Analysis of Bike Sharing Station Placement Leveraging Heterogeneous Urban Open Data

Team Member:
Linghong Tang
Xin Su
Kaiwen Xu

Problem authors studied: Formulate the bike station placement issue as a bike trip demand prediction problem.

Purpose of authors: This method can be applied to different cities to effectively recommend places with higher potential bike trip demand for placing future bike stations.

Method: Identify the most relevant datasets from a large pool of urban open data sources based on prior knowledge, and extract customized features to characterize bike station utilization in individual cities. Then, feed these features into predictive models to rank the potential of locations for placing future bike stations.

How to work with it:

Factor analysis: area function; human activity and demographics.

Data Selection: point of Interest (POI) dataset; check-in dataset and demographics dataset.

Area Functions and Bike Trip Demand: characterize an area's functions by the categorical distribution of POIs, and analyze the correlation between area functions and BTD.

Human Activity and Bike Trip Demand: take LBSN check-ins as the semantic proxy of human activities.

Demographics and Bike Trip Demand: validate the correlation between demographic factors and BTD.

Framework:

Feature Extraction: extract a set of city-specific features for each factor based on the analysis previous.

Model Selection: need to select models that are capable of incorporating heterogeneous features to effectively predict the potential BTD of candidate areas and rank them. We adopt the regression-and-ranking methodologies.

Experiment:

Experiment Settings: collect datasets about bike station utilization and the relevant factors from Washington, D.C., USA and Hangzhou, China

Baseline Methods: use the following two baseline methods in comparison with the proposed LRR and ANNRR algorithms. (1) Nearby Station Average (NEARBY); (2) Single Data Source Static Model (SS).

Parameter Settings: use a service area radius r = 200m for Washington, D.C., and r = 250m for Hangzhou, respectively, based on the corresponding surveys.

Evaluation Metrics: compare the ranking results with the ideal ranking list R^{-}, where candidates are sorted by their actual daily bike utilization.

Experiment Results:

Their methods (ANNRR and LRR) outperform the two baseline methods in terms of top recommendation accuracy and overall ranking quality in both cities.

Conclusion:

In this paper, they leverage open data to predict bike trip demand and recommend optimal placement of bike sharing stations. They propose a two-phase feature selection method to extract customized features from heterogeneous urban open data for bike trip demand prediction. The evaluation results show that semi-supervised method outperforms the state-of-the-art baseline approaches on recommending locations for optimal bike station placement.