Du-Parking: Spatio-Temporal Big Data Tells You Realtime Parking Availability

Yuecheng Rong, Zhimian Xu, Ruibo Yan, Xu Ma
Baidu, Beijing China
{rongyuecheng,xuzhimian,yanruibo,maxu}@baidu.com

ABSTRACT

Realtime parking availability information is of great importance to help drivers to find a parking space faster and thus to reduce parking search traffic. While there are limited realtime parking availability systems in a city due to the expensive cost of sensor device and maintaining realtime parking information. In this paper, we estimate the realtime parking availability throughout a city using historical parking availability data reported by a limited number of existing sensors of parking lots and a variety of datasets we observed in the city, such as meteorology, events, map mobility trace data and navigation data from Baidu map, and POIs. We propose a deep-learning-based approach, called Du-Parking, which consists of three major components modeling temporal closeness, period and current general influence, respectively. More specifically, we employ long short-term memory (LSTM) to model the temporal closeness and period, and meanwhile using two fully-connected layers to model the current general factors. Our approach learns to dynamically aggregate the output of the three components, to estimate the final parking availability of given parking lot. Using the proposed approach, we have provided the realtime parking availability information in Baidu map app, in nine cities in China. We evaluated our approach in Beijing and Shenzhen. The results show the advantages of our method over two categories of baselines, including linear interpolations, and the well-known classification model like GBDT.

CCS CONCEPTS

• **Information systems** → *Wrappers (data mining)*; • **Computing methodologies** → *Artificial intelligence*;

KEYWORDS

Parking Availability, DNN, LSTM, Spatial-temporal big data

ACM Reference Format:
Yuecheng Rong, Zhimian Xu, Ruibo Yan, Xu Ma. 2018. Du-Parking: Spatio-Temporal Big Data Tells You Realtime Parking Availability. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3219819.3219876

KDD '18, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08.
https://doi.org/10.1145/3219819.3219876

1 INTRODUCTION

Realtime parking availability information of parking lots is of great importance to help drivers to find a parking space faster and thus to reduce parking search traffic[3]. A significant portion of the traffic in big cities is caused by searching for parking spaces, which is responsible for up to 40% of the total traffic within cities[8]. In reality, however, there are insufficient realtime parking availability services in a city due to the expensive cost of sensor devices and maintaining realtime parking availability information.

In this paper, we estimate the realtime parking availability throughout a city using historical parking availability data reported by a limited number of existing sensors of parking lots and a variety of datasets we observed in the city, such as meteorology, events, map mobility trace data and navigation data from Baidu map, and POIs.

When estimating parking availability, spatial and temporal factors may have varying importance. Thus, the most important and challenging issues in parking availability estimation are:

- How to identify discriminative features from a variety of data sources.
- How to correctly model heterogeneous features to significantly improve the performance of estimation.

The advances in location-acquisition and wireless communication technologies have resulted in massive spatio-temporal data[17], which is used to address the big challenges in big cities[18]. In China, Baidu map possesses a dominant advantage in mobile map applications and our research is supported by the large-scale datasets collected from anonymous Baidu map app users. As an application of Du-Parking, we have provided the realtime parking availability information in Baidu map app, in nine cities(Beijing, Shanghai, Guangzhou, Shenzhen, Tianjin, Hangzhou, Nanjing, Suzhou and Jinan) in China. The Figure 1 shows that realtime parking availability displayed on Baidu map app.

In summary, our main contributions are as follows:

- We propose a DNN-based learning approach(Du-Parking), which consists of three major components modeling temporal closeness, period and general influence, respectively.
- We extract discriminative spatially-related and temporally-related features, contributing to not only our application but also the general problem of parking availability inference.
- We evaluated our approach using meteorological data, POIs, parking availability records of existing sensors of parking lots in Beijing and Shenzhen, events, and map mobility trace data and navigation data from Baidu map. The results demonstrate the advantages of our approach compared with 2 baselines.

The rest of this paper is organized as following. In Section 2, we discuss the related work of the proposed approach. Section

(a) Parking assistant in Beijing (b) Parking availability in Shenzhen

Figure 1: The visualization of parking availability in Baidu map app

3 presents the overview of the framework and definitions. We address the details concerning how to extract important features in Section 4. The proposed model is explained in detail in Section 5. Experimental results are presented in Section 6. Finally, we conclude this paper and suggest future work in Section 7.

2 RELATED WORK

Analysing parking data in terms of predicting parking availability in smart cities has received significant attention from researchers in last years. Most of them focus on parking availability prediction in the next intervals for those parking lots, which have parking sensors, and they use clustering[11, 13], Markov chain model[7], Poisson distribution[9], GBDT[1] and ANN[14, 19] to predict the parking availability based on the history parking availability data of parking sensors. However, the parking sensors and the service of maintaining realtime parking availability information are too expensive to cover a larger area. But little attention has been paid to parking availability estimation of the parking lots without parking sensors.

From the perspective of big data, it is possible to estimate parking availability information from crowdsourcing solutions like mobile applications[3]. In fact, learning an effective model from spatial-temporal data will significantly contribute to the big challenges in big cities[17]. Our work is inspired by the air quality inference based on the air quality data reported by existing monitor stations and a variety of data sources observed in the city[18].

Recurrent neural networks (RNNs) have been used successfully applied to sequence learning tasks[12]. The incorporation of long short-term memory (LSTM) allows RNN to continue to learn long-term temporal dependency. DNN based approach is proposed to model the the temporal properties of spatio-temporal (ST) data, which is summarized into three categories:closeness, period, and trend[16].

3 OVERVIEW

3.1 Preliminary

Definition 1: *Parking Availability.* Parking availability is used to describe the remaining parking space in a parking lot. In this paper, parking availability is divided into three levels by $parking\ remain\ ratio$ (PRR), and each level is assigned a descriptor. Let $parking\ free$ denote the number of free parking spaces, and $parking\ total$ denotes the total number of parking spaces in a parking lot. The $parking\ remain\ ratio$ is defined to be:

$$parking\ remain\ ratio = \frac{parking\ free}{parking\ total} \tag{1}$$

The parking availability level is described in detail in Table 1. The descriptor of each parking availability level is regarded as the class to be estimated, i.e. $\tau = \{L, M, H\}$. Each parking lot p has a parking availability level label $p.A$ ($p.A \in \tau$) to be estimated or already associated if having parking sensors.

Definition 2: *Point of Interest (POI).* A point of interest is a place in the physical world, which has a name, address, geographic location and other attributes. A parking lot is a POI. A parent POI of a parking lot is the building to which the parking lot belongs.

Table 1: Parking Availability Levels

Parking Remain Ratio	Parking Availability Level
$PRR \leqslant 0.15$	Low(L)
$0.15 < PRR < 0.30$	Medium(M)
$PRR \geq 0.30$	High(H)

3.2 Framework

Our system consists of three components: grid computing, realtime streaming computation and a online service. The whole framework is illustrated in Figure 2.

Grid computing: This component is an offline distributed system which is mainly for three kinds of functions:

1) Basic static feature extraction: POI related features such as POI categories, popularity of map queries and building age of each POI, are extracted from Baidu map data warehouse for both training and prediction.

2) Sample data processing: Our system converts realtime parking lots occupancy data which acquired from Baidu map every 30 minutes, into labeled sample data. Sample data stored in database for offline learning and evaluation.

3) Model training: Training dataset is generated by labeled sample data and all features which consist of static features and dynamic features. The proposed model are trained for each POI category, respectively.

Realtime streaming computation: A streaming computation system is to deal with geolocation coordinates, navigation data in realtime. Baidu map geolocation SDK and app are able to upload geolocation coordinates and navigation trajectory position. Our system map each coordinate into a geographic grid index. The mapped data is then stored in memory database for online prediction.

Online service: For result visualization, parking availability status should be updated periodically to present the real world parking availability. We first fetch POI related features, map query features etc. and retrieve realtime location and navigation trajectory from database, then all features are fused into a single feature vector for each POI at the current time interval, finally, we predict a parking availability level which is to be displayed in Baidu map app at the current time by trained model.

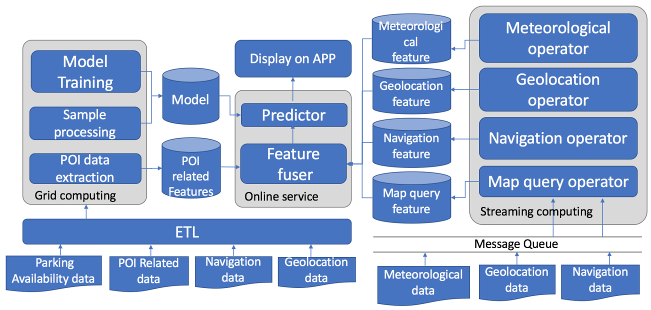


Figure 2: Framework of our system

3.3 Problem Statement

Given a collection of parking lots $P = P_1 \cup P_2 = \{p_1, p_2, ..., p_n\}$, where $p.A$ $(p \in P_1)$ is known and $p'.A$ is unobserved $(p' \in P_2)$, POIs corresponding to the parking lots of P, holiday events data affecting P, map mobility trace datasets and navigation datasets from Baidu Map users visiting P, and meteorological data of P, we aim to estimate $p'.A$, at periodic intervals, e.g., every 30 minutes.

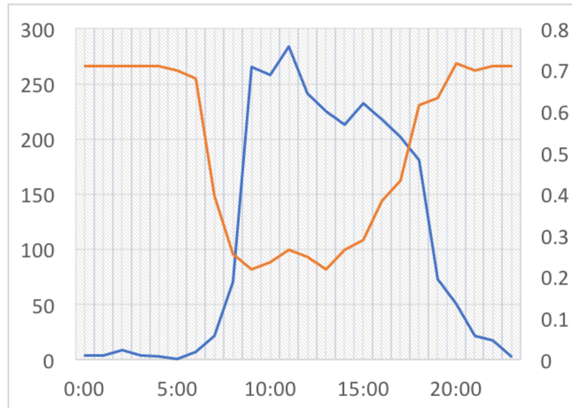
4 FEATURE EXTRACTION

4.1 Mobility features

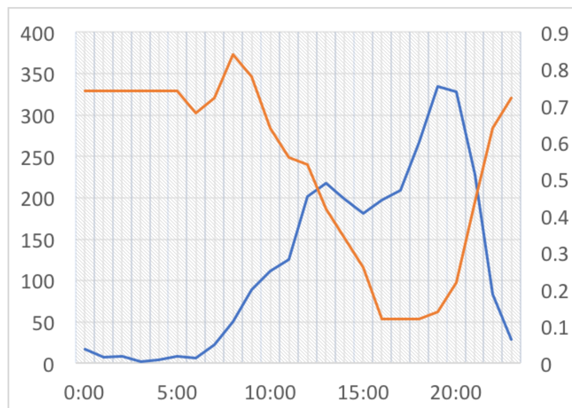
Mobility features F_{mob} is comprised of three types of features: geolocation features f_{loc}, navigation features f_{navi} and map query features.

- Geolocation features: Geolocation features is the number of unique visitors located in a parking lot area in a time interval, which presents human mobility in this area. It is believed that human mobility is strongly correlated to automotive mobility which is able to affect parking availability. The Figure 3 shows that the correlation between geolocation feature and parking availability. Apparently, geolocation features and parking availability are highly correlated.

- Navigation features: Navigation features f_{navi} concludes nav_{in}, the number of unique visitors arriving at the parking lot POI in a time interval, nav_{out}, the number of unique visitors leaving a parking lot in the same time interval. Navigation's arrival and departure represent automobile mobility which directly affects the parking availability. The Figure 4 shows the correlation between navigation features and parking availability.



(a) Office building



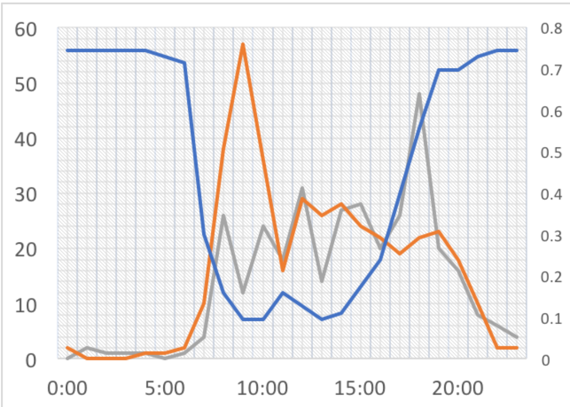
(b) Shopping mall

Figure 3: The Correlation between geolocation feature and parking availability of the same parking lot: (a): The orange line shows parking availability of an office building. During 8:00AM to 5:00PM, the parking availability is much lower than other times. The blue line denotes geolocation visitor number in the same building during 9:00 A.M. to 6:00 P.M., which is much more than other period. (b): The geolocation number peak of a shopping mall is from 5:00 P.M. to 8:00 P.M.. However, parking availability is low during this period.

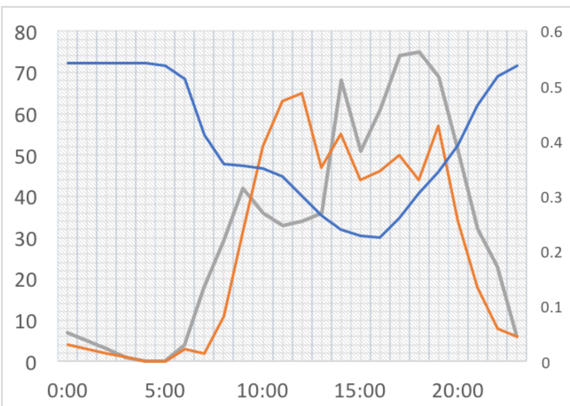
- Map query features: When a user searches for a place or plans a route using Baidu map app, the query keyword, i.e. the destination, and anonymous user's current location will be recorded and stored into Baidu's database[15]. The popularity of search queries shows the intention to transport to a new place of map users, so that it has connection with parking availability.

4.2 Meteorological Features

Traffic mobility could be affected by weather conditions. We defined three kinds of meteorological features(F_{mete}): temperature,



(a) Office building

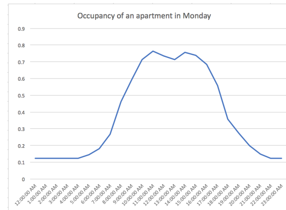


(b) Shopping mall

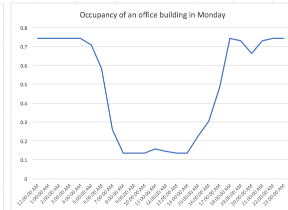
Figure 4: The Correlation between navigation feature and parking availability of the same parking lot. The orange line stands for nav_{in}, the gray line denotes nav_{out} and the blue line shows the PRR of the parking lot. (a): The popularity that navigate to a office building(nav_{in}) is more than nav_{out} during 7:00 A.M. to 10:00 A.M.. Meanwhile, the PRR become low. From 5:00 P.M. to 7:00 P.M., more people depart from this building than arrivals, more parking spots are available. (b):For a shopping mall, the nav_{in} increases at 11:00 A.M. and the parking remain ratio start to decrease at that time. The nav_{out} increases at 16:00 P.M., more parking spots are available from that time.

weather and wind speed. the whole map divided into 10km x 10km geographic grid, and realtime meteorological data of each grid are retrieved by the center coordinates from Baidu map developer platform. These data can be transform to a grid index of meteorological features.

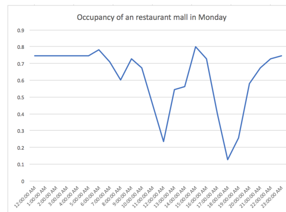
Each parking lot corresponds to a grid, so that it could be mapped to specified meteorological features. The bad weather conditions such as low temperature, raining, snowing or strong wind may lead to lower traffic flow than usual. Meanwhile, parking occupancy would be less, especially in scenic spots and shopping malls, but



(a) Apartment



(b) Office building



(c) Restaurant

(d) Shoping mall

Figure 5: The different categories corresponding to different parking occupancy patterns

office buildings and apartment may be less affected.

4.3 POI-related Features

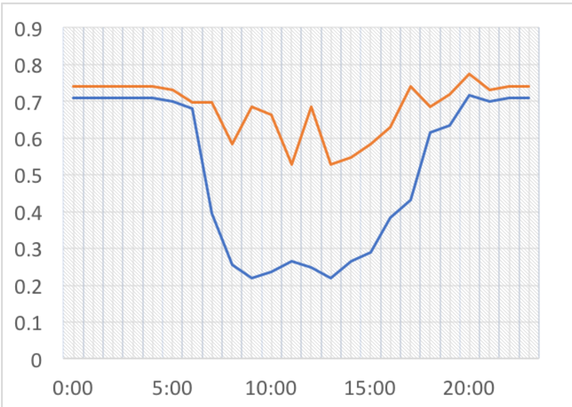
The parking availability of a parking lot is related to the category. The parking lot POIs are divided into many categories, such as "apartment", "hotel", "office building". There are too many categories for all the parking lots, so we merge similar POIs and divide them into seven categories. As shown in Table 2, they are "apartment", "office", "mall", "food", "hospital", "park", "entertainment" . The category is an important feature for this problem. The Figure 5 shows the parking availability of different category POIs on Monday in 24 hours. We also choose building age of POI as a feature which relating to the total spot number of current parking lot. The POI-related features are called F_{poi}.

Table 2: POI Categories

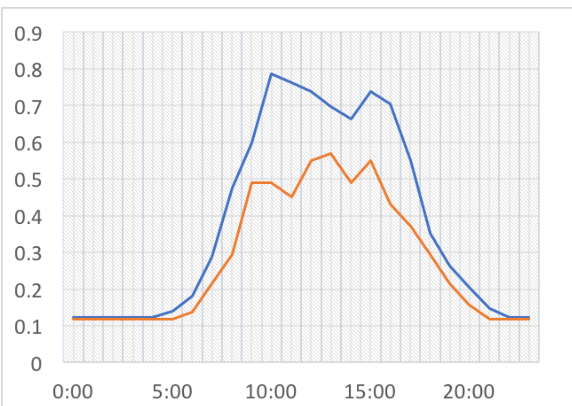
Categories	Summarized Categories
house, hotel, dormitory	apartment
office, company, government organization	office
shopping, super market	mall
restaurant, teahouse	food
clinic, nursing home, emergency center	hospital
museum, zoo, gallery	park
cinema, resort, KTV	entertainment

4.4 Holiday Event Features

The parking availability presents particularly high temporal regularity. Figure 6 shows the pattern of a parking lot in weekday and weekend/holiday. Obviously, the parking availabilities are different between weekday and weekend or holiday. Holiday event feature is highly discriminative feature for this problem. This feature should be included. The holiday event features are denoted as F_h.



(a) Office building



(b) Apartment

Figure 6: The orange line shows the PRR in weekend, the blue line denotes which in weekday. (a):The PRR in weekday is much lower than in weekend of an office building and it is more difficult to park during the day time. (b):the temporal dynamics of apartment parking lot is almost opposite of office building, but more parking lot are occupied in weekend.

5 PARKING AVAILABILITY ESTIMATION MODEL

The goal of the proposed model is to estimate parking availability of parking lots. Figure 7 shows the high-level overview of the proposed model, which is comprised of three major components modeling temporal closeness, period and general influence, respectively. We divide the time axis into two fragments, denoting recent time and near history. The temporal features of intervals in each time fragment are then fed into the first two components separately to model the two temporal properties: closeness and period, respectively. The first two components have the similar network structure with a LSTM followed by a fully-connected layer. In the current general component, the features at current time interval are fed into a two-layer fully-connected neural network. The outputs of closeness component $\boldsymbol{Y}_c$, the outputs of period component $\boldsymbol{Y}_p$ and the outputs of general component $\boldsymbol{Y}_g$ are fused as based on parameter vector, which assign different weights to the results of different components. Finally, the aggregation is fed through the softmax layer to produce the estimation of parking availability at the current time interval. Proposed neural network can be trained end-to-end.

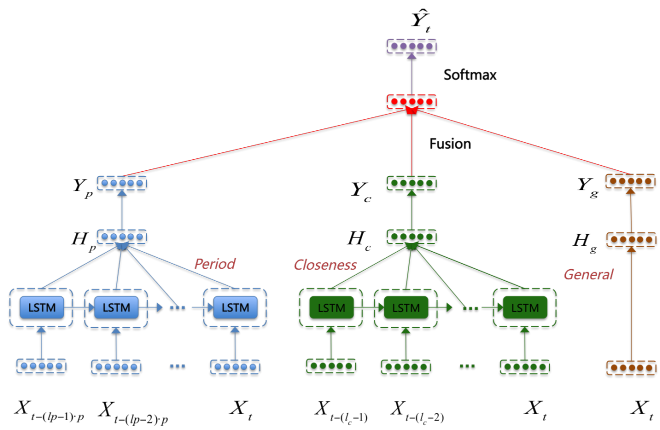


Figure 7: The Proposed Du-Parking Model

5.1 Structure of The First Two Components

The first two components (i.e. closeness and period) have the similar network structure, which is composed of LSTM and fully-connected layer, as shown in Figure 7.

The parking availability changes continuously over time, the same as the temporal factors that affect it. Intuitively, the past parking information may affect the current parking availability, which can be effectively handled by the Recurrent neural network (RNN) that has shown its powerful ability to capture the sequential structural information[12].

We use LSTM, as RNN variant dealing with the vanishing gradient problem[2], to adaptively capture temporal dependencies among temporally-related features at different times. In the online system of Baidu map, we update the model by the near historical datasets, so we wouldn't model the seasonal trend.

The closeness component of Figure 7 adopts a few features of intervals in the recent time to model temporal closeness dependence. In this paper, we denote the feature at time interval t as $\boldsymbol{X}_t = (F_{mob}, F_{poi}, F_{mete}, F_h, t)$ and let the recent fragment be $\left[\boldsymbol{X}_{t-l_c-1}, \boldsymbol{X}_{t-l_c-2}, ..., \boldsymbol{X}_t\right]$, which is also known as the closeness dependent sequence. $\boldsymbol{X}_i$ is fed into the LSTM, which output a hidden state $\boldsymbol{h}_i$. We concatenate the hidden state fragment $\left[\boldsymbol{h}_{t-l_c-1}, \boldsymbol{h}_{t-l_c-2}, ..., \boldsymbol{h}_t\right]$ as $\tilde{\boldsymbol{H}}_c$, which is followed by fully-connected layer as:

$$\boldsymbol{H}_c = ReLU\left(\boldsymbol{W}_{hc}\tilde{\boldsymbol{H}}_c + \boldsymbol{b}_{hc}\right) \tag{2}$$

$\boldsymbol{H}_c$ is fed through the fully-connected layer to output the parking availability at time interval t, defined as:

$$\boldsymbol{Y}_c = tanh(\boldsymbol{W}_c\boldsymbol{H}_c + \boldsymbol{b}_c) \tag{3}$$

where $\boldsymbol{W}_{hc}$, $\boldsymbol{b}_{hc}$, $\boldsymbol{W}_c$ and $\boldsymbol{b}_c$ are the parameters to be learned.

Likewise, we can construct the period component of Figure 7. Assume that there are l_p time intervals from the period fragment and the period is p. Therefore, the period dependent sequence is $\left[\boldsymbol{X}_{t-(l_p-1)\cdot p}, \boldsymbol{X}_{t-(l_p-2)\cdot p}, ..., \boldsymbol{X}_t\right]$. With the LSTM layer and concatenation layer like in Eqs. 2 and 3, the output of the period component is $\boldsymbol{Y}_p$. In the detailed implementation, p is equal to one-day that describes daily periodicity.

5.2 The Structure of General Component

Parking availability can be affected by many complex factors, such as weather, mobility trace, holidays, and event. We have analysed these features, which affects the parking availability in section 4. We stack two fully-connected layers to model the general factors that affect parking availability. The feature vector $\boldsymbol{X}_t$ is fed into the first layer, which outputs a hidden state $\boldsymbol{H}_g$ as follows:

$$\boldsymbol{H}_g = ReLU\left(\boldsymbol{W}_{hg}\boldsymbol{X}_t + \boldsymbol{b}_{hg}\right) \tag{4}$$

$\boldsymbol{H}_g$ is fed to produce parking availability at the given time interval $\boldsymbol{Y}_g$, defined as:

$$\boldsymbol{Y}_g = tanh\left(\boldsymbol{W}_g\boldsymbol{H}_g + \boldsymbol{b}_g\right) \tag{5}$$

where $\boldsymbol{W}_{hg}$, $\boldsymbol{b}_{hg}$, $\boldsymbol{W}_g$ and $\boldsymbol{b}_g$ are the learnable parameters.

5.3 Fusion

We here merge the output of the three components as shown in Figure 7, assigning different weights to different components. Finally, the softmax layer estimates the parking availability at time interval t denoted by $\hat{\boldsymbol{Y}}_t$, is defined as

$$\hat{\boldsymbol{Y}}_t = Softmax\left(\boldsymbol{V}_c \circ \boldsymbol{Y}_c + \boldsymbol{V}_p \circ \boldsymbol{Y}_p + \boldsymbol{V}_g \circ \boldsymbol{Y}_g\right) \tag{6}$$

where $\circ$ is element-wise multiplication, $\boldsymbol{V}_c$, $\boldsymbol{V}_p$ and $\boldsymbol{V}_g$ are the learnable parameters that adjust the degrees affected by closeness, period and general factors, respectively.

Our proposed model can be trained to estimate $\hat{\boldsymbol{Y}}_t$ from two sequences of temporal features and current features by minimizing the cross-entropy between the ground truth parking availability $\boldsymbol{Y}_t$ and the estimated parking availability $\hat{\boldsymbol{Y}}_t$:

$$\zeta(\theta) = -\boldsymbol{Y}_t^T log\hat{\boldsymbol{Y}}_t \tag{7}$$

where θ are all learnable parameters in our model.

5.4 Algorithm and Optimization

Algorithm 1 outlines the Du-Parking training process. We first construct the training instances from the historical sequence data. Then, Du-Parking is trained via back-propagation and Adam[6].

Algorithm 1 Du-Parking training algorithm

Input: The set of original sequences:$\{(\boldsymbol{X}_0, \boldsymbol{Y}_0), ..., (\boldsymbol{X}_{n-1}, \boldsymbol{Y}_{n-1})\}$;
lengths of closeness, period sequences: l_c , l_p;
peroid: p;
Output: Learned Du-Parking model
$D \leftarrow \emptyset$
for all available time interval $t(0 \le t \le n-1)$ **do**
  $S_c = \left[\boldsymbol{X}_{t-(l_c-1)}, \boldsymbol{X}_{t-(l_c-2)}, ..., \boldsymbol{X}_t\right]$
  $S_p = \left[\boldsymbol{X}_{t-(l_p-1)\cdot p}, \boldsymbol{X}_{t-(l_p-2)\cdot p}, ..., \boldsymbol{X}_t\right]$
  put an training instance $\left(\{S_c, S_p, \boldsymbol{X}_t\}, \boldsymbol{Y}_t\right)$ into D
end for
initialize all learnable parameters θ in Du-Parking
repeat
  randomly select a batch of instances D_b from D
  find θ by minimizing the objective (7) with D_b
until stopping criteria is met

6 EXPERIMENTS

In this section, we introduce the experimental dataset firstly, then based on the features introduced in Section 4, four groups of experiments are conducted. In the first part, we propose an effective modeling strategy which presents significant improvement in parking availability estimation. In the second part, we show the result of methods described in Section 6.2. Finally, proposed DNN models are employed to predict parking availability.

6.1 Datasets

In the evaluation, we use the following four real datasets, where these sources are available in Beijing and Shenzhen.

Real-time parking availaibility data: We employ availaibility data which collected by sensors of 2692 parking lots in Beijing and 3009 parking lots in Shenzhen from Baidu map every 30 minutes from Nov. 26 to Dec. 4, 2017. The total number of sample data is 1677314 in Beijing and 1799655 in Shenzhen.

POI-related data: A POI basic database which including POI categories, construct year etc. are employed from Baidu map. We also collect the historical map queries and calculate popularity of each POI. All POI-related data are mapped to each parking lots and stored in MySQL database.

Geolocation data: We collect GPS location data of 150 million Baidu mobile map anonymous users and three weeks duration. The historical locations are captured by Baidu mobile geolocation SDK since Nov. 24, 2017.

Navigation trajectories: We employ navigation data of 130 million Baidu map anonymous navigation users, which including navigation trajectory, timestamp and POI hashed ID from Nov. 24, 2017 to Dec. 4, 2017.

6.2 Baselines and Ground Truth

We compare our method (Du-Parking) with two baselines:

1) Linear interpolation: It is believed that the distance between two parking lots is shorter, they are more similar in parking availability[3]. This is a distance-weighted interpolation algorithm using the parking lot which has realtime availability data. as shown in Equation

8

$$g_x = \sum_i \frac{g_i \times \frac{1}{d_{xi}}}{\sum \frac{1}{d_{xi}}} \quad (8)$$

where g_i denotes the parking occupancy rate of parking lot i, d_{xi} stands for the geo-distance between parking lot x and the i-th parking place with occupancy data collected by sensors[18].

2) gradient boosting decision tree(GBDT): The reason we choose gradient boosting model as a baseline is that it has been widely adopted in many data mining competitions, and is well-known for its outstanding performance and efficiency for training. The XGBoost(eXtreme Gradient Boosting) is an open source gradient boosting library which also provides an optimized distributed version, implemented by Xboost[4]. It has been widely adopted in many data mining competitions.

Ground Truth: The total parking space data and occupancy data that retrieved from Baidu map is used as the ground truth to measure the estimation. We separate the training dataset from the test dataset by parking lot hashed ID to guarantee the parking lots of training dataset are not in the test dataset.

6.3 Results of Respective Modeling

In this part, we present an effective modeling strategy, which is inspired by the POI-related features analysis in Section 4. According to the analysis, the distinct category presents the huge characteristic differences of parking lots in availability temporal dynamics. It is easier to fit a single category than the combination of all categories, so we train a model for each POI category respectively instead of training a unified model for all POI categories.

The experimental result of two strategies based GBDT is shown in Table 3.

Table 3: Results of Respective Modeling

Categories	Respective Modeling		Unified Modeling	
	Precision	Recall	Precision	Recall
apartment	0.8560	0.8533	0.7329	0.4706
entertainment	0.8696	0.8639	0.7222	0.8065
food	0.9058	0.9147	0.7385	0.8335
hospital	0.8316	0.7861	0.5410	0.7149
mall	0.6814	0.7068	0.6536	0.6473
office	0.8274	0.8347	0.5495	0.6960
park	0.8140	0.8190	0.6915	0.7342
total	0.8402	0.8399	0.6613	0.7004

As shown in Table 3, proposed modeling strategy can significantly improve the accuracy of parking availability estimation. The following experiments are conducted based on this strategy.

6.4 Feature Importance

In order to evaluate the importance of features, six groups of experiments are conducted. We use GBDT model to study the performance of individual features and their combinations with Beijing dataset. Why we choose the GBDT is due to it's prominent performance and efficiency for classification. GBDT also has fewer parameters to adjust than DNN model. Firstly, we use only time of day feature(t) and mobility features, then we combine them in different way. The result is shown in Table 4.

Table 4: Results of Feature Importance

Features	Precision	Recall
t	0.7225	0.7390
F_{mob}	0.7043	0.7044
$F_{mob} + F_h$	0.7054	0.7045
$t + F_{poi}$	0.7324	0.7325
$t + F_{mob} + F_h + F_{poi}$	0.8257	0.8272
$t + F_{mob} + F_h + F_{poi} + F_{mete}$	0.8402	0.8399

As shown in Table 4, adding one feature into the model brings obvious improvement on both *recall* and *precision*.

6.5 Results of Baselines and Du-Parking

In this part, the Linear interpolation, GBDT algorithm and DNN models we proposed are adopted to train a classification model that is used to predict the parking availability of a parking lot. All the features are employed.

6.5.1 Results of Baselines. The experimental results of Beijing and Shenzhen are shown in Table 6 separately. From the first two rows in the table, we can see that GBDT algorithm outperforms Linear Interpolation. Both Beijing and Shenzhen present good performance on the *precision* and *recall*. Thus our features are general and robust for different cities.

6.5.2 Results of Du-Parking. The Du-Parking model with different components are described in detail in Table 5. The python libraries, including Theano [10] and Keras [5], are used to build our models. Du-Parking-G is a DNN model with two fully-connected layers with 50 neurons. Du-Parking-C and Du-Parking-P have similar network structures. They are LSTM models with a LSTM layer which has 32 neurons, following by a fully-connected layer with 16 neurons using activation function "relu". There are 3 extra hyperparamers in our Du-Parking, of which p is empirically fixed to one day. For lengths of the two dependent sequences(closeness sequences and period sequences), we set them as: $lc = 6$ and $lp = 3$. The same meaning in other situations also effect. The batch size is 100. We select 80% of the training data for training each model, and the remaining 20% is chosen as the test set.

Table 5: Model Description

Model Names	Description
Du-Parking-G	current general components
Du-Parking-C	closeness components
Du-Parking-P	period components
Du-Parking-GC	G + closeness components
Du-Parking-GCP	GC + period components

The experimental result is shown in Table 6. It shows the performance of Du-Parking in terms of the mean *precision* and mean *recall*.

As shown in Table 6, the merged DNN models, Du-Parking-GCP outperforms classifiers in baselines and other Du-Parking models.

Table 6: Results of Du-Parking

Classifiers	Beijing		Shenzhen	
	Precision	Recall	Precision	Recall
Linear Interpolation	0.7110	0.7233	0.7029	0.6806
GBDT	0.8402	0.8399	0.8322	0.8214
Du-Parking-G	0.8388	0.8389	0.8267	0.8282
Du-Parking-C	0.8408	0.8405	0.8301	0.8294
Du-Parking-P	0.8432	0.8427	0.8328	0.8305
Du-Parking-GC	0.8421	0.8413	0.8329	0.8306
Du-Parking-GCP	0.8447	0.8435	0.8315	0.8311

Compared with Du-Parking-G, Du-Parking-GC considers the closeness component. From the results, Du-Parking-GC is better than Du-Parking-G. Du-Parking-GCP which is adding the period component improves the performance than Du-Parking-GC, demonstrating the benefits of closeness and period components .

Compared with the previous state-of-the-art models, Du-Parking improves accuracy obviously in both cities, demonstrating that model we proposed has good generalization performance on different cities estimating tasks.

The results also indicate that Du-Parking-C and Du-Parking-P are both better than Du-Parking-G, pointing that temporal model is always beneficial for this problem.

6.5.3 Efficiency. Table 7 presents the online efficiency of our approach, which was tested on a 64-bit server with 8-core 1.8G CPU, 64GB RAM and NVIDIA TITAN X GPU with 12G VRAM. The batch size of prediction is 50000. The feature extraction accounts for up to 86% of total online processing time. On average, we can generate the parking availability for one time interval and entire Beijing in 4 minutes by Du-Parking model.

Table 7: Efficiency study

Procedures		Time(ms)
Feature extraction(per POI)	Online process	4.57
Inference (per POI)	Du-Parking-CPU	0.74
	Du-Parking-GPU	0.13

7 CONCLUSION

We propose a novel deep-learning-based model for estimating the parking availability, based on the parking availability reported by a limited number of online parking sensors and a variety of datasets, such as meteorological data, events, map mobility trace data and navigation data from Baidu map, and POIs. We applied it to provide the realtime parking availability information in the parking lot detail page in Baidu Map app, in nine cites in China. Experimental results on the datasets of Beijing and Shenzhen prove the effectiveness of the proposed model for parking availability estimation.

In the future, we will apply our approach to more cities and predict parking availability of next time intervals for those parking lots without parking sensors.

ACKNOWLEDGMENT

We would like to thank Zhang Chuanming and Yu Yongjian for their valuable suggestion that has helped to improve the quality of the manuscript. We would like to thank Shi Wenli, Liu Junwei to provide the experimental data for us.

We are greatly indebted to Zhu Lukun, Wu Mingjie who have helped us to develop the fundamental. At the same time, We would like to appreciate Yang Liang, Jin Jiaxing, Yang Chao, who have helped us on visualization. We also thank all the anonymous reviewers who gave helpful advice.

REFERENCES

[1] Walaa Alajali, Sheng Wen, and Wanlei Zhou. 2017. On-Street Car Parking Prediction in Smart City: A Multi-source Data Analysis in Sensor-Cloud Environment. In *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, Cham, 641–652. https://doi.org/10.1007/978-3-319-72395-2_58

[2] Y. Bengio, P. Simard, and P. Frasconi. 1994. Learning Long-term Dependencies with Gradient Descent is Difficult. *Trans. Neur. Netw.* 5, 2 (March 1994), 157–166. https://doi.org/10.1109/72.279181

[3] Fabian Bock and Monika Sester. 2016. Improving Parking Availability Maps using Information from Nearby Roads. *Transportation Research Procedia* 19 (2016), 207–214. https://doi.org/10.1016/j.trpro.2016.12.081

[4] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, and Yuan Tang. 2015. xgboost: Extreme Gradient Boosting. (2015).

[5] François Chollet et al. 2015. Keras. https://github.com/keras-team/keras. (2015).

[6] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980 http://arxiv.org/abs/1412.6980

[7] Andreas Klappenecker, Hyunyoung Lee, and Jennifer L. Welch. 2014. Finding Available Parking Spaces Made Easy. *Ad Hoc Netw.* 12 (Jan. 2014), 243–249. https://doi.org/10.1016/j.adhoc.2012.03.002

[8] Christoph Pflügler, Thomas Köhn, Maximilian Schreieck, Manuel Wiesche, and Helmut Krcmar. 2016. Predicting the availability of parking spaces with publicly available data. In *GI-Jahrestagung*.

[9] S. Pullola, P. K. Atrey, and A. El Saddik. 2007. Towards an Intelligent GPS-Based Vehicle Navigation System for Finding Street Parking Lots. In *2007 IEEE International Conference on Signal Processing and Communications*. 1251–1254. https://doi.org/10.1109/ICSPC.2007.4728553

[10] Ying Zhang Rami Al-Rfou, Guillaume Alain. 2016. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints* abs/1605.02688 (May 2016). http://arxiv.org/abs/1605.02688

[11] Felix Richter, Sergio Di Martino, and Dirk C. Mattfeld. 2014. Temporal and Spatial Clustering for a Parking Prediction Service. In *Proceedings of the 2014 IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI '14)*. IEEE Computer Society, Washington, DC, USA, 278–282. https://doi.org/10.1109/ICTAI.2014.49

[12] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14)*. MIT Press, Cambridge, MA, USA, 3104–3112. http://dl.acm.org/citation.cfm?id=2969033.2969173

[13] Tim Tiedemann, Thomas VÃűgele, Mario Michael Krell, Jan Hendrik Metzen, and Frank Kirchner. 2015. Concept of a Data Thread Based Parking Space Occupancy Prediction in a Berlin Pilot Region. In *Papers From the 2015 AAAI Workshop. Workshop on Ai for Transportation.*

[14] Eleni I. Vlahogianni, Konstantinos Kepaptsoglou, Vassileios Tsetsos, and Matthew G. Karlaftis. 2016. A Real-Time Parking Prediction System for Smart Cities. *Journal of Intelligent Transportation Systems* 20, 2 (2016), 192–204. https://doi.org/10.1080/15472450.2015.1037955

[15] Zhengwei Wu, Haishan Wu, and Tong Zhang. 2015. Predict User In-World Activity via Integration of Map Query and Mobility Trace. (2015).

[16] Junbo Zhang, Yu Zheng, and Dekang Qi. 2016. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. *CoRR* abs/1610.00081 (2016). arXiv:1610.00081 http://arxiv.org/abs/1610.00081

[17] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, and Xiuwen Yi. 2016. DNN-based Prediction Model for Spatio-temporal Data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '16)*. ACM, New York, NY, USA, Article 92, 4 pages. https://doi.org/10.1145/2996913.2997016

[18] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-Air: When Urban Air Quality Inference Meets Big Data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*. ACM, New York, NY, USA, 1436–1444. https://doi.org/10.1145/2487575.2488188

[19] Yanxu Zheng, S Rajasegarar, and C Leckie. 2015. Parking availability prediction for sensor-enabled car parks in smart cities. In *IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. 1–6.