

Intelligent Bike Sharing System: Usage Analysis and Prediction

Bin Xie Shimeng Jiang Liyan Fang Chenyu Shi Yuhao Lan Hongyu Ouyang

To tackle no-bike-to-borrow or no-dock-to-return problem!!!

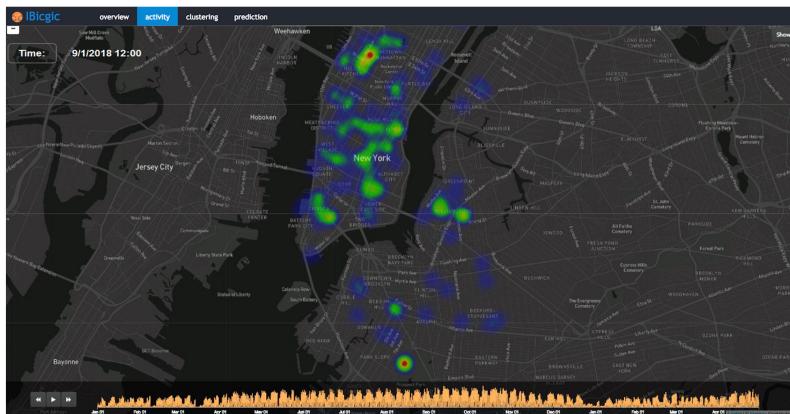


Summary

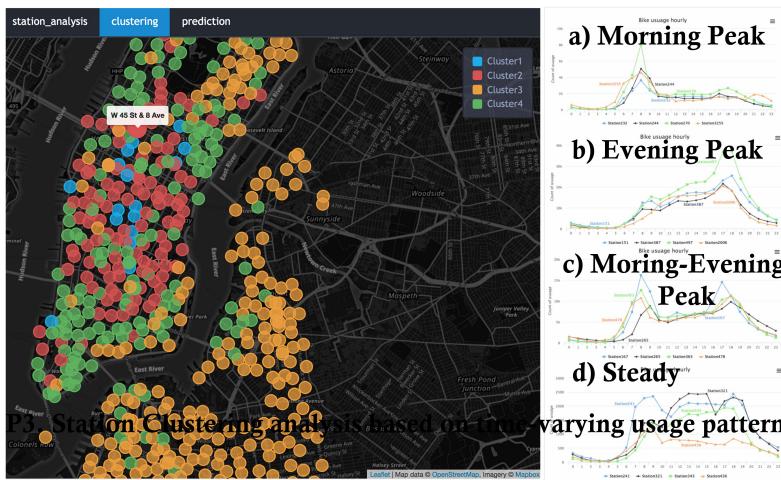
Unbalanced demand is the core problem in existing bike sharing system. Our intelligent bike sharing system is to help company runners **better planning the bike resource** which can benefit the city in traffic reduction and public health. The key function of the system is to **predict hourly bike usage for different stations** on the purpose of redistributing bikes when and where needed. Besides, the system can **present visualized summary data analysis** in an **interactive platform** for supporting daily operations, monitoring and decision-making.

System Design

Our system covers multi-angles analysis for the data. It contains the following contents: **Overview, activity, daily trip, station analysis, clustering, prediction**. Demo is shown below:



P2. Heatmap that gives the usage differences among stations.



P5. Station Cluster Analysis, different clusters shows different time-vary usage pattern.

Data

(Jan 2013 ~ Sep 2018)
65million rows of historical data (12.6G)

Citi Bike trip data download .csv Citi Bike AWS S3 holiday manually Office Holidays geographic district download shapefiles NYC.gov socio-demographic download .xlsx from NYC.gov bus station and subway download .csv from NYC.gov nearby facilities request API, response JSON Yelp weather request API, response JSON National Climate Data Center

Approaches

- Explore possible patterns of bike usage.
- Evaluate the efficacy of different predicting algorithms, choose the optimal for our system.

What's new?

- More features are included for prediction: facilities, subway, bus stations, geographic district and socio-demographic.
- Station clustering analysis to discover time-varying usage patterns.
- Algorithms with higher accuracy: NN
- Dynamic and Static Prediction can predict future usage for already exists stations and predict
- Improve visualization: provides comprehensive data analysis that enables users do timely planning.

1) Station Clustering : K-means

We do clustering analysis according to stations' hourly bike usage for exploring time-varying usage preferences so that bike resource scheduling can be performed at a cluster level. Bikes can be redistributed among stations when some are in their peak hours while others are not.

2) Bike Usage Prediction

- Dynamic Model

For existed stations, we use historical data and CNN time series model to predict its future usage.

- Static Model

Given the location of new station, we use Random Forest(which is chosen among several prediction models after evaluation) to predict its usage.