

Free-Floating Bike-Sharing Demand Prediction with Deep Learning

Ziyang Zhang, Lingye Tan, and Weiwei Jiang

Abstract—As a solution to the last mile problem in big metropolitan cities, free-floating bike-sharing service is becoming a new choice for short travels all over the world. Unlike the docked bikes which requires the users to borrow and return at fixed stations, free-floating bikes can be used everywhere. However, this feasibility also brings a higher management cost. The bikes should be scheduled from the regions with less demand to those with higher demand, based on a precise demand prediction. In this paper, we use deep learning techniques including Multi-Layer Perceptron and ConvLSTM networks for this task. We find that in the case of the insufficient training data, e.g., one-month data of Mobike, Multi-Layer Perceptron performs better than both ConvLSTM and two simple historical methods.

Index Terms—Free-floating bike-sharing, demand prediction, deep learning.

I. INTRODUCTION

In modern cities, automobiles have become common, but the congestion problem of urban roads has become increasingly serious. More and more commuters choose public transportation to travel, saving time and economic costs. However, in many cases, the destination is still a certain distance away from the bus station, subway station, etc., and the commuter needs to walk for a while. To fill in this gap, free-floating shared bicycles appeared in 2014. It has the advantages of scanning the QR code to ride the bicycle, convenient borrowing and return, and low price. Compared with traditional public bicycles with stations, it takes into account the needs of users at any location, which greatly solves the last mile needs of commuters in a real sense. According to statistics released by the Beijing Municipal Commission of Transportation, as of the end of 2019, the total number of shared bicycles in Beijing has stabilized at around 900,000. The average daily number of shared bicycles in Beijing is 1.272 million, and each bicycle will be used on an average of 1.4 times per day. According to statistics released in Shenzhen, the number of shared bicycles in operation in the city in 2019 was 480,000, and the average daily usage was about 849,000.

The excessive amounts of different free-floating shared bikes bring the problem of urban management. In order to

seize market share, bicycle-sharing companies aggressively invested and dropped new bikes on the streets, but neglected offline operation and maintenance management, which caused chaos such as random parking. Problems such as crowding of public entrances and exits, sidewalks, blind lanes, and non-motorized vehicle lanes such as subways have occurred. People can also often see that bicycles of various colors are discharged in patches or swaying around. Some bicycles have been left unattended for too long and are covered with dust. In order to solve this problem, the mainstream idea is to control the amount of shared bicycles. This also makes accurate prediction of changes in the demand for bicycles and bicycle scheduling become more important. Regarding how to dispatch ys, which areas have a large flow of people and high demand, when the shared bicycle industry first appeared, dispatchers used some existing experience to predict and dispatch. However, as the shared bicycle industry is increasing, deep learning and machine learning models are widely used. Now forecasting is increasingly based on data-driven methods, rather than human experience.

In this paper, we use deep learning techniques for free-floating bike-sharing demand prediction, which is important to the efficiency of the transportation system. Specifically, we compare the performance of a MLP model, a ConvLSTM model and two simple historical methods. We use a real-world shared bike usage dataset from Mobike to conduct experiments. The dataset contains 3765364 orders from May 10, 2017 to May 31, 2017 in Beijing, China. Our results show that MLP achieves a better prediction performance, both than two simpler methods and a more complex model, when the dataset used is insufficient, i.e., less than a month.

The following of this paper is organized as follows. In Section II, we review some latest related work. In Section III, we state the problem formulation. In Section IV, we describe the dataset and the preprocessing steps. In Section V, we describe the models we use. In Section VI, we show our experiments. In Section VII, we draw our conclusion.

II. RELATED WORK

Shared bicycles have recently become a very popular research object. Researchers in many countries have conducted extensive studies on the time and space distribution of shared bicycles, which can be roughly divided into two categories, prediction based on machine learning models and model prediction based on deep learning. Deep learning is successful in different fields, including computer vision problems [1]-[3] and time series related problems [4], [5]. Deep learning is also being used in the transportation for traffic forecasting [6]. In this section,

Manuscript received September 8, 2020; revised April 23, 2021.

Ziyang Zhang and Lingye Tan are with the Faculty of Engineering, University of New South Wales, Sydney, 2032, Australia (e-mail: zzy967230@163.com, tanlingyelynn@163.com).

Weiwei Jiang is with the Department of Electronic Engineering, Tsinghua University, Beijing, People's Republic of China (Corresponding author; e-mail: jwwthu@163.com).

both machine-learning based prediction and deep learning based predictions for bike-sharing demand are reviewed.

A. Machine Learning Based Prediction

In [7], the authors proposed a two-part clustering algorithm, which clusters bicycle stations into groups. The Gradient Boosted Regression Tree (GBRT) predicts the total number of bicycles that will be rented out in New York and Washington. Then, based on multi-similarity, the authors propose an inference model to calculate the rent ratio between clusters.

In Ref. [8], the authors used the two latest high-efficiency models, LSTM and GRU, to predict the short-term available number of bicycles in Suzhou through one-month historical data. Random forest is used for comparison as a benchmark. The results show that both RNN (LSTM and GRU) and random forest can achieve good performance with acceptable errors and relative accuracy. In terms of training time, random forest is more advantageous, while LSTM with complex structure can be more accurate and long-term prediction.

In Ref. [9], the researchers found that the prediction mechanism of fuzzy inference can well capture the highly variable trend of shared bicycle usage. The Wang-Mendel rule generation method is used to generate a rule base, and then only current information such as date-related information and weather conditions are used to predict the bicycle share demand at any given point in the future. The simulation results show that the fuzzy inference predictor may be better than the traditional feedforward neural network in terms of prediction accuracy.

Artificial immune system (AIS) and regression tree (RT) is combined in [10] for bicycle sharing system (BSS). Cells in AIS are the basic components. The model embeds the RT predictor model into AIS to form a cell bank, and uses a clone selection mechanism to generate cloned antibodies.

B. Deep Learning Based Prediction

In Ref. [11], the authors proposed a novel data-driven spatiotemporal graph attention convolutional neural network for bicycle station-level traffic prediction (Gbikes), and designed a novel attentional convolutional neural network (GACNN). It has an attention mechanism to capture and distinguish the correlation between stations, which improves the effectiveness and accuracy. At the same time, the researchers conducted extensive experiments on three large bike sharing systems in New York, Chicago, and Los Angeles with a total of 11 million trips.

In Ref. [12], the author proposed a new graph convolutional neural network with a data-driven graph filter (GCNN-DDGF) model, and explored two architectures of the GCNN-DDGF model. In addition, the author also proposed four types of GCNN models, including spatial distance matrix (SD), demand matrix (DE), average travel duration matrix (ATD) and demand correlation matrix (DC).

In Ref. [13], five architectures for implementing RNN are provided and compared with four evaluation indicators: average absolute percentage error, root mean square logarithmic error, mean absolute error and root mean square error to predict site-level pickup demand for shared bicycles.

In Ref. [14], it is proposed to develop a GCN-based

station-level bicycle usage hourly demand forecasting architecture, which uses two graphical structures (GCN-IDW and gcnup) to reflect different spatial characteristics and compare performance.

In Ref. [15], the authors used a novel spatio-temporal graph convolutional network (STGCN) to predict Wenling's pick-up demand for shared bicycles by exploring potential information from multiple demand points. At the same time, the graph convolutional neural network (CNN) is used to express the spatial dependency; in addition, according to the time series data representing the demand, the gated CNN is used to express the time correlation for picking up/returning the public bicycle. After comparing the results, STGCN consumes longer training time, but it needs the least time period to achieve convergence accuracy.

In Ref. [16], in the first stage, the authors established a spatio-temporal graph neural network (ST-GNN) model to predict bicycle demand throughout New York, while capturing spatial correlation and temporal dependence in a unified network architecture. In the second stage, the truck-based station rebalancing problem is formulated as an optimization problem with transportation cost targets, and the integer linear programming (ILP) algorithm is used to effectively solve the problem.

In Ref. [17], the authors proposed a new multi-graph convolutional neural network model to predict station-level bicycle traffic in a bicycle sharing system. The authors designed three different inter-station graphs to represent the bicycle sharing system, namely distance, interaction and correlation graphs; then proposed a fusion method to perform graph convolution operations on the three graphs at the same time.

In Ref. [18], specifically, the authors integrate CNN and GRU-Net into the structure to represent the influence of external variables on space and time, and summarize ConvGRU-Net to understand the temporal and spatial dependence of the use of shared bicycles. Based on the effectiveness of the MBH model, the authors divided the four data sets by 15, 30, 45, and 60 minutes. The comparison results show that 30 minutes is the best time interval for realizing bicycle sharing supply and demand prediction.

III. PREDICTION PROBLEM FORMULATION

In this section, we formulate the free-floating bike-sharing prediction problem as a regression problem, similar to [19]. We would divide the spatial region into M by N grids and divide the temporal range into K time slots. We denote $D(i, j, k)$ as the number of orders in the grid (i, j) and time slot k . The prediction problem is to predict $D(i, j, k+1)$ for all i and j , given historical usage logs before time slot k .

In our problem formulation, the historical data are the main input for prediction. In previous studies, the inclusion of time of the day or day of the week have been proven useful for improving model performance. Also, the meteorology factors, e.g., weather and air quality, could be influential in traveling. We plan to add these factors into consideration in the future work, when relevant data are available.

IV. DATASET DESCRIPTION AND PREPROCESSING

A. Dataset Description

In this paper, we use a real-world bike-sharing usage dataset provided by Mobike, which used to one of the largest free-floating bike-sharing company in China. The dataset contains the bike usage orders for 22 days in Beijing, China. The usage of real-world transportation data has been proven important and necessary in previous studies [20-22].

Each order contains the following fields: order id, departure time, geohased origin location. We decode the geohased string back to longitude and latitude for further usage. Totally, we use 3,765,364 orders for this paper.

The spatial region of the dataset is within the longitude range from 116.21 to 116.55 and the latitude range from 39.76 to 40.03. This spatial range is basically the size of Beijing within the Fifth Ring Road. We divide the spatial region with 20 by 20 grids.

The temporal range of the dataset lasts from May 10, 2017 to May 31, 2017. We divide the temporal range by 1 hour as the time slot. In total, we have 528 time slots, which corresponds to data from 22 days. We show the order statistics for each hour in May 10, 2017 and May 11, 2017 in Fig. 1 and Fig. 2. As we can tell from these figures, the share bike demand presents a periodic pattern, which can be further used for prediction. We can also find that free-floating bikes are used heavily by commuters, when there are two peaks during the morning and evening rush hours.

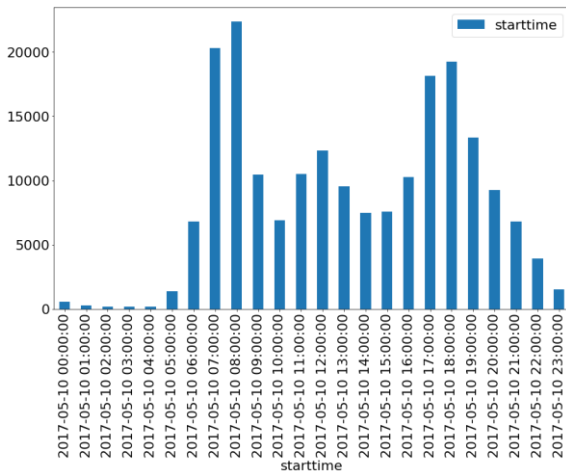


Fig. 1. Order statistics for each hour in May 10, 2017.

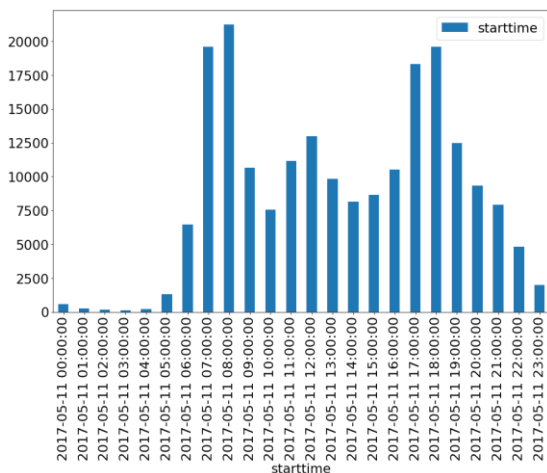


Fig. 2. Order statistics for each hour in May 11, 2017.

B. Dataset Preprocessing

We aggregate the order data into a matrix in 20 by 20 spatial grids and 528 time slots. For each element of the matrix, it represents the number of orders with the start location within the specific spatial grid and the start time in the time slot. We show the distribution of the matrix element values in Fig. 3.

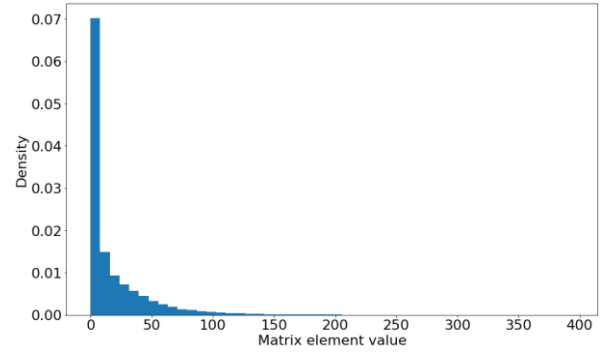


Fig. 3. The distribution of the aggregated order data.

The aggregated data follows a long-tail distribution, which is not suitable for machine learning or deep learning models. We use a log transformation to the data with $\log(i+1)$, where i is the original value, as our preprocessing step. For evaluation, we would transform the data back with the reverse function. The distribution of the transformed element values is shown in Fig. 4. The value range becomes much smaller and the tail is so obvious in Fig. 4.

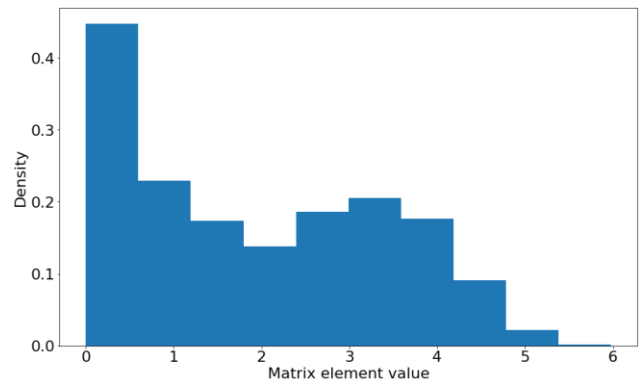


Fig. 4. The distribution of the transformed aggregated order data.

V. MODELS

In this study, we propose to use a simple MLP model to solve the free-floating bike-sharing demand prediction. We use the ConvLSTM in [19] and two simple historical methods as baselines.

The MLP model is a typical structure of deep neural networks. It contains at least one hidden layer and use activation functions for non-linear feature learning. MLP are fully connected and we use a four-layer MLP model in this study. Dropout is not used in this study.

For the baselines, we use a ConvLSTM model in [19] and the extreme gradient boosting (XGBoost) model in [23]. The first simple historical method uses the demand value from the last hour as the prediction and the second simple historical method uses the demand value the same hour from the previous day as the prediction.

We denote the two simple historical methods as

HIST_HOUR and HIST_DAY, the MLP model as MLP, the XGBoost model as XGBoost, and the ConvLSTM model as ConvLSTM.

VI. EXPERIMENTS

A. Parameter Settings

We use a four-layer densely connected MLP in this study. We show the specific model structure of MLP used in this study in Fig. 5. Notice that with the change of input historical length, the number of trainable parameters of the MLP models would also change.

Layer (type)	Output Shape	Param #
dense (Dense)	multiple	2458112
dense_1 (Dense)	multiple	262656
dense_2 (Dense)	multiple	262656
dense_3 (Dense)	multiple	205200
Total params: 3,188,624		
Trainable params: 3,188,624		
Non-trainable params: 0		

Fig. 5. The model structure of MLP.

For ConvLSTM and MLP models, we use the historical data from the last 6, 12, 18 or 24 hours as input frames and predict the one-hour ahead frame, where frame is used to represent the matrix in a time slot. The optimizer used in this study is Adam [24] and its learning rate is set to $1e-3$. The batch size is set to 10 and the number of epochs is set to 100. Python and TensorFlow are used for all experiments.

B. Evaluation Metrics

Considering the limited data amount, we only use the last 4 days of the whole dataset as the test set and the other data as the training set. The root mean squared error (RMSE) over the test set is used as our final evaluation metric. For a better prediction performance, we want the models to achieve a lower RMSE.

TABLE I: EXPERIMENTAL RESULTS OF DIFFERENT MODELS ON THE TEST SET

Model	Input Historical Length			
	6 hours	12 hours	18 hours	24 hours
HIST_HOUR	12.7706			
HIST_DAY	16.3164			
XGBoost	11.2539	11.7700	12.0636	12.0652
ConvLSTM	10.6952	11.1844	11.7517	10.9736
MLP	8.3258	8.6910	9.8477	9.6197

C. Results

We show the results in Table I. As we can tell from Table I, a simple MLP model performs better than the complex ConvLSTM model, with the different input historical lengths. The MLP model also outperforms the two simpler historical methods. The best performance is achieved for an input historical length of 6 hours, both for MLP and ConvLSTM models. Considering the limited training data for only 18 days, using a longer input data length increases the probability of model overfitting and damages the performance on the test set. To fight against the possible

overfitting problem, more data is necessary.

Our result is different from the previous study, e.g., in [25], the authors compared different models including XGBoost, MLP and LSTM and found that LSTM outperformed other models in their study. This indicates that the different datasets may present different characteristics and a single model cannot always win on all datasets.

VII. CONCLUSION

In this paper, deep learning techniques are used for free-floating bike-sharing demand prediction, which is important to the efficient operation of this newly appeared transportation mode. Specifically, a MLP model, a ConvLSTM model and two simple historical methods are used. We use a real-world shared bike usage dataset from Mobike to conduct experiments. Our results indicate that MLP achieves a better prediction performance, both than two simpler methods and a more complex model, when the dataset used is insufficient, i.e., shorter than a month. For further research, more data is needed for training the sophisticated deep learning models, e.g., LSTM used in [25].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Ziyang Zhang concuted the research; Ziyang Zhang and Lingye Tan analyze the data and write the paper; Weiwei Jiang guided the whole process and approved the final version.

REFERENCES

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT press, 2016.
- [2] W. Jiang and L. Zhang, "Edge-SiamNet and Edge-TripleNet: New deep learning models for handwritten numeral recognition," *IEICE Transactions on Information and Systems*, vol. 103, no. 3, pp. 720-723, 2020.
- [3] K. He, X. Zhang, S. Ren *et al.*, "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
- [4] W. Jiang, "Time series classification: Nearest neighbor versus deep learning models," *SN Applied Sciences*, vol. 2, no. 4, pp. 1-17, 2020.
- [5] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," arXiv preprint arXiv: 2003.01859, 2020.
- [6] W. Jiang and L. Zhang, "Geospatial data to images: A deep-learning framework for traffic forecasting," *Tsinghua Science and Technology*, vol. 24, no. 1, pp. 52-64, 2018.
- [7] Y. Li, Y. Zheng, H. Zhang *et al.*, "Traffic prediction in a bike-sharing system," in *Proc. the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015, pp. 1-10.
- [8] B. Wang and I. Kim, "Short-term prediction for bike-sharing service using machine learning," *Transportation Research Procedia*, vol. 34, pp. 171-178, 2018.
- [9] S. M. Salaken, M. A. Hosen, A. Khosravi *et al.*, "Forecasting bike sharing demand using fuzzy inference mechanism," *International Conference on Neural Information Processing*, Springer, Cham, pp. 567-574, 2015.
- [10] J. L. Wu and P. C. Chang, "A prediction system for bike sharing using artificial immune system with regression trees," in *Proc. 2015 IIAI 4th International Congress on Advanced Applied Informatics*, IEEE, 2015, pp. 511-516.
- [11] S. He and K. G. Shin, "Towards fine-grained flow forecasting: A graph attention approach for bike sharing systems," in *Proc. The Web Conference*, 2020, pp. 88-98.

- [12] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258-276, 2018.
- [13] P. C. Chen, H. Y. Hsieh, K. W. Su *et al.*, "Predicting station level demand in a bike-sharing system using recurrent neural networks," *IET Intelligent Transport Systems*, vol. 14, no. 6, pp. 554-561, 2020.
- [14] T. S. Kim, W. K. Lee, and S. Y. Sohn, "Graph convolutional network approach applied to predict hourly bike-sharing demands considering spatial, temporal, and global effects," *PloS One*, vol. 14, no. 9, p. e0220782, 2019.
- [15] G. Xiao, R. Wang, C. Zhang *et al.*, "Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks," *Multimedia Tools and Applications*, pp. 1-19, 2020.
- [16] R. Guo, Z. Jiang, J. Huang *et al.*, "BikeNet: Accurate bike demand prediction using graph neural networks for station rebalancing," in *Proc. 2019 IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, IEEE, 2019, pp. 686-693.
- [17] D. Chai, L. Wang, and Q. Yang, "Bike flow prediction with multi-graph convolutional networks," in *Proc. the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2018, pp. 397-400.
- [18] M. Xu, H. Liu, and H. Yang, "A deep learning based multi-block hybrid model for bike-sharing supply-demand prediction," *IEEE Access*, vol. 8, pp. 85826-85838, 2020.
- [19] L. Tan and Z. Zhang, *Ride-Hailing Service Prediction Based on Deep Learning*, 2020.
- [20] W. Jiang and L. Zhang, "Evaluating the effects of double-apping on the smartphone-based E-hailing service: A simulation-based study," *IEEE Access*, vol. 6, pp. 6654-6667, 2018.
- [21] W. Jiang, J. Lian, M. Shen *et al.*, "A multi-period analysis of taxi drivers' behaviors based on GPS trajectories," in *Proc. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2017, pp. 1-6.
- [22] W. Jiang and L. Zhang, "The impact of the transportation network companies on the taxi industry: Evidence from Beijing's GPS Taxi Trajectory Data," *IEEE Access*, vol. 6, pp. 12438-12450, 2018.
- [23] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.
- [24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv: 1412.6980, 2014.
- [25] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets,"

Transportation Research Part C: Emerging Technologies, vol. 95, pp. 47-60, 2018.

Copyright © 2022 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Ziyang Zhang was born in Jinan, Shandong province China in 1996, he received the bachelor's degree in road bridge and river crossing from the School of highway, Chang'an University, Xi'an, Shanxi province, China in 2018. In 2020, he received the master's degree in civil engineering from the University of New South Wales, Sydney, Australia. He has published 3 publications in international conferences and journals. Her current research is on civil engineering.



Lingye Tan was born in Chongqing, China in 1996. She received the bachelor's degree in bridge inspection and maintenance technology from the College of Civil Engineering, Chongqing Jiaotong University, Chongqing, China in 2018. In 2020, she received the master's degree in civil engineering from the University of New South Wales, Sydney, Australia. She is applying for a PhD in Nanyang Technological University. She has published 4 publications in

international conferences and journals. Her current research interests are not only bridge but the combination with AI technology.



Weiwei Jiang was born in Jiangsu, China in 1991. He received the B.Sc. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He was also a visiting scholar in the Department of Industrial Engineering and Operations Research, University of California, Berkeley in 2017.

He is currently a postdoctoral researcher with the Department of Electronic Engineering, Tsinghua University. He has more than 10 publications, in prestigious international journals and conferences. His current research interests include the intersection between big data and machine learning techniques and signal processing applications.