Battery Maintenance of Pedelec Sharing System: Big Data Based Usage Prediction and Replenishment Scheduling

Battery Maintenance of Pedelec Sharing System: Big Data based Usage Prediction and Replenishment Scheduling

Chaofeng Zhang, *Student Member, IEEE,* Mianxiong Dong, *Member, IEEE,*
Tom H. Luan *Senior Member, IEEE* and Kaoru Ota, *Member, IEEE*

Abstract—Pedelecs are an alternative of traditional share bikes by applying the battery-powered motor to assist pedaling and accordingly extend the riding coverage. The large scale deployment of pedelecs, however, requires a careful design of maintainance system to replace the batteries regularly, which can be costly. This paper investigates the maintainance of city-wide pedelec system by developing an offline solution in two steps. First, we develop an optimal and efficient hybrid prediction model which predicts the usage demand of pedelecs in every 48 hours on a scale of millions of pedelecs. Our proposal predicts the future usage increment of pedelecs by combining a local predictor, a global predictor and an inflection predictor, which captures both the short-term and long-term factors affecting the pedelec usage. Second, based on the developed predictor and results of big data analytics, an optimal path planning scheme for the replenishment of pedelec batteries is developed. As compared to other schemes, our scheme can save 40% of the maintenance cost. To verify our proposal, extensive real-data driven simulations are performed which show that the accuracy of the prediction process is high enough than each traditional method and our proposal solves the maintenance problem efficiently.

Index Terms—Intelligent transportation systems, Big data analytics, Artificial intelligence.

◆

1 INTRODUCTION

Bike-share has recently become a rising trend worldwide as the last-mile solution in cities from the bus/subway station to home [1]. In a bike-sharing system, riders can use their smartphones to locate and use a shared bike available in the nearby. When finish riding, riders can either return the bikes to a bike station or merely leave bikes on the roadside until the next rider.

An issue of the legacy bikes is that people can quickly get tired, especially in windy weather or on steep roads, making the service coverage of sharing bikes limited to about 2-4 km. Pedelec is a new type of e-bike to solve this issue. The pedelec applies a battery-powered motor to assist the pedaling as in Fig. 1 and accordingly saves the energy of riders. When the pedaling speed is above a reasonable speed (e.g., 25 kilometers per hour), indicating a smooth and pleasant ride, the pedelec bike cuts off the electric power to save its energy. As a result, pedelec can extend the ride range to 5-10 km and replace vehicles as the last mile solution.

A key challenge of the pedelec system is how to manage the motor batteries. For instance, a fully charged battery in pedelecs can provide power within 8 hours [2], indicating that a pedelec may need to have its battery replaced for three times a day when continually used to provide 24-hour service [3]. To keep the large-scale service of pedelecs in the city, replenishment trucks are typically adopted across
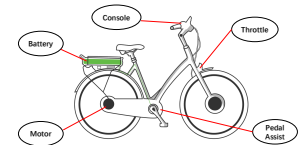
(a) Pedelec is stored at the roadside stations, instead, a maintenance truck stops at the roadside to provide battery exchange service

(b) One pedelec contains the components of the battery, console, controller, and motor; The installation and removal of the battery are less than 10 seconds

Fig. 1. Pedelec's share system, structure and battery maintenance

the bike sharing stations to change the batteries of pedelecs. Note that the schedule of replenishment is not the only key to the service quality of the entire pedelec system, but also determines the maintenance cost; frequent replenishment can guarantee the service of pedelecs but incur excessive maintenance cost to governments. This gives rise to the fundamental issue of replenishment path planning in pedelec systems, which is a challenging task due to the fact that with different locations and surrounding environments, pedelecs at different sharing stations have very different patterns of battery usage, and even within one station, the battery usage of pedelecs can be quite heterogeneous due to the different riding histories.

This paper targets to develop an optimal path planning scheme for replenishment trucks based on the big data analytics and prediction. To the best of our knowledge, we are the first to use the systematic methodology of modeling, analysis, and experiments to advance the e-bike prediction process using the captured datasets, and conduct the hybrid

- Chaofeng Zhang, Mianxiong Dong and Kaoru Ota are with the Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. E-mail: {16096014, mxdong, ota}@mmm.muroran-it.ac.jp
  Tom H. Luan is with the School of Cyber Engineering in the Xidian University, China. E-mail: tom.luan@xidian.edu.cn

prediction process which learns the internal connections between different dataset dimensions as well as optimizes the overall maintenance efficiency. Our contributions are three-fold:

- Big data based prediction: we develop an efficient offline prediction scheme based on big data analytics. We use a local predictor to analyze the characteristics of the station itself, a global predictor to analyze the global correlations among stations, and an inflection predictor to deal with unexpected situations. The prediction scheme therefore adapts to different factors and environments of deployments.
- Optimal path planning algorithm: based on the predicted results, we can analyze the degree of correlation between stations in real time. After making a station correlation map, a new pedelec maintenance algorithm is developed.
- Evaluation and verification: we select NYC's shared bicycle data and city-relevant big data for trace-driven simulation experiments. After comparing from the perspectives of prediction accuracy, average battery power, and overall service quantity, our solution is shown to outperform benchmark algorithms.

The remaining parts of this paper are structured as follows: Section 2 reviews the related works and highlight our contributions. Section 3 discusses the preliminaries of urban big data and then provides a brief overview of our hybrid pedelec system for prediction and maintenance. Section 4 discusses the local predictor and global predictor that complete the predictions using temporal and spatial information, respectively. Next, it discusses how the apparent inflection impacts the prediction of each station, and finally, an integrator is used to combine these predictive results. In Section 5, we develop a weight-based algorithm to solve the battery replacing problems of replenishment trucks. Section 6 conducts a large-scale simulation to test different methods, and Section 7 closes the paper with concluding remarks.

2 RELATED WORK

There are two main problems should be solved during our study. One is the traffic flow prediction, and another is the bike maintenance service optimization. The traffic flow prediction should enormously combine the big citywide data as much as possible, to fit well with the characters of a city and improve the prediction accuracy. Then, based on the reliable predictive results, the system could analyze the inner connection between the outputs and make a more efficient pedelec service system.

2.1 Urban Traffic Analyzing and Prediction

Pan et al. [5] detects the change of traffic using human mobility and social media, and shows that accidents, celebrations, disasters and the citywide events could result in a significant change of traffic flows. They develop a system that identifies the anomalies of driver's routing behaviors based on their regular patterns and representative terms on social media. Hoang et al. [6] propose a new planning method to describe the traffic flows in different regions. They use the common citywide big data, three different categories of movement patterns are decomposed and analyzed by the system, including periodic patterns, continuous changes, instantaneous changes. Then, Gaussian Markov Random Fields are adopted to exploit spatial-temporal dependence among different regions. Kong et al. [7] propose a new estimation and prediction for traffic congestions using car trajectories. They use a fuzzy comprehensive evaluation method to leverage the real traffic flow patterns. These methods are highly dependent on real-time data, which is similar to this paper. However, their customized predictions could not efficiently convert into ordinary citizens' cognitive services, which will be discussed deeper in this paper.

Another problem is how to make the flexible use of sensed data and collected information to improve the QoE. Torres et al. [1] aim to improve the sustainability and livability for the city and present a participatory sensing system called Becity to take the advantages of the collective knowledge of transportation cyclists, to improve the QoE of cycling. By combining the city usage information, cycling associations and government entities, the system provides bike routes with the shortest and accessible route, which is better than Google Maps. Caggiani et al. [4] discuss a multi-objective model to assist cyclists to keep lower travel cost, air pollution, and road safety. The new optimization model uses geolocated real-time data, which considering traffic congestions and flows, to select an optimal path and avoid air pollution or more safety. Then, they set up the graphics interface to guide users through the current optimized path, and finally, result in the improvement of the cycling experience.

Instead of using recently uploaded information to 'adjust' the original schedule and 'adopt' new changes, our work aims to 'predict' the future trend and make a countermeasure at the beginning, which is more efficient and less risky comparing with the 'problems occur then to be fixed' model.

3 SYSTEM MODEL AND OVERVIEW

3.1 Impacts of Urban Data on Station Usage

We first study the interplay between the big citywide data and the pattern of bike usage which shares insights on the developed prediction scheme. Through our preprocessing of urban big data, it is found that the bicycle usage data have a strong correlation with the characteristics of the station itself and the global influence: the station location determines whether the station service type is used for asymmetric travel during busy periods (unbalanced access), central transit station (high throughput) or edge extension node; weather, large-scale activities or temporal high throughput stations around the station also affects the use of the stations around an area and cause fluctuations in the use of the whole area.

A bike usage record set X includes time information, station information, customer information, and other decryption information. For a single record x, the duration time is noted as t_x, which is related to the start time $t_{s,x}$ and end time $t_{r,x}$. During the renting period, it forms a trajectory from start station ss_x to return station rs_x. Then we make statistics to build an index for each station, including the usage amount in different hours and dates. Fig. 2(a) and

(a) Number of bikes changed by date (station A)

(b) Number of bikes changed by date (station B)

(c) Number of bikes changed by hour (station A)

(d) Number of bikes changed by hour (station B)

(e) Three nearby stations of A

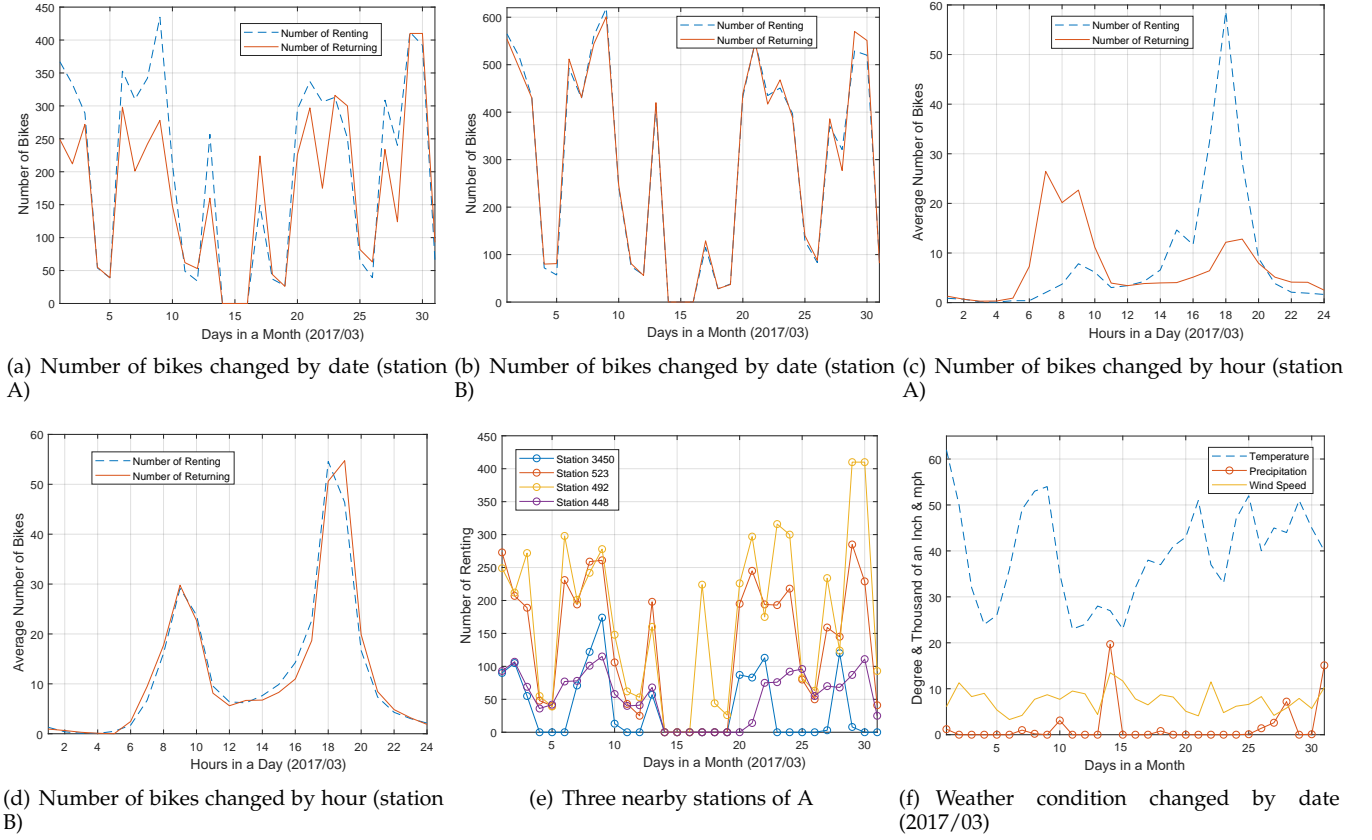(f) Weather condition changed by date (2017/03)

Fig. 2. Preliminaries for bike sharing records. The statistical result ranges from different hours and days, for analyzing the potential connections between bike usage and local meteorologies

Fig. 2(b) show the usage amount of starting from and returning to station A and B, respectively. Note that station B is a popular station downtown and keeps the highest usage record in March 2017. Station A is the second popular station this month. We can observe that station B is a 'center node' in the station network and reflects the usage trend of the whole city, while station A seems to be a functional area that forms unbalanced demand of bikes. Through Fig. 2(c) and Fig. 2(d), we find out that station A is in the resident area since the unbalanced usage ratio in the morning and closing time.

Another high relevant data set is the weather condition, which is similar to [8]. In Fig. 2(f), it shows the temperature, precipitation and wind speed in March 2017. In Fig. 3(a) and Fig. 3(b), the usage amount in 3 days (14th-16th day) is relatively low, as well as the temperature. Another observation is that the usage is more sensitive to unusual weather (e.g., rain). Fig. 2(e) is the usage of nearby stations compared with station A. The usage from stations nearby follows the same trend of increase and decrease with station A. The stations far away from station A have more or less opposite trends with it in.

In a word, these temporal datasets profoundly affect the usage amount of each station. Meanwhile, the spatial datasets such as coordinators and relative locations reflect the specific features in different areas [4].

3.2 Structure of Scheme

Fig. 4 shows the structure of proposed pedelec maintenance system, including three processes: (1) Preliminary Data Processing; (2) Prediction Process; (3) Maintenance Service.

Preliminary Data Processing: The process of preliminary data processing mainly handles with raw data from temporal and spatial datasets. It also provides the baseline of the predicting outcomes from local predictor and global predictor [9].

Prediction Process: Firstly, it loads the raw big data sets through different resources and formats them to meet the predictive model [10]. The raw temporal data includes all bike usage records and the weather report. The raw spatial data includes station coordinator and local meteorology.

First, we use two normal predictors: the local predictor and global predictor. The local predictor uses temporal data sets to predict the rough result as a baseline, The global predictor calculates the influence from nearby stations. Then, we analyze the sudden drop instances based on the observation [11] and develop a classification method that minimizes the variance while keeping most of them as ordinary cases [12]. This process of inflection detecting and predicting model is triggered when one of these features is beyond its threshold. Finally, it integrates the predictions of preliminary data and inflection data.

Maintenance Service: Instead, depending on the prediction of history data [13], the system follows the pattern of pedelec traffic flow and outputs the schedule. We propose
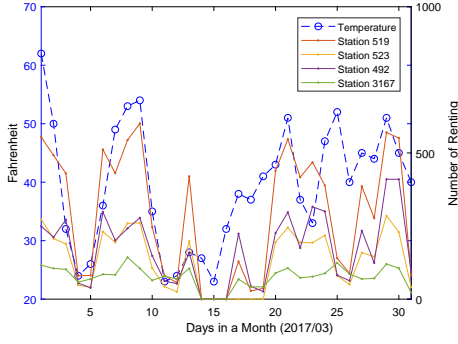
(a) Number of bikes influenced by temperature
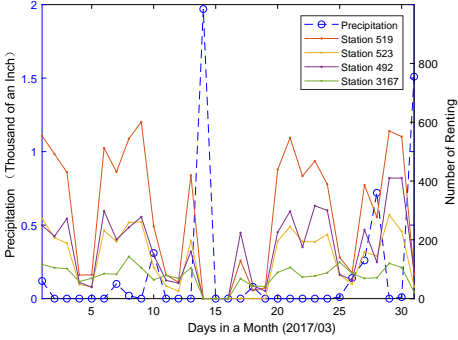
(b) Number of bikes influenced by precipitation

Fig. 3. Preliminaries for bike sharing records. The statistical result ranges from different hours and days, for analyzing the potential connections between bike usage and local meteorologies

an algorithm that calculates the optimized routes for replenishment trucks to exchange empty batteries. The maintenance monitor advises the administrative department with new updated predictions.

4 HYBRID PREDICTION MODEL

In this section, we discuss how the local predictor, global predictor and inflection predictor work to create predictions results, and finally gather them together using prediction integrator.

4.1 Local Predictor

Local predictors are primarily based on the characteristics of station itself to predict future usage. They are quantitatively analyzed for environmental and land form disturbances. The local predictor trains the temporal data sets including: 1) the usage record in the past days at the station; 2) the local meteorology (such as temperature/precipitation/snowfall/wind speed, or certain types of weather like Thunder/Ice Pellets/Blowing Snow); 3) calendar related information (holiday/weekend and workday); and 4) weather forecast information.

These features have correlation with the bike usage by our observation. All these features are formulated into numerical matrices, and are used as the original numerical format of National Weather Forecast Office of U.S.A: the data sets indicating one particular state of a period, such as thunder, ice pellets and even holiday, are formatted as

binary variables, e.g. $foggy = 1$ indicates this period is foggy; and other continuous data sets, e.g., temperature, precipitation, keep in standard matrices. The winds-peed is the average wind speed in miles per hour (mph). The precipitation is the total precipitation in one day to the nearest hundredth of an inch. PSBL is the percentage of possible sunshine, which sensed by dividing the minutes of sunshine by the total possible minutes; Some data sets related to human-made activities rules, are described by different binary variables even they belong to the same category.

Based on the input temporal data sets, the prediction on future few hours is computed. Here, we pick up the data set in the recent few days as the input. Since time is passing, the input of training sets changes every hour. We predict the recent six hours. For the next two days, we predict the intervals of each six hours. The reason is that the predictions of recent six hours are strongly influenced by the current data when the prediction of 7-48 hours roughly follows the historical trends.

Next, we discuss the details of linear regression modeling here. At first, as introduced in Section 3, we convert all the features into numeric documents: boolean values are changed into (0, 1) to represent the extreme weather (storm/thunder/snow); All of the date information is changed into timestamps and 1 - 7 to present the weekday. Notice that the already predicted results will not be put into the modeling process again since it may cause the iterative error due to the previous predictions. Also, the numeric weather forecast data set may change periodically, which affects the bike usage prediction distinctly.

4.2 Global Predictor

The traffic flows from nearby stations also impact on the bike demand [8] [14] [15]. The motivation for using a global predictor is that there are some connectivities between specific stations. Usually, a small surge in usage also affects the surrounding stations in the future. We infer indirectly the growth of bike usage based on the surrounding stations.

According to our statistics, users keep a speed of 10 to 15 km/h by cycling and 86.4% of the duration time is below 20 minutes (in 2016/04). To understand the geographical connections between stations, we develop a Bicycling Sharing Radar Cross Model (BSRC) to describe the nearby bike usage conditions. Fig. 5(a) shows the BSRC which represents the bike usage of nearby stations for station 491 (coordinator [40.741 -73.983]) in 2017/03. We divide the radar plane into fan shapes, noted as $Q_1, Q_2, Q_3, Q_4, \cdots$. The number in each fan shape means the number of renting records in this station. In one fan shape, we divide it into three regions based on the distance ([0, 500], [501, 1000], [1001 ∞] meters). Each region contains the sum of bike usage in the same fan shape with the specific distance range. The three different kinds of color (green, yellow, red) means the different number of levels ([0,20], [21 40], [41 ∞]).

Following reasons are why we turn associative zone to a BSRC and integrate the spatial data sets into the definition of regions:

1) It is impossible that we put all stations' data into the learning system individually [16].
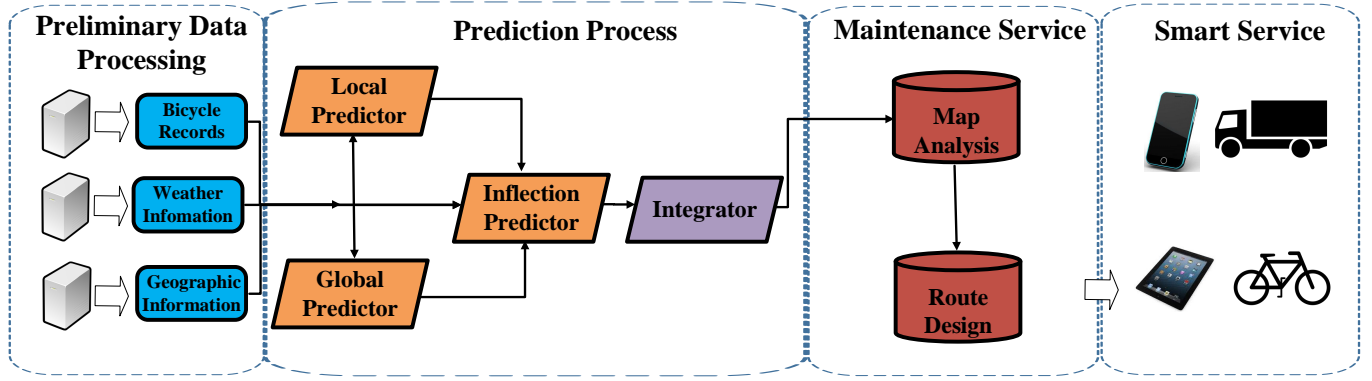
Fig. 4. The framework of the pedelec maintenance system. Our work contains three part, preliminary data processing for raw data, prediction process for usage trend prediction, and maintenance service for pedelec maintenance tasks

2) We use the total usage amount instead of the usage record that flows to the target station because it cannot reflect the real traffic flows.

3) Several pairs of stations have no direct path due to the river and highland. Therefore, these stations may become redundant in the prediction process.

4) The adjustable number of fan shapes provides the adaptability in different cities.

Then, we form the mentioned BSRC into arrays with the time series. Besides, not only the recent usage record is formed, other notable features are also considered in the BSRC: 1) weather features in the specific regions, including wind speed, humanity, temperature, precipitation, snowfall, minutes of sunshine, other unusual weather conditions (Fog/Thunder/Ice pellets/Hail/Glaze or Rime/Blowing Dust or Sand/Smoke or Haze/Blowing Snow/Tornado). 2) public service facilities, including those used for educational (schools, libraries), recreational(hospitals, parks, theaters, stadiums), and cultural (museums, memorial halls) buildings.

Fig.6 shows the structure of input data and ANN for prediction. All the datasets are formatted into numbers. The usage number of bicycle start $Start_x$ and return End_x records are counted based on the updating. The $Weather_x$ indicates the average temperature among these stations, and each station counts other parameters that represent states in this fan. As the same, the statistical results from the same fan are input to the full-connected neutral network group by group, with the order of i divided fan shapes and j distance based regions. Similar to Section 4.1, we predict the future hours based on the interval of six hours, respectively.

Then, we feed the data set into the 3-layer artificial neural network, which the time cost and accuracy are acceptable after adjusting several parameters. The output is the increment of bicycle usage $\Delta Usage$ comparing with the 'current' number of bicycle usage. Then, we can predict the usage in future k hours $Usage_{t+k} = Usage_t + \Delta Usage$. This value is also estimated by a threshold to detect the sudden change of bicycle usage in the city, which discussed later.

4.3 Inflection Predictor

An inflection predictor is proposed here to handle these sudden changes. Generally, The usage record of a station
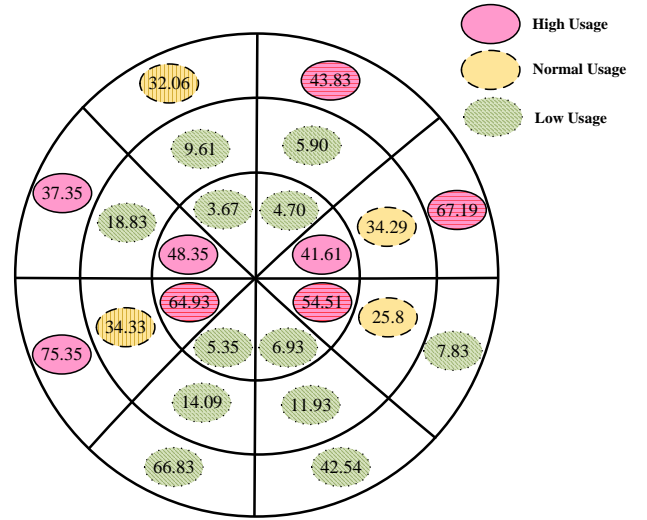
(a) A shared bike radar cross Model of station 491 (start)
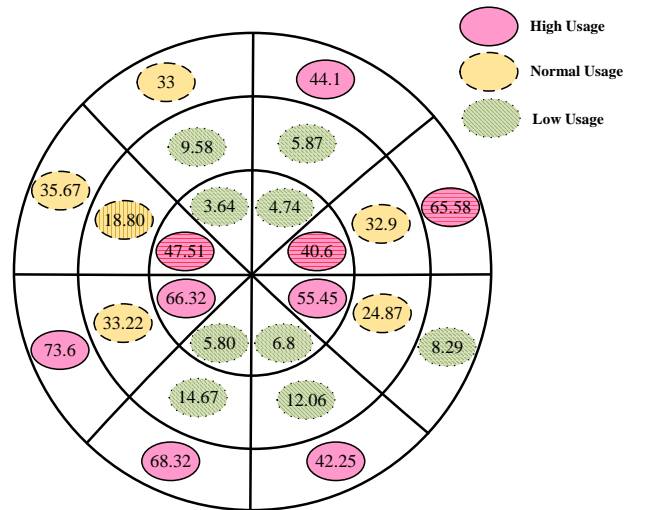
(b) A shared bike radar cross Model of station 491 (end)

Fig. 5. An example of BSRC. The red circle is high usage (> 40), yellow is normal usage ($20 \rightarrow 40$), and green is low usage (< 20). The distance of regions is classified as 500, 1000 and infinite in meter.

New Updated Dataset

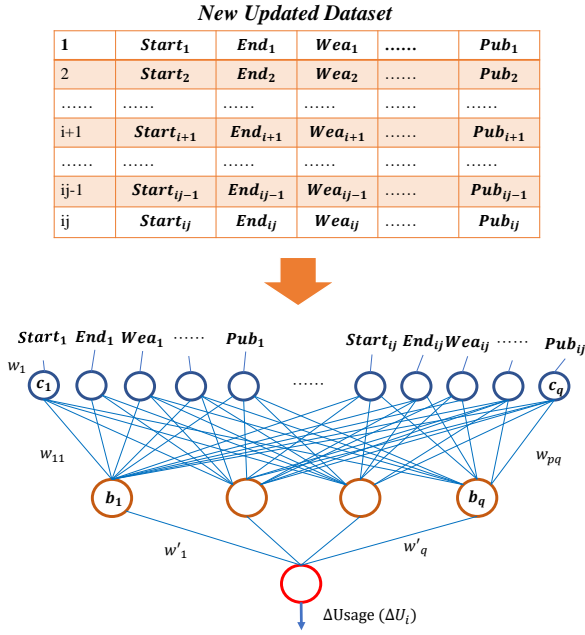| 1	$Start_1$	End_1	Wea_1	......	Pub_1
2	$Start_2$	End_2	Wea_2	......	Pub_2
......	......	......	......	......	......
i+1	$Start_{i+1}$	End_{i+1}	Wea_{i+1}	......	Pub_{i+1}
......	......	......	......	......	......
ij-1	$Start_{ij-1}$	End_{ij-1}	Wea_{ij-1}	......	Pub_{ij-1}
ij	$Start_{ij}$	End_{ij}	Wea_{ij}	......	Pub_{ij}

Fig. 6. Structure of input data set and a 3-layer artificial neural network for prediction. The table is an exsample of format for each input samples. The network structure shows the connection between input of train set, hidden neurons and output of future changes

may increase or decrease sharply during a few time intervals due to the sudden changes of terrible rainfalls. They genuinely show the apparent influence on people's daily life, and moreover, the reaction in emergencies. For instance, the heavy rain and strong breeze can stop most citizens' bike travel. As a result, the usage records in some stations keep 0. This kind of significant curve deformation we call them inflection situations, and we extract these data sets from the normal ones by the following steps:

1) Make the rough selection to separate the inflection dataset S' from the whole data sets S: At first, we set a threshold based on common sense to point out the unusual usage record. In a case of station A, if the number of records in one hour is 200% bigger or 10% than the average (603 and 30, respectively), we consider these records as the inflection dataset S' which may infect the usage curve deformation massively.

2) Find the notable sudden change ' interval': A sudden change interval for a feature (e.g., temperature) is a continuous numerical interval that the distribution of set S' in this feature is higher than the distribution of set $S-S'$. Similar with Fig. 7(a), We plot the probability distribution of set S' and the complementary set $S - S'$ for each feature. Then we can observe the intervals from features that the probability distribution in dataset S' is greater than S. These intervals are considered as the distinctive sudden change intervals for describing inflection data (or the distinctive category for discrete feature). When the feature of one new record falls into the distinctive interval, we can classify this record to dataset S' and active the inflection predictor.

3) Assign the distinctive intervals as thresholds. The range

of multiple distinctive intervals of a feature may be suitable for a given data set S'. However, we still need to find the precise connections if one feature holds multiple intervals. To find the precise set of distinctive intervals, we define an environmental fitness Q to form the notable interval selection problem into a coverage ratio optimization problem, where we have

$$Q = \frac{|D(y') - D(y)|}{|D(x') - D(x)|} \cdot \left(\frac{|x|}{|S'|} - \frac{|y|}{|S - S'|} \right) \quad (1)$$

where the absolute value signs $|S'|$ indicates the number of records in S'. x and y are the records belong to S and $S - S'$. $D(x') - D(x)$ indicates the difference of data square error before and after adding a new given set of distinctive intervals. Through the fitness Q, we can search for the notable interval set which covers most of the inflection records and affects them. This problem can be solved by a genetic algorithm or particle swarm optimization, which depends on the data scale.

4) Train an inflection predictor with the highest fitness interval set. After finding the highest fitness interval set, we update the inflection dataset S', which could satisfy all the thresholds (or intervals) and discrete features (e.g., thunder). Then we put the dataset S' back to local predictor and global predictor train a new hybrid predictive model independently, the process is similar to the mentioned predictor.

To distinguish the difference, we claim as the inflection predictor, and the output is the increment of bicycle usage $\Delta Usage$, comparing with the defined 'current' data set. When the time goes, and newly updated records come to the prediction system, it estimates whether they fall into the threshold and triggers on the inflection predictor if necessary. The system separates the records in two sets, $S - S'$ and S', and fed them into regular predictive process or inflection predictor, respectively.

4.4 Prediction Integrator

The prediction integrator combines the results of three predictors. We use a decision-making tree as the final step of our prediction process. Fig. 7(b) shows an example of a decision-making tree, and both continuous and discrete features are considered in this process. To process the continuous features, we use the influence of variance to detect the partition thresholds. The feature which has the most significant decrease of variance will be selected as the first ramification point. Then, the process estimates the other features in the next two separated branches and continues to create branches. Finally, it generates a level limited decision-making tree.

For a specific leaf i in the decision-making tree, we use a linear regression of bike renting usage N_i and returning usage N_i' to combine these prediction results and features, where we have

$$N_i = a_{i,1} \cdot Local + a_{i,2} \cdot Global + a_{i,3} \cdot Inflection \quad (2)$$

where a_i indicates the different coefficient impacted by predictors and weather features. We use a simple linear
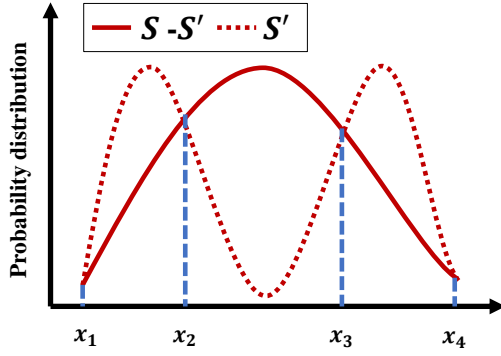
(a) An example of finding the sudden change interval (b) An example of using decision making tree for three outputs. Each pair $x_1 \rightarrow x_2$ and $x_3 \rightarrow x_4$ (the dash line shows the probability of lines shows an option based on different constraints, and each leaf is a distribution of inflection data set and full line show the linear regression function for final output. usual dataset)

Fig. 7. Examples of sudden change interval and decision making tree

regression to train these coefficients for each leaf N_i. For training these coefficients, the input is considered as the prediction results from each predictor, and output is considered as the real increment comparing with current usage. Based on our observation, the prediction result from local, global act as the similar behaviors and inflection predictors perform relatively higher coefficient when the weather seems unusual (e.g., the temperature is high). Other features (Wind-speed, PSBL) make the prediction a fine-tuning.

Algorithm 1 is the pseudo code of the overall prediction process. When new usage records updated by the online service, the system predicts the usage in future time intervals immediately.
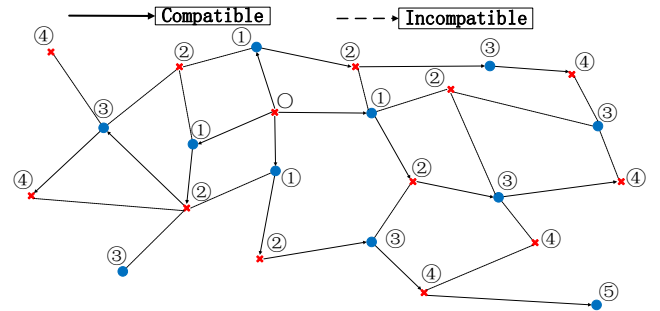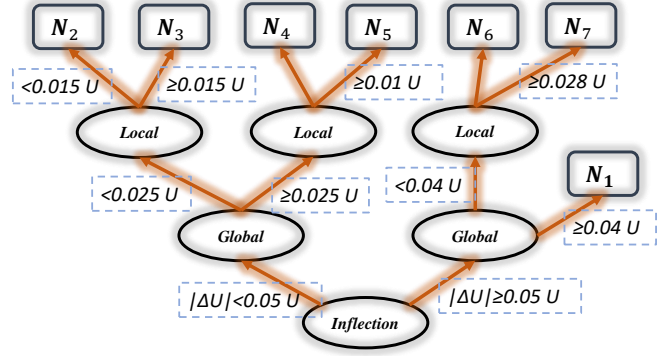

Fig. 8. An example of less incompatible maintenance map (solid and dashed lines are compatible and incompatible links between stations. Numbers from 1 to 4 present feasible routes for trucks.)

Algorithm 1 Prediction Precess using Three Predictors and An Integrator

Input: U: updated new bike usage record from the service; t_{max}: the number of targeted predictive results based on different time interval; C: thresholds to trigger on the inflection prediction process; $Info_t$: other environmental information in future time interval t;

Output: Predictive result of usage increment $\Delta Usage$;
    Array $\Delta Usage = \{0\}$;
    for $t = 1 : t_{max}$ **do**
        if $Info_t$ satisfies one of the conditions in C **then**
            Input U to $Inflection\ Predictor$ and get $\Delta U_{t,3}$;
        else
            $\Delta U_{t,3} = 0$;
        end if
        Input U to $Local\ Predictor$ then get $\Delta U_{t,1}$;
        Input U to $Global\ Predictor$ then get $\Delta U_{t,2}$;
        Active Prediction Integrator to combine the results $\Delta U_t = Integrator(\Delta U_{t,3} + \Delta U_{t,2} + \Delta U_{t,1})$;
    end for
    Update the prediction array ΔU using ΔU_t;

---

5 PEDELEC MAINTENANCE OPTIMIZATION ALGORITHM

We use these predictions to design a feasible optimization algorithm for every replenishment truck, which exchanges the batteries at each station. Then, divide our solution into two steps: (1). Construct a less incompatible maintenance map to offer suitable station-to-station links; (2). Reduce the dynamic replenishment problem into a classical TSP. (3). Develop a heuristic algorithm called intelligence sharing bike maintenance optimization algorithm (IMOA) to improve the performance.

5.1 Less Incompatible Maintenance Map

It may cause the overlapped problem if several trucks go around in the same area. The bikes may happen to follow the same route with the replenishment truck and be served twice [17] [18].

For such a problem, we set an indicator $r_{i,j}$ to indicate the proportion of returning at station i from the renting station j. Then we set a threshold $r \le r_{i,j}$ to indicate a less incompatible link from station i to j. We use the average distance of the collected distances of each most extended incompatible station, noted as l_1. Considering the time cost on the road, and there are more than 400 stations

in the city, one replenishment truck completes more than k maintenance missions within an hour. We set a maximum range for each mission. Therefore, the trucks can guarantee k missions within one hour, noted as l_2. Therefore, we can construct a less incompatible maintenance map to indicate the candidate links (distance $\in [l_1, l_2]$) from stations to stations that the trucks can follow. The stations that nearby enough (less than l_1) are considered as incompatible links and the stations that far away are considered out of truck's moving range.

Fig. 8 shows an example of less incompatible maintenance map, where the solid line is the feasible links between stations and dashed line means the indicator is too high for constant maintenance. The numbers indicate the feasible steps of each route. Through it, the system can narrow down the optional range of next stations and reduce computation time.

5.2 Reduced Dynamic Replenishment Optimization Problem

After making a less incompatible maintenance map, we design routes for trucks that cover lower battery remaining stations as many as possible. We assume there are w replenishment trucks from the warehouse and $s \in G$ stations are available for bike maintenance. We consider this warehouse as an origin point, which considered as an additional 'station' on the map. Therefore, this problem is reduced into a traveling salesman problem (TSP): Given a map list D of vertices (stations) and arcs (links) between each pair of cities, find the optimal route that satisfies maximized demands D.

Different from the classical static TSP problem, the replenishment demand of each station is continuously changing [19]. Although the calculated schedule may be the optimal solution in current period [20], however, as time goes by, the solution cannot fit with the demand change in each station. One potential solution is the genetic algorithm, which is a bionic optimization algorithm and has the capability of random large-scale search [21]. The calculation speed is adjustable and comfortable to combine with other algorithms, and is suitable for static TSP problems [22].

With the observation above, combining the advantages of ACO (Ant Colony Optimization) and GA [23] [24], we propose an intelligent pedelec maintenance optimization algorithm, which is based on the pheromones on the links left by the former trucks [25], to improve the efficiency of the whole system. This algorithm can make trucks no longer confined to the pheromones when selecting the paths, and make the use of existing information and search new path synchronously. It correctly increases the randomness and diversity of solutions and dramatically enhances the robustness of the basic heuristic algorithm.

Then, we describe the TSP problem again: There is a total number of n stations in the city, noted as $(1, 2, \cdots, n \in V)$. One truck departs from station 1 to start its tasks, where the probability travels to city i is noted as p_{1i}, and the probability travels to another city p_{1j} is independent. No matter how many time it takes at the station i, the time serving in other stations will be delayed by t_{1i}. For instant, one truck starts at station 1, then $p_{11} = 0$ and $t = 0$. It costs

time t_{ij} for the truck travel from station i to station j, where $t_{ij} = vd_{ij}$ and v is average speed of replenishment trucks. If the task at station i is completed, the truck travels to another station k with the probability p_{ik}. The optimization goal is that we design a schedule which contains the stations with the highest expected number of lower battery bikes.

Algorithm 2 Intelligence Sharing Pedelec Maintenance Optimization Algorithm

---

Input: $G = (V, A)$: created incompatible maintenance map contains vertex and arc; w: number of replenishment fleet; U: prediction results of bike usage; q: set qth best replenishment schedule;

Output: Schedule P;

Set incompatible indicator array $\sigma = IncIn(U)$;
Set travel probability array $p = TravelArc(z_0, \sigma)$;
Pheromone array $z = z_0$;
for each $g = 1; g \leq Generation; g++;$ **do**
    $Path = \{0\}$;
    Temporary Station Set $V' = V$;
    for each $i = 1; i \leq w; i++$ **do**
        $Path_i = RouteSelection(V', p)$;
        Delete stations in $Path_i$ from V';
    end for
    Sort $Path$ and Select best qth schedules, $Path'$;
    Pheromone array $z = UpdatePhe(z, Path')$;
    Travel probability array $p = TravelArc(z, \sigma)$;
end for
$P = Path$;

5.3 Intelligent Maintenance Optimization Algorithm

Next, we take the demand of each station into consideration. Trucks tend to select a path with higher demand, lower distance and higher pheromone provided by former trucks. Our prediction process predicts the demand for replenishment task, and pheromone is the pattern that describes the relativity of one link in a traditional optimized solution.

According to the above concerns, we define the probability that the truck travels to another station,

$$p_{ij} = \begin{cases} \dfrac{z_{ij}\sigma_{ij}^{\beta}}{\sum z_{ik}\sigma_{ik}^{\beta}}, & \text{where} \quad i, j, k \in V' \\ 0, & \text{where} \quad i = j \end{cases} \quad (3)$$

where z_{ij} indicates the pheromone of the link from station i to station j and set a parameter z_0 as the initial value. β is the adjustable parameter and $V' = \{1, 2, \cdots, n\} - V_0$ is the candidate stations, where V_0 is the set of stations served recently. σ_{ij}^{β} indicates the importance of stations' maintenance order. The importance σ_{ij}^{β} we define here is,

$$\sigma_{ij}^{\beta} = \frac{U \cdot S_r}{S_b \cdot t_{ij}} \cdot \frac{1}{c \cdot CS + d \cdot MS} \quad (4)$$

where U, S_r, S_b is the predicted bike storage, renting ratio and returning ratio. CS is the number of nearby served stations and MS is the number of nearby marked stations that are planned to be served. c and d is the adjustable parameter. t_{ij} is the time cost if travels to station j. This

σ_{ij}^{β} considers the possibility of dry battery bike storage in current stations, as well as the spreading of bikes before and after maintenance.

The algorithm allows the trucks to find the optimal path for updating the pheromones, and this parameter is updated globally, noted as

$$z_{ij} \leftarrow (1 - \phi) \cdot z_{ij} + \phi \cdot \triangle z_{ij} \tag{5}$$

where $\phi \in [0, 1]$ and $\triangle z_{ij} = D_{ij}$ which indicates the total number of served bikes at station j using the journey i to j.

However, before the system converges to global optimal solution, the current solution may fall into the local minimum, which leads to a stop of iterations. In order to avoid this, we update the rule of pheromones. At each iteration, only trucks with the best solutions can update z_{ij}. We select q trucks with the best q solution, and the definition of pheromones becomes,

$$z_{ij} \leftarrow (1 - \phi) \cdot z_{ij} + \sum_{q=1}^{m} \phi \cdot \triangle z_{ij,q} \tag{6}$$

where the new $\triangle z_{ij,q}$ we define here is,

$$\triangle z_{ij,q} = \begin{cases} \dfrac{D_{ij}}{q}, & \textbf{where} \quad i, j \in V' \\ 0, & \textbf{where} \quad i = j \end{cases} \tag{7}$$

where q indicates the qth optimal solution in current system.

When the new updated of prediction results come, firstly, the system modifies the incompatible map and reset the incompatible indicator array σ and travel probability array p. Then, set the initiate pheromone array as z_0 and enter into the iterations process. The main steps of iteration in the algorithm are described as the following: (1) We set w trucks and input the current prediction result. These trucks travel to another station following the less incompatible map and use Eq.3 to select the next station. (2) When all the trucks finished their journeys, the system sorts the qth best solution and uses Eq.6 to update the pheromones Δz on the arcs. (3) Reset all the trucks into the initial state and update the travel probability p between stations using new pheromones Δz. (4) Repeat the above steps until the number of iterations runs out. Finally, our IMOA outputs the routes with specific stations and time stamp for each time interval.

6 SIMULATION

This section verifies the performance of our proposal using simulations. In each simulation, we apply the proposed pedelec maintenance system in a real-world city map. The settings of pedelecs and city big data sets are extracted from the open data of governments.

6.1 Settings

We use the open data of bike sharing system 'Citibike' in NYC for our simulations. 'Citibike' is a widely used data set that has already been visualized. Besides, the geographic information and weather information are also well discussed by other studies:

'Pedelec' and Battery: As mentioned above, Pedelec is a kind of electric bike that is very useful when riding in

hills or strong winds. Since there is a revaluation about the battery in these e-bikes, we consider three different types of power replenishment system in the simulations. For an adult, he can keep the speed of 6 km/h without pedaling or keep the speed of 10 to 13 km/h within the range of 100 kilometers under the support of electronic controller. Therefore, we assume one lithium ion battery can keep working for 6 to 8 hours. When a Pedelec needs maintenance, the replenishment truck can install a new battery on it instead of the empty battery if stopping at the station. Since they all use rechargeable batteries, we assume one replenishment truck can replace all the batteries at the station.

'Citibike': It is a set of sharing bike records starting with 2013/12 and ending with 2017/04 in the website of Citibike [27], and we have shown the statistical result of total usage in Section 3.1. One essential record contains trip duration, start time & stop time, start station & end station, station locations, start time & end time, coordinator, etc. As counted in the introduction, there are more than 600 stations. We choose the data for this year (2017/03-2017/05) as our testing data, and the others are considered as training data.

Weather Report and Forecast: We use the data set of National Weather Service Forecast Office [28] for our study. One essential record in a specific area (e.g., central park) contains the date, max & min & average temperature, precipitation, snowfall, direction & speed & peak of wind, particular weather (fog, thunder, tornado).

Geography Information: Google API is applied in our study to gather useful geography information that helps the system understand the geographical correlations between stations.

6.2 Accuracy of Predictor

First, we analyze the accuracy of each predictor and discuss how a citywide big dataset performs for the pedelec maintenance system. Fig. 9(a) and Fig. 9(b) shows the average probability and error of bike usage at each station. The linear regression method reaches the accuracy of 67.39% within recent 6 hours, where the accuracy is defined as $\overline{A} = 1 - \sum_{i=1}^{n}(\overline{U} - U_i)/\sum_i U_i$, where U indicates the number of bike usage record. However, the additional week data (noted as 1 to 7) shows no help for the prediction, which means no strong connection with bike usage. Weather information makes the prediction reach 81.68% within new 6 hours. The average error is defined as $\overline{E} = \sum_{i=1}^{n}(\overline{U} - U_i)/n$.

It shows the prediction result is below 10.200 for all the stations, while linear regression method and additional week data can help reach 19.011% and 18.846%, respectively. Fig. 9(e) and Fig. 9(f) show the prediction result of global predictor. The average error of $ANN + BSRC$ is 7.275 within new 6 hours, which is the half of ANN only prediction (13.86). Fig. 9(c) and Fig. 9(d) show the accuracy of our inflection predictor and integrator. The upper and lower limits of the result are very high, therefore, the accuracy can reach 55.68% about new 6 hours and the average error is 22.15.

6.3 Results Comparisons of IMOA

In the experiment, we apply four maintenance strategies to evaluate our system. One strategy is the optimal solution

(a) Local predictor using citywide big data (average accuracy in each station)

(b) Local predictor using citywide big data (averaged error in each station)

(c) Global predictor using citywide big data (average accuracy in each station)

(d) Global predictor using citywide big data (averaged error in each station)

(e) Inflection predictor and integrator using citywide big data (average accuracy in each station)

(f) Inflection predictor and integrator using citywide big data (averaged error in each station)
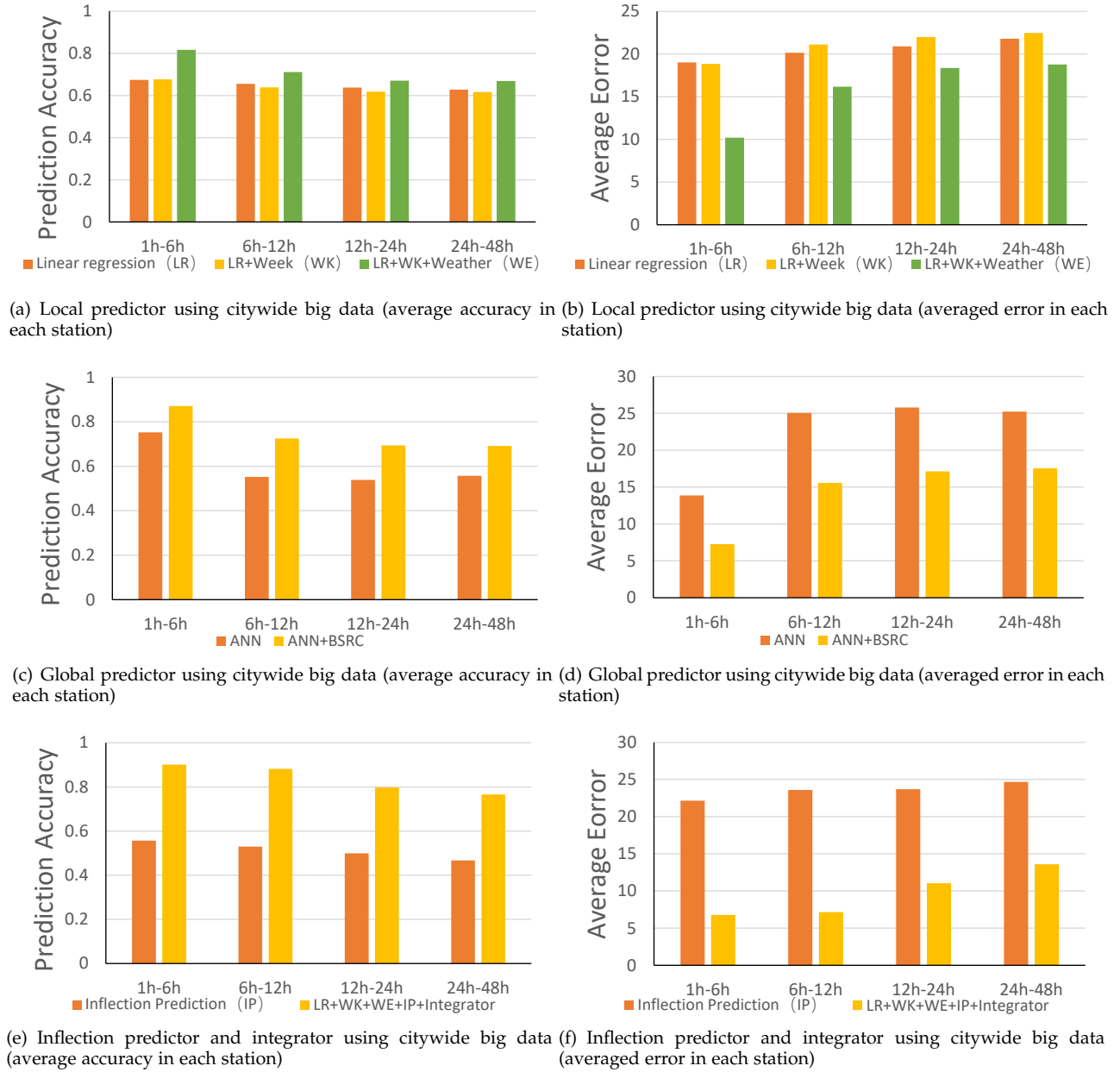
Fig. 9. Comparison of different prediction model.

that assuming all the future usage records have already been foreseen, and replenishment trucks ignore distance to travel immediately (noted as Optimal Solution, OS). IMOA method uses our proposed algorithm and the predicted results. The third algorithm uses the standard nearest neighbor procedure (NNP) with highest bike storage. The fourth one completes replenishment tasks at stations without strategy (randomly selection, RS).

Fig. 10(a), it shows the number of charged bikes during a day in 2017/04. The x-axis is the time series and we account every hour. The y-axis is the number of charged bikes, and 9765 bikes in 612 stations are monitored in that day. From the figure, we can observe that our proposed algorithm outperforms the other two since we take the emission traffic trend of charged bikes into consideration. Fig. 10(b) shows the average remaining battery using the same setting. The

y-axis is the change of the average remaining battery among all bikes. Similar to the former figure, we can observe that our IMOA algorithm can detect the traffic flows smartly. However, the remaining battery can reflect the active maintenance. Therefore, the slope about remaining battery is more smooth than the number of charged bikes. Finally, the remaining battery stays at a moderate level as well as the others. Fig. 10(c) and Fig. 11(a) are the average remaining battery using different settings. The former one is the average remaining battery changed by different replenishment trucks. The latter one is about the different batteries. The lead-acid battery can support average 3 hours while the up to date lithium iron phosphate battery keeps working for about 12 hours. The other two can finally keep average 76.3% and 88.7% batteries if the replenishment trucks keep working.
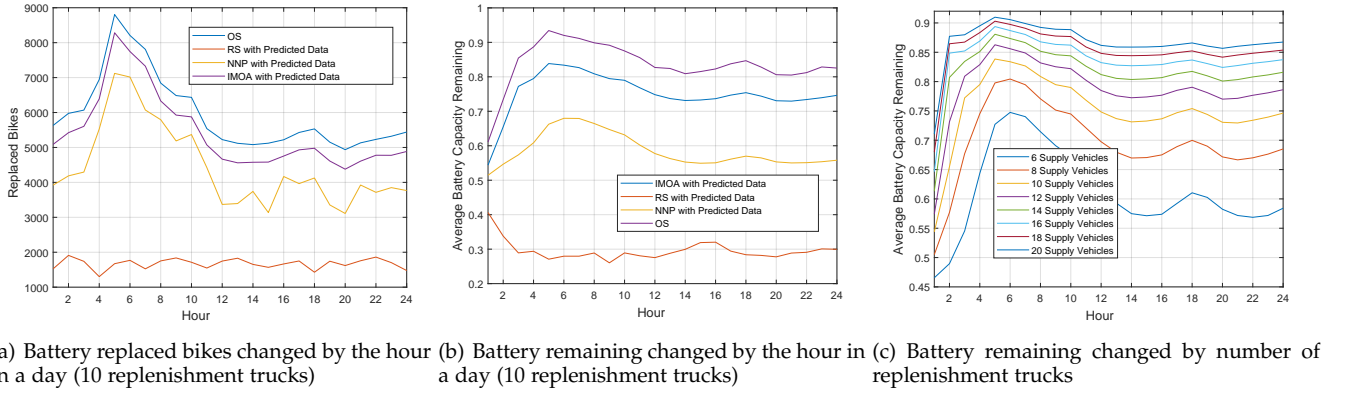
(a) Battery replaced bikes changed by the hour in a day (10 replenishment trucks)

(b) Battery remaining changed by the hour in a day (10 replenishment trucks)

(c) Battery remaining changed by number of replenishment trucks

Fig. 10. Performance of four strategies by different settings.

(a) Battery remaining changed by different batteries

(b) Number of low battery bikes changed by the hour (10 replenishment trucks)

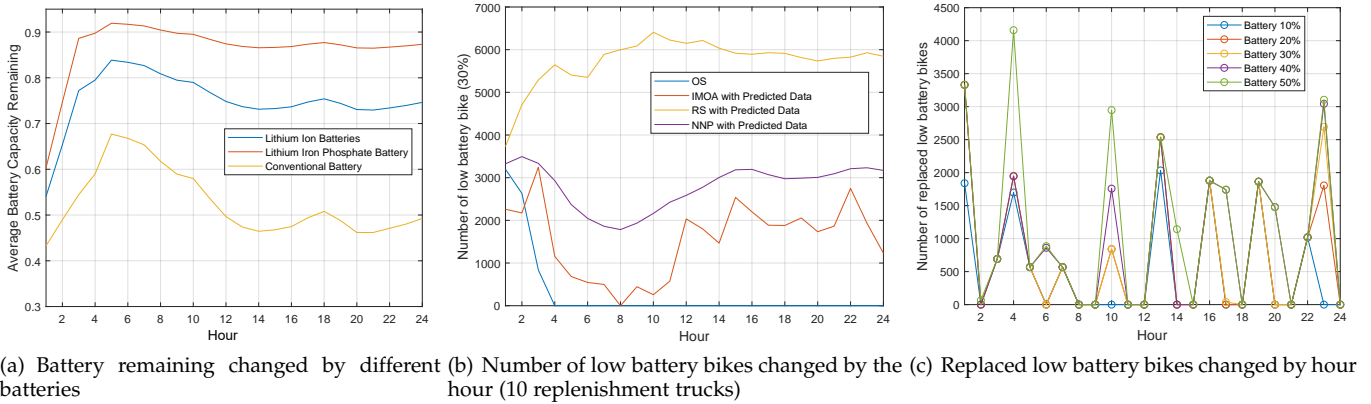(c) Replaced low battery bikes changed by hour

Fig. 11. Performance of battery replacing demand.

Next, we discuss the impact of low battery bike. Fig. 11(b) shows the number of low battery bikes using the same setting and algorithms. Our proposed IMOA algorithm performs better than the former one due to the better maintenance schedule, and reach 0 minimum at the start of 8 o'clock. Fig. 11(c) shows the number of replaced low battery bikes for one day, and we distinguish the different level of the remaining battery to observe the result. The higher number means more low battery bikes are satisfied, which creates the opposite slope with the former figure in most cases.

In a word, our proposed algorithm performs better than the other classical solutions, and the average remaining battery keeps at a relatively higher standard.

7 CONCLUSION

This paper has studied the maintenance of pedelec system based on big data analytics, which meets the growing maintenance demand of new types. We have first studied the real-world urban data sets from all collectable dimensions to explore the relations of different ingredients on the usage of sharing bikes. After that, a new hybrid predictive model to digitize the useful information in various fields and predict the future trend of bike usage has been developed. Finally, a heuristic algorithm called intelligent share bike maintenance optimization algorithm (IMOA) has been proposed to solve the maintenance problem for replenishment trucks. Through

simulations, we have verified our proposal from all angles and validate that the accuracy of our prediction model is high enough and IMOA solves the maintenance problem efficiently.

ACKNOWLEDGEMENT

REFERENCES

[1] S. Torres, F. Lalanne, G. del Canto, F. Morales, J. Bustos-Jimenez, and P. Reyes, "Becity: sensing and sensibility on urban cycling for smarter cities," in *2015 34th International Conference of the Chilean Computer Science Society (SCCC)*, pp. 1–4, Nov. 2015.

[2] C. T. Liu and R. C. Hsu, "A fuzzy q-learning based assisted power management method for comfortable riding of pedelec," in *2015 6th International Conference on Automation, Robotics and Applications (ICARA)*, pp. 580–585, Feb. 2015.

[3] I. Minakov, R. Passerone, and M. Rossi, "Design and energy optimization of a multifunctional iot solution for connected bikes," in *2017 Global Internet of Things Summit (GIoTS)*, pp. 1–6, Jun. 2017.

[4] L. Caggiani, R. Camporeale, and M. Ottomanelli, "A real time multi-objective cyclists route choice model for a bike-sharing mobile application," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 645–650, Jun. 2017.

[5] P. Bei , Y. Zheng, D. Wilkie, and C. Shahabi. "Crowd sensing of traffic anomalies based on human mobility and social media." *in the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 344-353, 2013.

[6] H. Minh, Y. Zheng, and A. K. Singh. "FCCF: forecasting citywide crowd flows based on big data." *in the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, p. 6, 2016.

[7] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," in *Future Generation Computer Systems*, pp. 97-107, Volume 61, 2016.

[8] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in *the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. SIGSPATIAL'13, pp. 344–353. New York, NY, USA: ACM, 2013.

[9] K. Ota, M. Dong, H. Zhu, S. Chang, and X. Shen, "Traffic information prediction in urban vehicular networks: A correlation based approach," in *2011 IEEE Wireless Communications and Networking Conference*, pp. 1021–1025, Mar. 2011.

[10] Y. Li, J. Gao, P. P. C. Lee, L. Su, C. He, C. He, F. Yang, and W. Fan, "A weighted crowdsourcing approach for network quality measurement in cellular data networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 300–313, Feb. 2017.

[11] X. Wang, Z. Zhou, Z. Yang, Y. Liu, and C. Peng, "Spatio-temporal analysis and prediction of cellular traffic in metropolis," in *2017 IEEE 25th International Conference on Network Protocols (ICNP)*, pp. 1–10, Oct. 2017.

[12] F. Lv, H. Zhu, S. Chang, and M. Dong, "Synthesizing vehicle-to-vehicle communication trace for vanet research," in *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 1–3, May. 2017.

[13] Q. Chen, M. Liu, and X. Liu, "Bike fleet allocation models for repositioning in bike-sharing systems," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 1, pp. 19–29, Spring 2018.

[14] K. Ota, M. Dong, S. Chang, and H. Zhu, "Mmcd: Cooperative downloading for highway vanets," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 34–43, Mar. 2015.

[15] X. Lu, K. Ota, M. Dong, C. Yu, H. Jin, "Predicting Transportation Carbon Emission with Urban Big Data," *IEEE Transactions on Sustainable Computing (T-SUSC)*, vol. 2, no. 4, pp. 333-344, October-December, 1 2017.

[16] M. Dong, K. Ota, and M. Sakai, "A novel information dissemination system for vehicle-to-rsu communication networks," in *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 918–919, Dec. 2013.

[17] Z. Liu, M. Dong, B. Zhang, Y. Ji, and Y. Tanaka, "Rmv: Real-time multi-view video streaming in highway vehicle ad-hoc networks (vanets)," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec. 2016.

[18] F. Paul and K. Bogenberger, "Evaluation-method for a station based urban-pedelec sharing system," *Transportation Research Procedia*, vol. 4, pp. 482 – 493, 2014, sustainable Mobility in Metropolitan Regions. mobil.TUM 2014. International Scientific Conference on Mobility and Transport. Conference Proceedings.

[19] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, p. 38, 2014.

[20] D. Gong and X. Ruan, "A hybrid approach of ga and aco for tsp," in *Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788)*, vol. 3, pp. 2068–2072 Vol.3, Jun. 2004.

[21] Z. Liu, M. Dong, B. Zhang, Y. Ji, and Y. Tanaka, "Rmv: Real-time multi-view video streaming in highway vehicle ad-hoc networks (vanets)," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec. 2016.

[22] W. Elloumi, N. Baklouti, A. Abraham, and A. M. Alimi, "Hybridization of fuzzy pso and fuzzy aco applied to tsp," in *13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, pp. 105–110, Dec. 2013.

[23] K. Honda, Y. Nagata, and I. Ono, "A parallel genetic algorithm with edge assembly crossover for 100,000-city scale tsps," in *2013*

[24] D. Chen, "Solving a new type of tsp using genetic algorithm," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 3333–3339, Oct. 2017.

IEEE Congress on Evolutionary Computation, pp. 1278–1285, Jun. 2013.

[25] M. M. A. Mohammed, C. He, and J. Armstrong, "Performance analysis of aco-ofdm and dco-ofdm using bit and power loading in frequency selective optical wireless channels," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Jun. 2017.

[26] M. H. Almannaa, M. Elhenawy, A. Ghanem, H. I. Ashqar, and H. A. Rakha, "Network-wide bike availability clustering using the college admission algorithm: A case study of san francisco bay area," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 580–585, Jun. 2017.

[27] https://www.citibikenyc.com/system-data

[28] https://w2.weather.gov/climate/index.php?wfo=okx

Chaofeng Zhang is currently a Ph.D student in Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. His research interests include cloud computing, full-duplex communication, wireless positioning technology.

Mianxiong Dong is currently an Associate Professor in the Department of Information and Electronic Engineering at the Muroran Institute of Technology, Japan. He serves as an Editor for IEEE Communications Surveys & Tutorials, IEEE Network, IEEE Wireless Communications Letters, IEEE Cloud Computing, and IEEE Access. He is currently the Member of Board of Governors and Chair of Student Fellowship Committee of IEEE Vehicular Technology Society.

Tom H. Luan is now a professor at the School of Cyber Engineering in Xidian University, China. His research mainly focuses on the content distribution and media streaming in vehicular ad hoc networks and peer-to-peer networking, and protocol design and performance evaluation of wireless cloud computing and fog computing.

Kaoru Ota is currently an Assistant Professor with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. Her research interests include Wireless Networks, Cloud Computing, and Cyber-physical Systems. Dr. Ota has received best paper awards from ICA3PP 2014, GPC 2015, IEEE DASC 2015, IEEE VTC 2016-Fall, FCST 2017 and IET Communications 2017.