Spatially Embedded Co-offence Prediction Using Supervised Learning

Mohammad Tayebi\textsuperscript{1}, Martin Ester\textsuperscript{1}, Uwe Glässer\textsuperscript{1}, Patricia Brantingham\textsuperscript{2}

\textsuperscript{1}Simon Fraser University, School of Computing Science
\textsuperscript{2}Simon Fraser University, School of Criminology
Crime

- Crime generates substantial cost.
- Crime reduction and prevention
- Predictive policing
- Data mining approaches for crime prevention
- Forward-thinking, proactive versus reactive
- Micro-level analysis in addition to the macro-level analysis
Predictive Policing

- Predicting offenders
- Predicting victims
- Predicting crime locations
- Predicting criminal collaborations
  - Co-offending network disruption
  - Organized crime detection
  - Co-offence prediction
Co-offending Networks

- Network of offenders who have committed crimes together.
  - Node: offender/ Edge: co-offence

  "Understanding co-offending is central to understanding the etiology of crime and the effects of intervention strategies". [Reiss, 1988]

- Most co-offending groups are unstable and the relationships are shortlived [Reiss, 1988].

- There is some stability in co-offending relationships over time [McGloin, 2008].
Co-offence Prediction Problem

Problem: To predict whether a potential co-offence in $G_t$ belongs to the positive class or the negative class.

Potential co-offence

$G_t(V_t, E_t)$

$G_{t+1}(V_{t+1}, E_{t+1})$

$(u, v) \notin E_t \land (u, v) \in E_{t+1}$

$(u, v) \notin E_t \land (u, v) \notin E_{t+1}$
Co-offence Prediction

- Link prediction problem for co-offending networks
- Modeled using a binary classification problem that adopts a set of **prediction features**
- **Main challenge** is heavily skewed distribution of negative and positive classes.
  - The prior probability of link formation is very small.
  - Negative class: 890M, Positive class: 11k
  - Classifier overfits to negative samples
  - For co-offence prediction the recall of positive samples is important.
Geographic and Network Proximity

- Crime is strongly linked to geographical characteristics
  - British Columbia crime dataset
    - 39% of the co-offenders live in less than 2 km apart.
    - 46% of the crimes happen in less than 2 km distance from the home location.
  - Co-offenders tend to be geographically confined.
Common Activity Space

Activity Space: $A_u^R = \{a_u^1, a_u^2, \ldots, a_u^k\}$

Common Activity Space: $A_{u,v}^R = \{a_u^{i,j}, a_v^{i,j} | a_u^i \cap a_v^j \neq \emptyset \land \delta_u^i \cap \delta_v^j \neq \emptyset\}$
Criminal Cooperation Opportunities

- Offenders do not select their collaborators accidentally.
  - **Socially**-related: distance in the co-offending network (smaller than \( N \))
  - **Geographically**-related: common activity space (with radius \( R \))
  - **Experience**-related: similar criminal experience (\( P \) similar crime types)

- Prediction spaces
  - **SR Space**: socially-related, and \( N = 2 \).
  - **GR Space**: geographically-related but not socially-related, and \( R = 2 \text{km} \).
  - **ER Space**: Experience-related but not socially-related, and \( P = 2 \).

- Effects of prediction space division
  - Clearer understanding of the effect of criminal cooperation opportunities
  - Reducing class imbalance ratio
Reducing Class Imbalance Ratio

- Reducing class imbalance ratio and keeping as many positive samples as possible.
- With \((N=2, R=2\text{km}, P=2)\) the class imbalance ratio decreases significantly for SR, GR and ER spaces.
- We keep about 30% of positive samples in each prediction space, and 50% in total.

![Bar chart showing prediction space size and class imbalance ratio](chart.png)
Prediction Features

Social features: Derived using only the topology of co-offending networks and the position of offenders in the network.

Similarity features: Derived from offenders’ demographic attributes.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferential</td>
<td>$</td>
</tr>
<tr>
<td>Common</td>
<td>$</td>
</tr>
<tr>
<td>Overlap</td>
<td>$\frac{</td>
</tr>
<tr>
<td>Adamic</td>
<td>$\sum_{z \in \Gamma^1_u \cap \Gamma^1_v} \log(\Gamma^1_z)$</td>
</tr>
</tbody>
</table>

|                          | $|Age(u) - Age(v)|$                           |
|--------------------------|-----------------------------------------------|
| Age                      |                                               |
| Gender                   | $\begin{cases} 1, & \text{if Gender}(u) = \text{Gender}(v) \\ 0, & \text{if Gender}(u) \neq \text{Gender}(v) \end{cases}$ |
| Ethnic                   | $\begin{cases} 1, & \text{if Ethnic}(u) = \text{Ethnic}(v) \\ 0, & \text{if Ethnic}(u) \neq \text{Ethnic}(v) \end{cases}$ |
| CrimSim                  | $\frac{\sum_{i=1}^{K} P^i_u P^i_v}{\sqrt{\sum_{i=1}^{K} (P^i_u)^2} \times \sqrt{\sum_{i=1}^{K} (P^i_v)^2}}$ |
**Prediction Features**

**Geographic features:** With increasing the overlap of the activity space of offenders the chance of forming new criminal collaboration increases.

**Geo-Social features:** Combines the social and geographic characteristics of offenders.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDN</td>
<td>[ \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{e^{-D(h^i_k,h^j_k)}}{</td>
</tr>
<tr>
<td>HDT</td>
<td>[ \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{e^{-D(h^i_k,h^j_k)}}{</td>
</tr>
<tr>
<td>CDN</td>
<td>[ \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{e^{-D(c^i_u,c^j_v)}}{</td>
</tr>
</tbody>
</table>

**Formulas:**

- **OCT**
  \[ \sum_{i=1}^{p} \sum_{j=1}^{k} |\chi^i_{u,v}| \]

- **OCTT**
  \[ \sum_{i=1}^{p} \sum_{j=1}^{k} |\chi^i_{u,v}| \times |\delta^i_u \cap \delta^j_v| \]

- **OCN**
  \[ \sum_{i=1}^{p} \sum_{j=1}^{k} |\phi^i_{u,v}| : [t_0, t] \]

- **CCT**
  \[ \sum_{i=1}^{p} \sum_{j=1}^{k} |\phi^i_{u,v}| \times |\delta^i_u \cap \delta^j_v| \]

- **CCN**
  \[ \sum_{i=1}^{p} \sum_{j=1}^{k} |\phi^i_{u,v}| : [t_0, t] \]
British Columbia Crime Data

- Police arrest data
  - Covers all of British Columbia policed under contract with the RCMP
  - Between mid-2001 and mid-2006
  - 4.4 million events, 1000 crime types

- Extracted co-offending network
  - 150,000 nodes
  - Average degree of four
  - 50% of the nodes have degree one.
  - The largest connected component links about 18% of the nodes.
Experimental Design

- Train (first 50 months) and test set (last 10 months)
  - 1.8 M and 800K records
  - 67K and 17K offences with more than one involved offender
- Classification methods: Naïve Bayes, C4.5, random forests, and bagging
- Running 10-fold cross validation over 10 different randomly sampled training sets
- Evaluation measures
  - ROC and AUC
Single Feature Significance

- In SR, Preferential attachment is the best among social feature.
- In GR, performance of geographic and geo-social features are weaker.
- Similarity features works better in SR compared to GR and ER.
- Crime locations distance works better than home location distance.
Prediction Evaluation

- All classifiers for all spaces outperform single features.
- Generally, prediction works best in the ER space.
- Two ensemble methods, bagging and random forest classifiers, work better than the other classifiers.
- Naïve Bayes is the weakest one in all spaces.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Space</th>
<th>Precision</th>
<th>Recall</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>J48</td>
<td>SR</td>
<td>0.888</td>
<td>0.807</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>0.869</td>
<td>0.834</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>ER</td>
<td>0.935</td>
<td>0.81</td>
<td>0.898</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>SR</td>
<td>0.836</td>
<td>0.514</td>
<td>0.825</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>0.825</td>
<td>0.441</td>
<td>0.817</td>
</tr>
<tr>
<td></td>
<td>ER</td>
<td>0.945</td>
<td>0.706</td>
<td>0.895</td>
</tr>
<tr>
<td>Random Forest</td>
<td>SR</td>
<td>0.898</td>
<td>0.843</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>0.864</td>
<td>0.883</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>ER</td>
<td>0.941</td>
<td>0.944</td>
<td>0.982</td>
</tr>
<tr>
<td>Bagging</td>
<td>SR</td>
<td>0.908</td>
<td>0.84</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>0.863</td>
<td>0.854</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>ER</td>
<td>0.946</td>
<td>0.942</td>
<td>0.986</td>
</tr>
</tbody>
</table>
Prediction Strength

- Geo-social feature set outperforms the other three sets.
- Geographic feature set has the worst performance.
- Prediction performance is the best when we integrate all three feature sets.

<table>
<thead>
<tr>
<th>Features Set</th>
<th>Precision</th>
<th>Recall</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td>0.903</td>
<td>0.792</td>
<td>0.919</td>
</tr>
<tr>
<td>Geographic</td>
<td>0.721</td>
<td>0.786</td>
<td>0.811</td>
</tr>
<tr>
<td>Geo-social</td>
<td>0.863</td>
<td>0.853</td>
<td>0.942</td>
</tr>
<tr>
<td>Similarity</td>
<td>0.849</td>
<td>0.851</td>
<td>0.928</td>
</tr>
<tr>
<td>All Features</td>
<td>0.908</td>
<td>0.84</td>
<td>0.951</td>
</tr>
</tbody>
</table>
Conclusion

- Our defined prediction spaces reduces the class imbalance ratio significantly.
- Our novel geo-social feature set outperforms the other feature sets.
- The proposed framework can correctly predict roughly 90% of the co-offences in the prediction spaces.
- Data mining provides valuable insights and novel methods for short-term and long-term crime reduction and prevention strategies.
In theory, there is no difference between theory and practice. But, in practice, there is.

Jan L. A. van de Snepscheut
Thanks!