Use Rate Prediction For Charging Stations

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

***Abstract*—With the development of electric vehicles, there comes a great need of charging stations for the recharging demand. However, where to locate a station and what are the main elements that operators should take into account when planning a setting, are still bothering problems that wait to be solved. In a common view, a better place to set a station ought to gurantee a relatively higher use rate of that station. Therefore, the problem changes into how to gain a higher use rate, and what are the factors that have important impact on it. In this paper, we propose a time frame based prediction framework of use rate for charging stations in Shanghai. The approach proposed in this paper takes both station's geographical information, such as longitude, latitude and Point of Interests(POIs), and working elements including price, AC/DC type and private or public to use, into consideration. We also seperate our datasets into different time frames including total time, weekday, weekend, morning, evening, moring_rush hours, evening_rush hours and travel hours. The aim is to classify whether it's a high-use-rate station or a low-use-rate one during different time periods. Experimental results show that our method performs well on LR, Random Forest, SVM and XGBOOST, which demonstrates that featrues as geographical information and working elements play an important role in use rate of charging stations.**

***Index Terms*—charging station, use rate, POIs, price, AC/DC, private, public, time frames**

I. Introduction

An electric vehicle charging station, also called EV charging station is an element in an infrastructure that supplies electric energy for the recharging of electric vehicles, such as plug-in electric vehicles and plug-in hybrids. At home or work, some electric vehicles have onboard converters that can plug into a standard electrical outlet or a high-capacity appliance outlet.

However, in most cases, others require a charging station that provides electrical conversion, monitoring, or safety functionality. These stations are also needed when travelling, and many support faster charging at higher voltages and currents that are available from residential EVSEs. Public charging stations are typically on-street facilities provided by electric utility companies or located at retail shopping centers and operated by many private companies.

Nowadays, electric vehicles become more and more popular as people want to cut down the pollution and cost of the usage of traditional energy. When most people considering whether to switch to an electric car, the most worrying aspect is the development of charging stations, which directly impact the usability of the electric car.

EVs in China is experiencing an overall growth in the past few years, which directly results in the massive construction and modification of charging stations and current road networks. Fig.1 shows the global distribution of an operators' charging stations in Shanghai, where there are over 600 stations. However, the cost of construction of charging stations is considerable and are often very costly or even impracticable to reallocate. This raise the question of how to select the locations for building the charging stations. In a typical view, a charging station no matter where it is located, its 'success' is often determined by the use rate of a station.

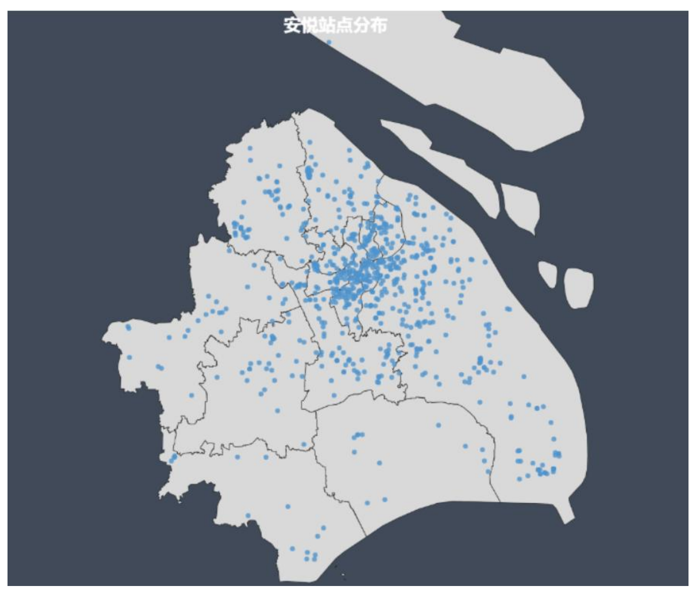


Fig. 1. Distribution of charing stations in Shanghai, China

In order to help with station setting strategies, we explore a time frame based prediction framework to classify stations into high use rate and low use rate in various time frames, using features like georaphical information as well as other working elements. We set 8 time frames in total, they are total time, weekdays, weekends, mornings, evenings, moring_rush hours, evening_rush hours and travel hours. As for features, we choose longitude, latitude, Point of Interests(POIs), charging price, AC/DC charging types and whether it's private or public to use.

We make use of an operator's charging station data in Shanghai and have kept collecting the use rate value for about a month, then we seperate the whole use rate dataset into various time frames that we have already planned. At the same time, we also collect important featrure data as formerly said. Both of the two datasets require a data cleaning procedure, in order to filter some invalid data or outliers. Furthermore, we make some analyses on features that we confirm to have significant impact on station's use rate.

We evaluate our method with four machine learning algorithm, they are RL, Random Forest, SVM and XGBOOST respectively. In each time frame, with features put into it, the model will tell which level of use rate that a station belongs to. Experimental results show that geopraphical information as well as working elements of a station do have great influence on its use rate, which can bring operators some enlightment on location choosing for station construction.

In summary, the contributions of this work are listed as follows:

- We propose a time frame based prediction framework to classify whether a staion is of high use rate or low use rate based on operator's charging station data and important features;

- We make detailed analyses on both of the two datasets to obtain basic information and find the relationships between station's use rate and those features;

- We make use of four learning models for improving the accuracy of use rate prediction and achieve relatively favorable results.

The rest of this paper is organized as follows: Section 2 gives definition of the problem. Section 3 reviews some related work done before. Section 4 describes work on collected dataset and feature analyses. Section 5 provides experiments with four machine learning algorithms and the results. In Section 6, we draw a conclusion to the paper and make discussions on future work.

II. Design Overview

A. Problem Formulation

There are many elements that affect location selection for charging stations when operators make the decision. In this paper, we hold that use rate of a station is the key factor which determines the 'success' of the station. Therefore, the original problem turns into how to get a higher use rate and what are the factors behind it.

The main objectives of our work is three-fold. First, we aim to explore some important featrues that have great impact on use rate of a station using spatio-temporal data of operator's charging station data. Second, we propose to study stations' different 'behaviours' during diverse time frames. Finally, on the basis of time frame based framework, we aim to predict that one station is of high use rate or low use rate according to its georaphical information and working elements during different time periods.

B. Design Methodology

Since we consider that diffrent features may have different influence on station's use rate, we make detailed analyses on features like geographical information and stations' working elements. Furthermore, station's use rate may differ during different time periods, so that we also study on each time frame to find the changes. Based on all the analyses mentioned above, we propose the time frame based framework to help with use rate prediction for charging stations. We then apply four machine learning algorithms including RL, Random Forest, SVM and XGBOOST to implement our experiments on different features and time frames.

III. Related Work

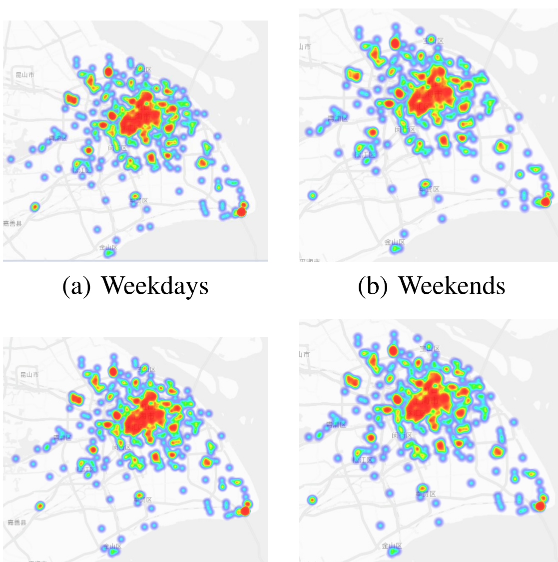
Hence, we studied the optimization problem of how to deploy charging stations. Existing works mainly fall into the domain of bike-sharing. [1] provieds a data-driven apporoach to deal with bike lane construction problem. It takes government constraints of planning bike lanes, such as budget limitations, construction convenience and bike lane utilization into consideration to formulate the problem. Furthermore, the problem is proved to be NP-hard so that they propose a greedy network expansion algorithm to help work out a scalable and approximate solution to bike lane planning problem. The approach performs well in the given problem, however it doesn't make use of learning models. [4] introduces a reinforcement learning algorithm to help solve the problem of repositioning sharing-bikes. First it uses an inner-balance clustering algorithem to cluster stations into groups, then the reinforcement learning algorithm is conducted in each group to learn a reposition policy. They make a good use of spatio-temporal data while don't take advantages of useful geographical and station-self features.

Current works of location selection are usually based on the flow prediction of a single station. Futhermore, they rely heavily on the historical data. [9] introduces a model for bicycle mobility prediction. It relis on historical bike-sharing data and a per-station basis with sub-hour granularity. It makes use of the randonm forest prediction model to implement their experiments and obtain a rather good result. [6] gives an optimization to this problem. In this work, traffic prediction no longer focus on the history data only, but can use location-based socail media to collect a much larger area of the traffic data for predicting traffic conditions. [8] is also a good example of prediction model for spatio-temporal mobility event. It encodes each POI's spatio and temporal dependencies rather than neglect the correlations between POIs. In this paper, we argue that the surrounding point of interests(POIs), distances to important POIs(e.g, metro stations, estates, etc.), station charging price, AC/DC station types as well as whether a station is private of public for use, play important roles in selecting the optimal location for stations.

IV. Feature Description

A. Hotspots

From the data we gathered, we made several observations that benefits the features to be included in the model we present. We separate the data into different time frames in order to determine the overall difference among them, since charging network is a dynamic system. From Fig.2, it is easy to notice that in different time frames, the charging hotspots stays at almost the same locations, meaning that during different time periods, the use rate of a given charging station is determined by its spatiotemporal context.



(a) Weekdays (b) Weekends

(c) Mornings (d) Evenings

Fig. 2. Charging Hotspots in Shanghai in different time frames

B. Point Of Interest

To the prosperity of Shanghai, there are so many points of interest(e.g., shopping malls, schools. estates, companies, etc.) located in the city. Understanding the purpose of the trip by each person who participates in charging system will help us to analyze the use rate prediction. So, we need to extract the POI around each existing charging station. There are a lot of POIs in Shanghai. The POIs are very close to each other. We set a radius around each station and then collect the POIs within the radius. Based on our experiment, choosing a radius as 300 meters is proper. In our work, we get 80 different types of POI totally. Some of the POIs are very similar to each other. Therefore, we group 80 POIs in further step. By grouping POIs, we get 10 groups at last. Table.I gives the groups and POIs in detail.

TABLE I
Groups of POIs

Group	Points of interests
Food	chinese & foreign restaurant, snack bar, cake & dessert shop, cafe, bar
Hotels	star hotel, express hotel, apartment hotel
Shopping	shopping centers, department stores, supermarkets, convenience stores, home building materials, home appliances digital, shops, markets
Education	institutions of higher learning, middle schools, primary schools, kindergartens, adult education, parent-child education, special education schools, study agencies, research institutions, training institutions, libraries, science and technology museums
Cultural venue	press and publication, radio and television, art groups, art galleries, exhibition halls, cultural palaces
Medical	general hospitals, specialist hospitals, clinics, pharmacies, medical examination institutions, nursing homes, emergency centers, disease control centers
Car service	car sales, car repair, car beauty, auto parts, car rental, car inspection field
Transportation	airport, railway station, subway station, subway line, long-distance bus station, bus station, bus line, port, parking lot, refueling station, service area, toll station, bridge, charging station, roadside parking space
Estates	Office building, residential area, dormitory

C. Distance

In use rate prediction, we need to consider distance. People will not choose to park their electric cars for charging if the destination they planned to go is far away. In the system, a station with nearer distance to metro stations, financial centers and major functional buildings would easily be used more often. We select the nearest distance to the following to be the distance we considered as features: company, estate, hospital, metro station, shopping center and university.

D. Price

By futher digging into the data, we find that there is a 0.3 correlation between price for charging and the use rate. Since there are two types of charging ports: DC and AC. We would include the number of ports and the price of both types in a charging station as one of its feature.

E. Private or public

By observing the data we've collected, it can be seen that most of the charging stations are private charging stations, which means they are typically used by electric public buses and rent cars, or used by specific companies for their employees, accounting for over 70% of the total charging stations. Also, since the private are used by more regular users(e.g., buses, companies employees), its use rate are 5% higher compared to public ones. This alongside other observations will be taken into consideration.

V. Time Frame Based Model

A. Time Frames

We separate the dataset into different time frames, including total time, weekday, weekend, daytime, evening time, morning_rush hours, evening_rush hours and travel_hours.Fig.3 shows the average use-rate of the time frames above.

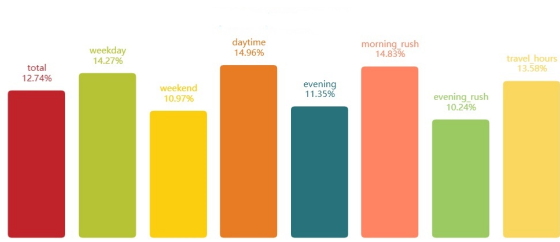


Fig. 3. Average use-rate of different time frames

B. Model Description

In each time frame, we add the geographical information as well as working elements into the prediction model, in order to predict whether it's a high use_rate station or a low use_rate one.

VI. Experiment

A. Datasets

In this paper, we gather the charging stations logs from all the existing charging station companies who provide their services in Shanghai, with a total of over 2,000,000 lines. The log has a length of one month, from 2018/10 to 2018/11, in which an hourly summarize of each charging station is recorded, showing whether it's occupied or not. After preprocessing, the training set has 80% of valid data, the test set has 20% left.

B. Implementation

With features extracted from our collected dataset, there are mainly three groups of them, including POIs, charing price and station types, that are added into the time frame based models. Fig.4 shows the general flow path of our work.

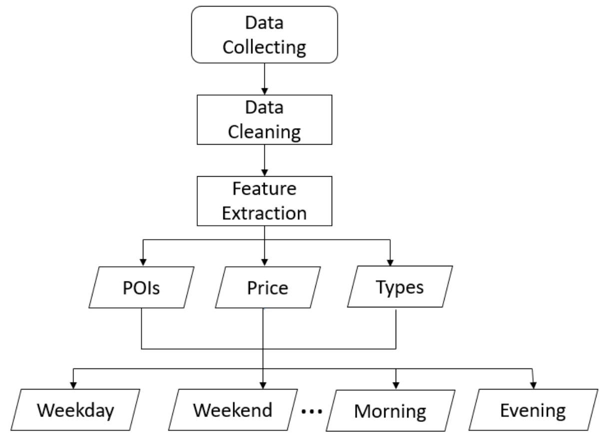


Fig. 4. Implementation of the work

C. Results

We run our dataset on LR, Random Forest, SVM and XGBOOST respectively, along with the features mentioned above added into time frame based models. Table.II shows the prediction accuracy of different models. We can see that they can all perform well based on our settings, and XGBOOST achieves the most favourable result.

TABLE II
EVALUATION RESULTS ON DIFFERENT MODELS

Model	Accuracy
LR	60.9%
SVM	68.9%
Random Forest	70.1%
XGBOOST	73.6%

VII. Conclusion and Future Work

In this paper, we propose the time frame based prediction model for use rate of charging stations. We study on some important features like station's surrouding POIs, charing price and station type. In experiments, we separate our dataset into different time frames and add the features into them, which obtains a relatively favourable result, indicating that the use rate of charing station is highly influenced by its geographical information and working elements. Furthermore, it also has a greate impact on location choosing problem.

There is still a lot of work to be continued in the future. First, we only consider three main types of features that might affect station use rate, many other important features also need to be extracted and incluede. Furthermore, we might explore a properly modified model to execute those features and gain a more satisfactory result.

REFERENCES

[1] Jie Bao, Tianfu He, Sijie Ruan, Yanhua Li, and Yu Zheng. Planning bike lanes based on sharing-bikes' trajectories. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*, pages 1377–1386, 2017.

[2] Minh X. Hoang, Yu Zheng, and Ambuj K. Singh. FCCF: forecasting citywide crowd flows based on big data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS 2016, Burlingame, California, USA, October 31 - November 3, 2016*, pages 6:1–6:10, 2016.

[3] Yaguang Li, Kun Fu, Zheng Wang, Cyrus Shahabi, Jieping Ye, and Yan Liu. Multi-task representation learning for travel time estimation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1695–1704, 2018.

[4] Yexin Li, Yu Zheng, and Qiang Yang. Dynamic bike reposition: A spatio-temporal reinforcement learning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1724–1733, 2018.

[5] Binbing Liao, Jingqing Zhang, Chao Wu, Douglas McIlwraith, Tong Chen, Shengwen Yang, Yike Guo, and Fei Wu. Deep sequence learning with auxiliary information for traffic prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 537–546, 2018.

[6] Xinyue Liu, Xiangnan Kong, and Yanhua Li. Collective traffic prediction with partially observed traffic history using location-based social media. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*, pages 2179–2184, 2016.

[7] Zhaoyang Liu, Yanyan Shen, and Yanmin Zhu. Where will dockless shared bikes be stacked?: - parking hotspots detection in a new city. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 566–575, 2018.

[8] Bilong Shen, Xiaodan Liang, Yufeng Ouyang, Miaofeng Liu, Weimin Zheng, and Kathleen M. Carley. Stepdeep: A novel spatial-temporal mobility event prediction framework based on deep neural network. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 724–733, 2018.

[9] Zidong Yang, Ji Hu, Yuanchao Shu, Peng Cheng, Jiming Chen, and Thomas Moscibroda. Mobility modeling and prediction in bike-sharing systems. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys 2016, Singapore, June 26-30, 2016*, pages 165–178, 2016.

[10] Xiao Zhou, Anastasios Noulas, Cecilia Mascolo, and Zhongxiang Zhao. Discovering latent patterns of urban cultural interactions in wechat for modern city planning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1069–1078, 2018.