Diagnosing New York City’s Noises with Ubiquitous Data

Dr. Yu Zheng yuzheng@microsoft.com
Lead Researcher, Microsoft Research
Chair Professor at Shanghai Jiao Tong University
Background

• Many cities suffer from noise pollutions
  – Traffic, loud music, construction, AC...
  – Compromise working efficiency
  – Reduce sleep quality
  – Impair both physical and mental health
  – ...

• Urban noise is difficult to model
  – Change over time very quickly
  – Vary by location significantly
  – Depends on sound level and people’s tolerance
  – The composition of noises is hard to analyze

311 in NYC

• 311 Data
  – A platform for citizen’s non-emergent complaints
  – Associated with a location, timestamp, and a category
  – Human as a sensor → crowd sensing
  – Implies people’s reaction and tolerance to noises
**Goal**

- Reveal the noise situation of each region in each hour
  - A noise indicator denoting the noisy level
  - Composition of noises in each location
Methodology

- Partition NYC into regions by major roads

A) Raster-based map  B) Dilation operation  C) Thinning operation
• Build a 3D tensor to model the noises

\[ \mathcal{L}(S, R, C, T) = \frac{1}{2} \| \mathcal{A} - S \times_R R \times_C C \times_T T \|^2 + \frac{\lambda}{2} (\|S\|^2 + \|R\|^2 + \|C\|^2 + \|T\|^2) \]
311 in NYC

• Data sparseness
  – 90% regions receive $\leq 1$ complaint per day on weekdays
  – 75% regions receive $\leq 1$ complaint per day on weekends

---

Figure 6. Proportion of regions with complaints smaller than a number

Figure 7. The proportion of regions with complaints of a noise category

68 weekdays
Methodology

POIs and Road Networks

Check-ins

\[ \mathcal{L}(S,R,C,T,U) = \frac{1}{2} \| A - S \times_R R \times_C C \times_T T \|^2 + \frac{\lambda_1}{2} \| X - RU \|^2 + \frac{\lambda_2}{2} \text{tr}(C^T L Z C) + \frac{\lambda_3}{2} \| Y - TR^T \|^2 + \frac{\lambda_4}{2} (\| S \|^2 + \| R \|^2 + \| C \|^2 + \| T \|^2 + \| U \|^2) \]
\[ \mathcal{L}(S, R, C, T, U) = \frac{1}{2} \| \mathbf{A} - S \times_R R \times_C C \times_T T \|^2 + \frac{\lambda_1}{2} \| X - RU \|^2 + \frac{\lambda_2}{2} \text{tr}(C^T L_Z C) + \frac{\lambda_3}{2} \| Y - TR^T \|^2 + \frac{\lambda_4}{2} \| S \|^2 + \| R \|^2 + \| C \|^2 + \| T \|^2 + \| U \|^2 \]

\[ Y = T \times R^T \approx L \]

\[ \frac{1}{2} \sum_{i,j} \| c_i - c_j \|_2 Z_{ij} = \cdots \]

\[ = \text{tr}(C^T (D - Z)C) = \text{tr}(C^T L_Z C) \]
Methodology

• Geographical Features
  – Road networks
    • Number of intersections $f_s$
    • Length of road segments in different levels $f_r$
  – POIs
    • Total number of POIs $f_n$ and density of POIs $f_d$
    • Distribution over different categories $f_c$

A) Loud talking  B) POI: Food  C) Loud Music  D) POI: Entertainment
Methodology

- **Check in data in NYC**

- **Correlation with noises**

- **Y implies**
  - correlation between different time slots
  - correlation between different regions

\[ Y = \begin{bmatrix} r_1 & r_2 & \cdots & r_i & \cdots & r_N \\ t_1 \\ t_2 \\ \vdots \\ t_k \\ t_L \end{bmatrix} \begin{bmatrix} d_{11} & \cdots & d_{ki} & \cdots & d_{LN} \end{bmatrix} \]
Methodology

- Correlation between different noise categories

<table>
<thead>
<tr>
<th>Categories</th>
<th>%</th>
<th>Categories</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$. Loud Music/Party</td>
<td>42.2</td>
<td>$c_8$. Alarms</td>
<td>1.7</td>
</tr>
<tr>
<td>$c_2$. Construction</td>
<td>17.2</td>
<td>$c_9$. Private carting noise</td>
<td>0.8</td>
</tr>
<tr>
<td>$c_3$. Loud Talking</td>
<td>14.6</td>
<td>$c_{10}$. Manufacturing</td>
<td>0.3</td>
</tr>
<tr>
<td>$c_4$. Vehicle</td>
<td>13.7</td>
<td>$c_{11}$. Lawn care equipment</td>
<td>0.3</td>
</tr>
<tr>
<td>$c_5$. AC/Ventilation</td>
<td>3.9</td>
<td>$c_{12}$. Horn Honking</td>
<td>0.2</td>
</tr>
<tr>
<td>$c_6$. Banging/Pounding</td>
<td>2.1</td>
<td>$c_{13}$. Loud Television</td>
<td>0.1</td>
</tr>
<tr>
<td>$c_7$. Jack Hammering</td>
<td>2.1</td>
<td>$c_{14}$. Others</td>
<td>0.8</td>
</tr>
</tbody>
</table>

$Z = \begin{bmatrix} c_1 & c_2 & \cdots & c_j & \cdots & c_M \\ c_1 & & & & & \\ c_2 & c_{11} & & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ c_j & & & & c_{jj} & \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & c_{MM} \end{bmatrix}$

A) weekday  
B) weekend
Experiment

• Datasets

<table>
<thead>
<tr>
<th>Data sets</th>
<th>Period</th>
<th>Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>311 noise data</td>
<td>5/23/2013-1/31/2014</td>
<td>67,378; 14 categories</td>
</tr>
<tr>
<td>Foursquare</td>
<td>4/24/2009-10/13/2013</td>
<td>173,275</td>
</tr>
<tr>
<td>POIs</td>
<td>2013</td>
<td>26,884; 15 categories</td>
</tr>
<tr>
<td>Road Network</td>
<td>2013</td>
<td>87,898 nodes, 91,649 edges</td>
</tr>
</tbody>
</table>
Evaluation of Tensor

- **Accuracy of the inferences**
  - Remove 30% non-zero entries
  - Metrics: RMSE & MAE

<table>
<thead>
<tr>
<th>Methods</th>
<th>Weekdays</th>
<th>Weekends</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>AWR</td>
<td>4.736</td>
<td>2.582</td>
</tr>
<tr>
<td>AWH</td>
<td>4.631</td>
<td>2.461</td>
</tr>
<tr>
<td>MF</td>
<td>4.600</td>
<td>2.474</td>
</tr>
<tr>
<td>Kriging</td>
<td>4.59</td>
<td>2.424</td>
</tr>
<tr>
<td>TD</td>
<td>4.391</td>
<td>2.381</td>
</tr>
<tr>
<td>TD+ X</td>
<td>4.285</td>
<td>2.279</td>
</tr>
<tr>
<td>TD+ X + Y</td>
<td>4.160</td>
<td>2.110</td>
</tr>
<tr>
<td>TD+ X + Y + Z</td>
<td>4.010</td>
<td>2.013</td>
</tr>
</tbody>
</table>

Yu Zheng, et al. [Diagnosing New York City’s Noises with Ubiquitous Data](#). UbiComp 2014.
Experiments

• Relative ranking performance
  – Ranked by the inferred noise indicators
  – Ground truth: measured by a mobile phone running a client program
  – Metric: NDCG
311 in NYC

- Correlation between 311 complaints and real noise levels
  - Measured the real noise levels of 36 locations in Manhattan
  - People’s tolerances vary in time of day

A) Locations
B) Correlation in 6am-6pm
C) Correlation in 7pm-11pm
Results

• 311 vs. Inferences
  – 2.75% of entries on weekdays are from 311 data (97.25% by inference)
  – 1.83% of entries on weekends are from 311 data (98.17% by inference)

A) Vehicles (6am-6pm)  B) Loud Talking (0-5am)
Results

• Overall noise situation in NYC
  – Weekdays and weekends
  – Different time of days

• Example
  – Wall Street, central park,
  – Time square, NYU, Columbia
Results

- Noises of particular categories

Conclusion

• 311 data helps reveal the noise situation of the city but not accurately enough
• Fusing other data sources is important
• Spatio-temporal correlation between noises of different locations and time slots
New York City's Noise

Data Panel
- Weekday
- All Categories
- Noise Layer

Time Panel
0:00 ~ 22:00

Top 5 Noisiest Regions
1. Wall Street: 179.01
2. Columbia University: 132.15
3. East Village & Stuyvesant Town: 100.87
4. Washington Heights: 95.21
5. Audubon Ave: 93.95

New York University Noise Analysis

This demo is based on 311 data from May 23, 2013 to Jan. 31, 2014.
New York City's Noise

Data Panel
- Weekday
- All Categories
- Noise Layer

Time Panel
- 0:00 ~ 22:00

Top 5 Noisiest Regions
1. Wall Street: 179.01
2. Columbia University: 132.15
3. East Village & Stuyvesant Town: 100.87
4. Washington Heights: 95.21
5. Audubon Ave: 93.95

New York University Noise Analysis

POI:
- Business to Business: 4
- Education: 4
- Food & Dining: 8
- Government: 3
- Health & Beauty: 9
- Home & Family: 7
- Legal & Finance: 2
- Estate & Construction: 4
- Shopping: 3
- Travel: 2

Check in: 424

Road Networks:
- Number of intersections: 196
- Length of roads (km):
  - Level 1&2: 0
  - Level 3: 0
  - Level 4: 0.87
  - Level 5: 4.15
  - Level 6: 16.25

The inferred noise in region 999 is 69.1
Weekday at 22:00
New York University

Add to Compare

This demo is based on 311 data from May 23, 2013 to Jan. 31, 2014.
Region Noise Comparison (Weekday)

New York University Noise Analysis (%)
- Loud Music/Party 30.4%
- Construction 13.7%
- AC/Ventilation 8.1%
- Loud Talking 7.8%
- Vehicle 6.2%

Columbia University Noise Analysis (%)
- Loud Music/Party 25.5%
- Loud Talking 13%
- Vehicle 12%
- Construction 10.7%
- AC/Ventilation 5.7%

Central Park Noise Analysis (%)
- Loud Music/Party 18.8%
- Construction 12.7%
- Loud Talking 11.3%
- Vehicle 8.4%
- AC/Ventilation 7.2%

Private carting noise: 4.6%
Jack Hammering: 4.6%
Others: 4.3%