Evaluating Crime Prediction Maps
MONSURU ADEPEJU & ROSSER GABRIEL

Big Data and Intelligent Policing
CPC Project Closing Workshop
7th June 2016
Motivation

Given the increasing number of predictive hotspot methods, crime analysts often find it difficult to determine **which method is most appropriate** for different data scenarios.

For example,

- Which method is the best for predicting geographically constraint crime such as shoplifting?
- Which method is most robust to highly spatio-temporal sparse crime types?
Proposed solution:

A systematic evaluation protocol by which the performance of multiple predictive methods can be assessed and compared.
Evaluation Framework

We combined measures of four different aspects of hotspot performance namely;

- Accuracy,
- Compactness,
- Variability,
- Complementarity
Evaluation metrics

1. Predictive accuracy (hit rate)

measures the number of crimes captured within the hotspot area

$$\text{Pred. Acc.} = \frac{\text{no of crimes captured}}{\text{Total no of crime}}$$

A method producing high predictive accuracy means more crimes can be intersected before they actually occur.
Evaluation metrics

2. Compactness Index (CI)

measures the ease at which a defined hotspot can be patrolled

✓ From police standpoint, Map A is easier to patrol than Map B because of higher connectivity ratio of hotspot units.
3. **Dynamic Variability Index (DVI)**

measures the extent to which the predicted locations change between consecutive predictions.

- DVI helps to distinguish between different methods based on the type of crime risk they detect e.g. short-term risk, long-term risks.
Evaluation metrics

4. Complementarity

measures the extent to which different methods detect the same and/or different crimes

✓ helps to reveal how much improvement a method made relative to other methods.

Case Study - London Borough of Camden

**Aim:** To demonstrate the utility of the proposed evaluation framework

**Dataset:** 3 crime types of varied level of ST sparseness

**Predict-evaluate routine:**
- For each method, generate hotspots at day $t_n$,
- Evaluate the hotspot for one day ahead ($t_{n+1}$)
- Repeat for 100 consecutive days
Results

mean hit rate, mean CI and mean DVI

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Method</th>
<th>Accuracy</th>
<th>Hotspot compactness</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Hit rate</td>
<td>Mean CI</td>
<td>Mean DVI</td>
<td></td>
</tr>
<tr>
<td>Shoplift</td>
<td>PSTSS</td>
<td>81.3</td>
<td>0.42</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>PKDE</td>
<td>74.3</td>
<td>0.55</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>SEPP</td>
<td>91.5</td>
<td>0.31</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>PHotspot</td>
<td>85.1</td>
<td>0.37</td>
<td>19.2</td>
</tr>
<tr>
<td>Violence</td>
<td>PSTSS</td>
<td>46.5</td>
<td>0.46</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>PKDE</td>
<td>51.7</td>
<td>0.54</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>SEPP</td>
<td>59.7</td>
<td>0.12</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>PHotspot</td>
<td>52.2</td>
<td>0.32</td>
<td>21.1</td>
</tr>
<tr>
<td>Burglary</td>
<td>PSTSS</td>
<td>34.4</td>
<td>0.51</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>PKDE</td>
<td>38.8</td>
<td>0.50</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>SEPP</td>
<td>47.4</td>
<td>0.02</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>PHotspot</td>
<td>34.9</td>
<td>0.30</td>
<td>5.3</td>
</tr>
</tbody>
</table>

Evaluation metrics of Camden crime prediction at 20% coverage level.
Results

Complementarity

Venn diagram showing the total number of crimes identified by each method at a fixed coverage of 20% in Camden

- **Shoplifting**
  - Total = 223
  - SEPP captures 213

- **Violence**
  - Total = 526
  - PHotspot captures 220
  - SEPP captures 271

- **Burglary**
  - Total = 165
Discussions

- Trade-offs:
  - Predictive accuracy vs. Compactness (ease of patrol)

- DVI reveals that certain methods are best suited for specific type of risk patterns. For example, PSTSS capture emerging risks patterns while PKDE captures persistent risk patterns.

- Complementarity suggests that results of some methods can be combined for better performance - ensemble predictions.