Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems

Longbiao Chen, Daqing Zhang, Leye Wang, Dingqi Yang, Xiaojuan Ma, Shijian Li, Gang Pan, Thi-Mai-Trang Nguyen, Jérémie Jakubowicz

Zhejiang University, Hangzhou, China
Institut Mines-Télécom, Paris, France

UbiComp’16, Heidelberg, Germany
PCC’16 Invited Presentation, Shanghai, China
August 28, 2016
Background

- **Bike Sharing Systems**: widely deployed in more than 700 cities
  - Urban sustainability, healthy lifestyle
  - Rent-Ride-Return: usually free for short trips

![Bike Sharing Systems Image]

Global expansion of bike-sharing

![Global Expansion Map]

- **EMBARQ**
The Users’ Problem: Availability

• **Over-Demand Bike Stations**
  • No bikes (empty) or docks (full) for a period of time\(^1\)
  • Possible causes: rush hours, social events, etc.
  • More than 60% users are unsatisfied with station availability in DC\(^2\)
  • Operators get fined when over-demand occurs in NYC\(^3\)

---

\(^1\) In this paper, we set the duration to 10 minutes.
\(^2\) Capital Bikeshare 2014 Member Survey Report
\(^3\) S. Kaufman, Citi Bike: The First Two Years, 2015
The Operator’s Task

- Preventing the occurrence of over-demand stations
  - **Prediction**: which station will be full/empty in the next hour?
  - **Action**: send trucks to move bikes, setup temporary docks, etc.

![Trucks moving bikes](image1)

![Temporary docks](image2)
Predicting the demand of a station is very difficult.
- densely distributed stations → ad-hoc bike usage demand
• Predicting the demand of a station is very difficult
  • densely distributed stations → ad-hoc bike usage demand
Challenge

- Predicting the demand of a station is very difficult
- densely distributed stations \(\rightarrow\) ad-hoc bike usage demand
Challenge

• Predicting the demand of a station is very difficult
  • densely distributed stations $\rightarrow$ ad-hoc bike usage demand

![Map with station clusters and bike usage graphs]

- Station 1, 8:00--9:00, #Rented
- Station 2, 8:00--9:00, #Rented
- Station 3, 8:00--9:00, #Rented

cluster

station rush hour
Challenge

- **Predicting the demand of a station** is very difficult
  - densely distributed stations → ad-hoc bike usage demand

![Map and station rush hour graphs]

Cluster of 3 Stations, 8:00--9:00, #Rented

Station 1, 8:00--9:00, #Rented

Station 2, 8:00--9:00, #Rented

Station 3, 8:00--9:00, #Rented
Basic Idea

1. Group neighboring stations into clusters
   - stations around a residential area, etc.

2. Predict bike demand of clusters
   - stable and predictable

3. Predict cluster over-demand probability
   - at least one station is over-demand
   - still useful for operators
• Obtaining demand-predictable clusters is not trivial
• Obtaining demand-predictable clusters is not trivial

A neighborhood  Morning rush hours
• Obtaining demand-predictable clusters is not trivial

A neighborhood  Morning rush hours  Basketball game
Issue

- Obtaining demand-predictable clusters is not trivial

  - A neighborhood
  - Morning rush hours
  - Basketball game

- Bike demand of stations is highly context dependent
  - common contextual factors: time, weather, etc.
  - opportunistic contextual factors: social event, traffic event, etc.
• Obtaining demand-predictable clusters is not trivial

A neighborhood  Morning rush hours  Basketball game

• Bike demand of stations is highly **context dependent**
  • **common** contextual factors: time, weather, etc.
  • **opportunistic** contextual factors: social event, traffic event, etc.

• Clusters need to be formed **dynamically** according to the context
Proposed Framework

- **Dynamic** cluster-based over-demand prediction according to context
Proposed Framework

- **Dynamic** cluster-based over-demand prediction according to context
Phase I: Context Modeling

- Understanding what contextual factors are relevant to bike demand
  - **common** factors: time, weather, etc.
  - **opportunistic** factors: social event, traffic event, etc.

- Modeling context from **urban data**
  - bike sharing system open data: \#rented, \#returned
  - weather API, social event listing, traffic alert tweets
Phase I: Context Modeling

- Common contextual factors: frequent, all stations
  - date and time vs. overall bike demand

<table>
<thead>
<tr>
<th>Table 1. Groups for modeling temporal context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day type</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Weekdays</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Weekends/Holidays</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

(a) Weekdays

(b) Weekends and holidays
Phase I: Context Modeling

- Common contextual factors: frequent, all stations
  - **date and time** vs. overall bike demand
  - weather condition and air temperature vs. hourly bike demand
Phase I: Context Modeling

- Common contextual factors: frequent, all stations
  - **date and time** vs. overall bike demand

- **weather condition and air temperature** vs. hourly bike demand

<table>
<thead>
<tr>
<th>Day type</th>
<th>Group name</th>
<th>Time span</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekdays</td>
<td>morning rush hours</td>
<td>07:00–11:00</td>
</tr>
<tr>
<td></td>
<td>day hours</td>
<td>11:00–16:00</td>
</tr>
<tr>
<td></td>
<td>evening rush hours</td>
<td>16:00–20:00</td>
</tr>
<tr>
<td></td>
<td>night hours</td>
<td>20:00–24:00</td>
</tr>
<tr>
<td>Weekends/Holidays</td>
<td>day hours</td>
<td>09:00–19:00</td>
</tr>
<tr>
<td></td>
<td>night hours</td>
<td>19:00–01:00</td>
</tr>
</tbody>
</table>

Table 2. Air temperature groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARM</td>
<td>$\geq 22^\circ C$</td>
</tr>
<tr>
<td>COMFORTABLE</td>
<td>[10$^\circ C$, 22$^\circ C$)</td>
</tr>
<tr>
<td>COOL</td>
<td>[0$^\circ C$, 10$^\circ C$)</td>
</tr>
<tr>
<td>BELOW ZERO</td>
<td>$&lt; 0^\circ C$</td>
</tr>
</tbody>
</table>

- e.g. 8:00 AM, WED, SUNNY, COMFORTABLE
Phase I: Context Modeling

- Opportunistic contextual factors: irregular, some stations
  - social event: pre- and post-event impacts
  - traffic event: post-event impacts
- Impacting zone: walking radius
- Impacting factor: event-time demand/normal demand

- e.g., 1:00 PM, RADIO CITY, ADELE CONCERT
Proposed Framework

- **Dynamic** cluster-based over-demand prediction according to context
Phase II: Dynamic Station Clustering

- Given a context, group neighboring stations with similar bike demand into clusters

8:00 AM, WED, SUNNY, COMFORTABLE

3:00 PM, STUYVESANT SQUARE, BASKETBALL GAME

- Quantifying bike demand similarity of stations
  - Common context: historical demand patterns
  - Opportunistic context: impacting zone based
Phase II: Dynamic Station Clustering

- Steps: **weighted correlation network**-based clustering

* Detailed technical contributions can be found in our paper
Phase II: Dynamic Station Clustering

- Steps: **weighted correlation network**-based clustering
  1. station network construction: nodes and links

* Detailed technical contributions can be found in our paper
Phase II: Dynamic Station Clustering

- Steps: **weighted correlation network**-based clustering
  1. station network construction: nodes and links
  2. link weight assignment: based on bike demand similarity

* Detailed technical contributions can be found in our paper
Phase II: Dynamic Station Clustering

- Steps: **weighted correlation network**-based clustering
  1. Station network construction: nodes and links
  2. Link weight assignment: based on bike demand similarity
  3. Geographically-constrained station clustering

* Detailed technical contributions can be found in our paper
Proposed Framework

- **Dynamic** cluster-based over-demand prediction according to context
Phase III: Over-Demand Cluster Prediction

- Predict which clusters are more likely to be over-demand
  1. Cluster-level bike demand estimation
     - common factors: based demand
     - opportunistic factors: inflation rate
  2. Monte Carlo simulation-based cluster over-demand prediction
     - repeat $N$ times
**Evaluation**

- **Datasets**: New York City and Washington D.C., 2014–2015
  - bike sharing data: bike trips + station status
  - meteorological data: Weather Underground API
  - social event data: Eventful API
  - traffic alert data: 511 traffic feeds + Twitter accounts

- **Plan**: 2014 data for training and 2015 data for test

- **Metrics**: precision, recall, F1-score, ROC curve
  - evaluation at cluster level

- **Baselines**

<table>
<thead>
<tr>
<th>Station Level Prediction</th>
<th>Cluster Level Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA, demand</td>
<td>static cluster</td>
</tr>
<tr>
<td>Bayesian Network, over-demand</td>
<td>common-factors only</td>
</tr>
<tr>
<td>ANN, over-demand</td>
<td>cluster + ANN, over-demand</td>
</tr>
</tbody>
</table>
Evaluation Results

- Overall scores

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision NYC</th>
<th>Recall NYC</th>
<th>F1 NYC</th>
<th>Precision DC</th>
<th>Recall DC</th>
<th>F1 DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>0.548</td>
<td>0.506</td>
<td>0.526</td>
<td>0.520</td>
<td>0.541</td>
<td>0.530</td>
</tr>
<tr>
<td>B-MC</td>
<td>0.753</td>
<td>0.656</td>
<td>0.692</td>
<td>0.636</td>
<td>0.539</td>
<td>0.583</td>
</tr>
<tr>
<td>ANN-S</td>
<td>0.776</td>
<td>0.571</td>
<td>0.658</td>
<td>0.667</td>
<td>0.428</td>
<td>0.521</td>
</tr>
<tr>
<td>SC-MC</td>
<td>0.790</td>
<td>0.647</td>
<td>0.711</td>
<td>0.793</td>
<td>0.821</td>
<td>0.807</td>
</tr>
<tr>
<td>CCF-MC</td>
<td>0.833</td>
<td>0.832</td>
<td>0.828</td>
<td>0.815</td>
<td>0.880</td>
<td>0.846</td>
</tr>
<tr>
<td>ANN-C</td>
<td>0.673</td>
<td>0.852</td>
<td>0.752</td>
<td>0.857</td>
<td>0.600</td>
<td>0.706</td>
</tr>
<tr>
<td>WCN-MC</td>
<td>0.882</td>
<td>0.938</td>
<td>0.909</td>
<td>0.857</td>
<td>0.923</td>
<td>0.889</td>
</tr>
</tbody>
</table>

(a) Neighborhood threshold
(b) ROC curves
Case Studies

- Case 1: clusters under different common contextual factors
  - Weekday morning rush hours
    - Empty: Penn Station, Brooklyn Heights, etc.
    - Full: Wall Street, Union Square, etc.
  - Sunny spring weekend
    - Full: Central Park, Union Square Park, Battery Park, etc.
  - Useful for operators to schedule bikes among areas
Case Studies

- Case 2: clusters under a social event
  - NYC Summer Streets Festival (08/08/2015)
    - Events: bike tours, block parties, and street markets
    - Empty: from Central Park to City Hall (along Park Avenue)
    - Full: Union Square Greenmarket
  - useful for operator to set up temporary bike corrals
Conclusion and Thank you

- **Take-away message**
  - Over-demand problem in bike sharing systems
  - Dynamic cluster-based over-demand prediction
  - Real-world evaluation in NYC and DC
Conclusion and Thank you

- Take-away message
  - Over-demand problem in bike sharing systems
  - Dynamic cluster-based over-demand prediction
  - Real-world evaluation in NYC and DC

- Advertisement: Harry Potter Series Bike Sharing System Series
  - Bike Sharing System and Placement Planing (UbiComp’15)
  - Bike Sharing System and Social Event Detection (Trans. HMS)
  - Bike Sharing System and Over-Demand Prediction (UbiComp’16)
Conclusion and Thank you

• Take-away message
  • Over-demand problem in bike sharing systems
  • Dynamic cluster-based over-demand prediction
  • Real-world evaluation in NYC and DC

• Advertisement: Harry Potter Series Bike Sharing System Series
  • Bike Sharing System and Placement Planing (UbiComp’15)
  • Bike Sharing System and Social Event Detection (Trans. HMS)
  • Bike Sharing System and Over-Demand Prediction (UbiComp’16)

Longbiao CHEN

http://longbiaochen.com