Time Series Data Clustering of Minnesota Bike Sharing System and Operation Strategy

Yufeng Zhang
zhan4879@umn.edu
Department of Civil, Environmental, and Geo-Engineering
Outline

1. Introduction
2. Bike Station Clustering
3. Rebalancing Vehicle Routing Strategy
4. Conclusions
Introduction

- Bike Sharing System in MN: Nice Ride
- 202 (190) stations
- 430,210 trips in 2016
Goal and Methods

• Goals
1) To identify usage patterns in BSS
2) To find a routing strategy for rebalancing vehicle

• Methods
1) Bike station clustering using Poisson Mixture Model
2) Vehicle routing optimization problem for finding the routing strategy
Station Clustering

• Input: time-series data

\[ X_{sdh}^{out} \]: departure count for station \( s \in \{1, \cdots, S\} \) during day \( d \in \{1, \cdots D\} \) and at hour \( h \in \{1, \cdots, 24\} \)

\[ X_{sdh}^{in} \]: arrival count for station \( s \in \{1, \cdots, S\} \) during day \( d \in \{1, \cdots D\} \) and at hour \( h \in \{1, \cdots, 24\} \)

A vector \( X_{sd} = (X_{sd1}^{in}, \cdots, X_{sd24}^{in}, X_{sd1}^{out}, \cdots, X_{sd24}^{out}) \) describes the arrival and departure activity of station \( s \) during day \( d \)

• Assumption: \( X_{sdt} | \{ Z_{sk} = 1, W_{dl} = 1 \} \sim \text{Poisson}(\alpha_s \lambda_{klt}) \)

• Output: Pick up and drop off rates on both weekend and weekday, \( \lambda_{klt} \)

• EM algorithm
Clustering Results (1)

• Weekend Destination

10 stations
Clustering Results (2)

- Bimodal

97 stations
Clustering Results (3)

- **Single Peak**

83 stations
Rebalancing vehicle routing strategy

• Mathematical formulation

\[
\begin{align*}
\text{Min} & \quad \sum_{u} \sum_{i,j} c_{ij} x_{ij}^{u} - \alpha \sum_{s} D S_{s} \\
\text{s.t.} & \quad \sum_{j} x_{ij}^{u} - \sum_{j} x_{ij}^{v} = \begin{cases} -1 & i = R \\
0 & i \in S \setminus \{R, S\} \\
1 & i = S \end{cases}, \quad \forall u \in U \\
y_{ij}^{u} & \leq CV^{u} x_{ij}^{u} \quad \forall (i,j) \in A, \forall u \in U \\
L_{i}^{u} & \leq \min\{\sum_{j \in S} y_{ij}^{u}, \bar{n}\} \quad \forall i \in S, \forall u \in U \\
\sum_{v \in S} L_{i}^{u} & \leq CS_{i} - I_{i} - D_{i} \quad \forall i \in S \\
U_{i}^{u} & \leq \min\{CV^{u} - \sum_{j \in S} y_{ij}^{u}, \bar{n}\} \quad \forall i \in S, \forall u \in U \\
\sum_{v \in V} U_{i}^{u} & \leq I_{i} + D_{i} \quad \forall i \in S \\
L_{i}^{u} - U_{i}^{u} & = \sum_{j \in S} y_{ij}^{u} - \sum_{j \in S} y_{ij}^{v} \quad \forall i \in S, \forall u \in U \\
d_{ij}^{u} & \leq T x_{ij}^{u} \quad \forall (i,j) \in A, \forall u \in U \\
\sum_{j \in S} d_{ij}^{u} - \sum_{j \in S} d_{ij}^{v} & = \sum_{j \in S} x_{ij}^{u} t_{ij} + (L_{i}^{u} + U_{i}^{u}) h \quad \forall i \in S, \forall u \in U \\
D S_{s} & \leq \min\{0, I_{s} + D_{s}\} \quad \forall s \in S \\
x_{ij}^{u} & \in \{0, 1\} \quad \forall (i,j) \in A, \forall u \in U \\
y_{ij}^{u}, I_{i}, L_{i}^{u}, U_{i}^{u} & \in Z_{+} \quad \forall (i,j) \in A, \forall i \in S, \forall u \in U \\
d_{ij}^{u} & \in \mathbb{R}_{+} \quad \forall (i,j) \in A, \forall u \in U
\end{align*}
\]
Rebalancing vehicle routing strategy

- Rebalancing vehicle travelling among stations to pick up and drop off bikes to meet the demand
- Decide the route that minimizes travel cost but satisfy demands
- Problem size:
  12,000 nodes
  420,000 links
  1 hour planning horizon
Routing Results (1)

- Routing- Time Expanded Network
Routing Results (2)

- **Routing - Physical Network**

<table>
<thead>
<tr>
<th>Station ID</th>
<th>Demand</th>
<th>Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>179</td>
<td>-6</td>
<td>6</td>
</tr>
<tr>
<td>95</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>192</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>41</td>
<td>-3</td>
<td>6</td>
</tr>
<tr>
<td>50</td>
<td>-5</td>
<td>6</td>
</tr>
<tr>
<td>173</td>
<td>-2</td>
<td>6</td>
</tr>
<tr>
<td>106</td>
<td>-6</td>
<td>6</td>
</tr>
<tr>
<td>166</td>
<td>-3</td>
<td>6</td>
</tr>
<tr>
<td>136</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>137</td>
<td>-5</td>
<td>6</td>
</tr>
<tr>
<td>53</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>104</td>
<td>-4</td>
<td>8</td>
</tr>
<tr>
<td>44</td>
<td>-1</td>
<td>6</td>
</tr>
<tr>
<td>93</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>105</td>
<td>-7</td>
<td>7</td>
</tr>
<tr>
<td>158</td>
<td>-4</td>
<td>6</td>
</tr>
<tr>
<td>39</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>164</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>183</td>
<td>-7</td>
<td>7</td>
</tr>
<tr>
<td>175</td>
<td>-5</td>
<td>6</td>
</tr>
<tr>
<td>163</td>
<td>-3</td>
<td>6</td>
</tr>
<tr>
<td>142</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>-4</td>
<td>4</td>
</tr>
<tr>
<td>167</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>67</td>
<td>-6</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>115</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>35</td>
<td>-5</td>
<td>6</td>
</tr>
</tbody>
</table>
Conclusion

• Bike stations are clustered into 3 categories based on its usage profile
• Using estimated demand from clustering, the vehicle routing problem gives optimal rebalancing routes that meet the demands and at the same time lowers the travel cost.
Thanks!