y. J., Ko, Y. D., & Lee, M. S. (2017). System optimization for dynamic wireless charging electric vehicles operating in a multiple-route environment. *IEEE Transactions on Intelligent Transportation Systems*.

Jang, Y. J., Jeong, S., & Ko, Y. D. (2015). System optimization of the On-Line Electric Vehicle operating in a closed environment. *Computers & Industrial Engineering, 80*, 222–235.

Jeong, S., Jang, Y. J., & Kum, D. (2014). Design optimization of the OLEV system considering battery lifetime. *Intelligent transportation systems (ITSC), 2014 IEEE 17th international conference on* (pp. 2492–2498). IEEE.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*, 237–285.

Ko, Y. D., & Jang, Y. J. (2013). The optimal system design of the online electric vehicle utilizing wireless power transmission technology. *IEEE Transactions on Intelligent Transportation Systems, 14*(3), 1255–1265.

Ko, Y. D., & Jang, Y. J. (2018). Efficient design of an operation profile for wireless charging electric tram systems. *Computers & Industrial Engineering,* accepted.

Ko, Y. D., Jang, Y. J., & Lee, M. S. (2015). The optimal economic design of the wireless powered intelligent transportation system using genetic algorithm considering nonlinear cost function. *Computers & Industrial Engineering, 89*, 67–79.

Lai, M. (2015). Giraffe: Using deep reinforcement learning to play chess. arXiv preprint arXiv:1509.01549.

Lee, J. W. (2001). Stock price prediction using reinforcement learning. *Industrial electronics, 2001. Proceedings. ISIE 2001. IEEE international symposium on* (pp. 690–695). IEEE.

Liu, Z., & Song, Z. (2017). Robust planning of dynamic wireless charging infrastructure for battery electric buses. *Transportation Research Part C: Emerging Technologies, 83*, 77–103.

Liu, F., & Zeng, G. (2009). Study of genetic algorithm with reinforcement learning to solve the TSP. *Expert Systems with Applications, 36*(3), 6995–7001.

Lukic, S., & Pantic, Z. (2013). Cutting the cord: Static and dynamic inductive wireless charging of electric vehicles. *IEEE Electrification Magazine, 1*(1), 57–64.

Mabu, S., Hirasawa, K., & Hu, J. (2007). A graph-based evolutionary algorithm: Genetic network programming (GNP) and its extension using reinforcement learning. *Evolutionary Computation, 15*(3), 369–398.

Matarić, M. J. (1997). Reinforcement learning in the multi-robot domain. *Robot colonies* (pp. 73–83). Boston, MA: Springer.

Mikami, S., & Kakazu, Y. (1994). Genetic reinforcement learning for cooperative traffic signal control. *Evolutionary computation, 1994. IEEE world congress on computational intelligence. Proceedings of the first IEEE conference on* (pp. 223–228). IEEE.

Moreno, T., Reche, C., Rivas, I., Minguillón, M. C., Martins, V., Vargas, C., ... Ealo, M. (2015). Urban air quality comparison for bus, tram, subway and pedestrian commutes in Barcelona. *Environmental Research, 142*, 495–510.

Samma, H., Lim, C. P., & Saleh, J. M. (2016). A new reinforcement learning-based memetic particle swarm optimizer. *Applied Soft Computing, 43*, 276–297.

Sarker, A., Qiu, C., Shen, H., Gil, A., Taiber, J., Chowdhury, M., ... Rindos, A. J. (2016). An efficient wireless power transfer system to balance the state of charge of electric vehicles. *Parallel processing (ICPP), 2016 45th international conference on* (pp. 324–333). IEEE.

Sierzchula, W., Bakker, S., Maat, K., & van Wee, B. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy, 68*, 183–194.

Tan, Z., Quek, C., & Cheng, P. Y. (2011). Stock trading with cycles: A financial application of ANFIS and reinforcement learning. *Expert Systems with Applications, 38*(5), 4741–4755.

Tesauro, G. (1995). Td-gammon: A self-teaching backgammon program. *Applications of neural networks* (pp. 267–285). Boston, MA: Springer.

van Otterlo, M., & Wiering, M. (2012). Reinforcement learning and markov decision processes. *Reinforcement learning* (pp. 3–42). Berlin, Heidelberg: Springer.

Vaz, W. S., Nandi, A. K., & Koylu, U. O. (2016). A multi objective approach to find optimal electric-vehicle acceleration: Simultaneous minimization of acceleration duration and energy consumption. *IEEE Transactions on Vehicular Technology, 65*(6), 4633–4644.

Zhang, W., & Dietterich, T. G. (1995, August). A reinforcement learning approach to job-shop scheduling. In IJCAI (Vol. 95, pp. 1114–1120).

Zhang, J., & Maringer, D. (2016). Using a genetic algorithm to improve recurrent reinforcement learning for equity trading. *Computational Economics, 47*(4), 551–567.