Electric vehicles management aims to offer a reasonable EVs dispatching plan, which intendeds to maximize the total revenue of the ride-hailing platform (dispatching center) and experience of passengers. Dispatching center is able to send three kinds of command to EVs in the fleet: assign an EV to serve an order; arranging an EV to charge at a charging station; repositioning an EV to a high demand area.

In details, dispatching center collects riding orders in normal working hours, assigning an EV drive to the origin of an order, picking up the passengers and delivering them to the destination, considering the location, state of charge and service state of every EV. Whether an EV can sever an order depends on its service state, where "empty" means it is available and "on service" means unavailable. The EVs will execute commands one by one until the end of the day. Usually, EVs in the fleet will fully charge their battery during the night and start to work at 8 am.

In the working period, the dispatching center needs to monitor the state of charge of EVs, or an EV may encounter issue with low battery power and inability to serve passengers or even utilize the remaining battery to go to the nearest charging station. Meanwhile, given the complex flow of population in a busy city, the price of different areas may vary from time to time. Specifically, the unit price of areas where there is a great amount of riding orders at peak time will arise, while the unit price of areas where the amount of "empty" EVs substantially more than the amount of riding orders may decline. Therefore, it may be suitable for dispatching center to arrange some EVs reposition to some areas with high riding demand.

The service of a fleet usually covers the urban area of a city. The entire area will be divided into M nonoverlapping hexagonal sub areas, $G = \{g|1,2,\dots,M\}$. Note that each sub area could be the origin or destination of riding orders. Although passengers submit riding orders in continuous time, the decision period of the dispatching center will be discrete. That is, let $t = 1,2,\dots,T$ be the slices of decision period, and the dispatching center will all the riding orders submitted in $t-1$, determining whether assign EVs to serve them or not. The unsatisfied riding orders will be quit at the end of t, which means the passengers lose patience and turn to another type of transportation. Assuming that the number of EVs in the fleet is fix to N, while the full SOC of each

EV is E. When the SOC of an EV is lower than 5%, it needs to go to the charging station.

Therefore, the management of a dispatching center can be defined as a Markov decision process abbreviated as MDP. A MDP contains five fundamental definitions, which are described as a tuple: < S, A, P, R, γ>. Following are the details.

(1) State

Let $s_{g,t} \in S_t$ be the state of each sub area in a city, and $s_{g,t}$ contains the amount of available EVs, the amount of riding orders, unit price and current time of sub area g in period t. Note that the available EV means not only the EV is empty but also it can satisfy several constrains which will be introduced later. In this study, each EV will be seen as an independent agent, and the EVs in the same sub area will share the $s_{g,t}$. In reality, the EVs in a fleet can be divided into several model, corresponding to different levels of service quality. For example, when a passenger orders a riding service on DiDi Chuxing, he/she can choose to select several types of models like “Express”, “DiDi Flash”, “Select” and “Premier”. The expense of a riding order will vary from low to high given the different choice of model. Assuming that the SOC and speed of different type of EVs are the same, the only difference is the fix cost. Suppose there are three types of models in the fleet, that is “Express”, “Select”, “Premier”, as the fix cost of “Express” is lowest and “Premier” is highest. Passengers need to choose the model of EV when they submit the riding orders. Let $s_{i,t}$ be the state of agent i in period t, considering its current location, SOC, type of model and sub area state, i.e., $s_{i,t} = [l_i, E_i, h_i, s_{g,t}]$.

(2) Action

Let $a_t^E = [a_{1,t}, a_{2,t}, \dots, a_{N,t}]$ be the joint decision for all EVs in the fleet. The action of EVs contains three types.

- $a_{i,t} = Assign\ d_j$: means assign the ith EV to serve the jth order in period t. $d_j \in D_t$ represent the riding order collected by dispatching center from $t-1$ to t. An EV is able to serve an order when it can reach the origin within the maximum waiting time, and the SOC of the EV is enough for it to send the passengers to the destination as well as move to the nearest charging station. What's more the type of model of the EV should be the one the passengers choose.

- $a_{i,t} = Charge$: means arrange ith EV move to the charging station in period t. Usually, an EV needs to charge its battery when its SOC is lower than 5%, thus the dispatching center have to decide when to arrange the EV to charge and which charging station it should go to. Assuming that an EV become available again when it fully charges its battery. The consuming speed of battery should be a fix value ρ. Therefore, we can calculate the charging time of ith EV, as $t_i^C = \rho \times (E - E_i)$.
- $a_{i,t} = Reposition$: means reposition ith EV to another sub area in period t where the riding demand is higher. Repositioning an EV to other sub area will consume electricity, thus it should be a constrain of travel range to ensure that the future revenue of this EV exceed the cost of electricity.

(3) Reward and Return

Let $r_{i,t}$ denotes the reward obtained by ith EV in period t. Note that $r_{i,t}$ is the immediate revenue given an EV taking action $a_{i,t}$ while its state is $s_{i,t}$. Let $G_{i,t}$ denotes the cumulative discounted reward obtained by ith EV from period t until the end of the whole period.

$$G_{i,t} = r_{i,t} + \gamma r_{i,t+1} + \gamma^2 r_{i,t+2} + \cdots + \gamma^{T-t-1} r_{i,T-1}$$

γ is the discount factor in the equitation.

The reward functions of individual EVs are defined as follows.

- The reward for assign action is:

$$r_{i,t} = \rho^f + p_{g,t} \cdot D_{i,j}$$

where ρ^f is the fix cost of riding order j. As mentioned above, ρ^f differ in the type of models. $p_{g,t}$ is the unit price of the sub area g in period t, and the origin of d_j is g. $D_{i,j}$ denotes the distance start from the location of ith EV to the destination of the riding order.

- The reward for charge action is:

$$r_{i,t} = -(\rho^c + p_{g,t}^c \cdot (E - E_i - \Delta E))$$

where ρ^c is the base fare for charging. $p_{g,t}^c$ is the cost paid to charge, corresponding to the location of charging station in period t. ΔE represent the energy cost during the transportation from the current location to charging station. The reward for charging is negative but necessary.

- The reward for reposition action is:

$$r_{i,t} = 0$$

Since repositioning for a long distance may consume more electricity, in this study the destination of reposition is the adjacent area. Meanwhile, staying in the current is also a kind of special repositioning. We set the reward of repositioning equal to zero because it's not gainable or payable.

(4) State transition

Assuming that the state transition is deterministic, which indicates that $s_{i,t+1}$ is the state ith EV taking action $a_{i,t}$ with its current state $s_{i,t}$. Specifically, while the dispatching center commands the ith EV to serve riding order j, the travel time and consuming energy can be calculated. At the end of the assignment, the location l_i will be updated to the destination of riding order j, and the new SOC is $E_i - \rho \cdot D_{i,j}$. Additionally, since the EV moves to a new sub area, its $s_{g,t}$ is going to be updated as well.

Except for the MDP, several characteristics of the problem is needed to be discussed. Firstly, the objective of this study is maximizing the total revenue of the fleet as well as offering good experience to the customers, which means it has to satisfy riding orders as much as possible. In reality, the unit price of a sub area is determined by the ride-hailing platform, corresponding to the difference between supply and demand. For example, if the demand of a sub area is significantly higher than supply, the unit price may increase, indicating a more expensive fare. While whether the passengers are willing to paid for the fares or not depends on the actual cost of travel results from the origin to the destination. The passengers may accept higher prices within a certain range during the peak time, as many other passengers hope to obtain the riding service as well. The goal of ride-hailing platform is to find out the highest price that the passengers accept.

Let p^{accept} be the maximum unit price that the passengers are willing to accept, and p^{accept} cannot be obtained by the ride-hailing platform directly, which means it's private to passengers. Assuming that p^{accept} is uniformly distributed within $[p_{min}, p_{max}]$. The probability that the passengers of riding order j in sub area g accept the unit price p_g set by the ride-hailing platform is $G_j(p_g) = 1 - F_j(p_g)$, where $F_j(p_g)$ is the cumulative distribution function, i.e., $F_j(p_g) = P(p^{accept} \leq p_g)$.

The ride-hailing platform dynamically determines unit price p_g for each sub area g based on the difference between supply and demand. In details, let $p_g^b = \underset{p}{\operatorname{argmax}}\, pG(p)$, when the number of EVs is greater than the number of riding orders, which means the platform tries to find the highest unit price of sub area g. The pricing rules are as follows:

$$p_g = \begin{cases} p_g^b, & |V_{g,t}| \geq |M_{g,t}| \\ p_g^b(1 + 2e^{|V_{g,t}| - |M_{g,t}|}), & |V_{g,t}| \leq |M_{g,t}| \end{cases}$$

where $V_{g,t}$ is the number of EVs of sub area g in period t. $M_{g,t}$ is the number of EVs of sub area g in period t.

Assuming that the passengers are impatient and they will cancel the orders if the EV is not able to reach the origin in time. This phenomenon is very common in reality, for some passengers who are in emergency tend to switch to another transportation given the long waiting time. Let t^w be the waiting time of passengers, and the EVs who are unable to reach the origin within t^w will be regarded as unavailable.