Detecting Patterns of Crime with *Series Finder*

Tong Wang and **Cynthia Rudin**
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Daniel Wagner and **Rich Sevieri**
Cambridge Police Department
Cambridge, MA 02139, USA

Abstract

Many crimes can happen every day in a major city, and figuring out which ones are committed by the same individual or group is an important and difficult data mining challenge. To do this, we propose a pattern detection algorithm called *Series Finder*, that grows a pattern of discovered crimes from within a database, starting from a "seed" of a few crimes. Series Finder incorporates both the common characteristics of all patterns and the unique aspects of each specific pattern. We compared Series Finder with classic clustering and classification models applied to crime analysis. It has promising results on a decade's worth of crime pattern data from the Cambridge Police Department.

Introduction

The job of a crime analyst is to find patterns of crime. If crime analysts locate an ongoing pattern of crime committed by the same offender (a "series"), preemptive measures may be applied to prevent the next crime or to apprehend a suspect. Using a database of past crimes, Series Finder processes information similarly to how crime analysts process information instinctively: it searches through the database looking for similarities between crimes in a growing pattern and in the rest of the database, and tries to identify the modus operandi (M.O.) of the particular offender or group committing these crimes. As more crimes are added to the set, the M.O. becomes more well-defined. Our approach to pattern discovery captures several important aspects of patterns:

- *Each M.O. is different.* Different criminals can have very different M.O.'s. Some offenders operate during weekdays and target apartment buildings, others may operate mainly on the weekends, targeting single family houses. Different combinations of crime features can be more important than others for characterizing different M.O's.
- *General commonalities in M.O. do exist.* Each pattern is different but, for instance, similarities in time and space are important and should generally be weighted highly.
- *Patterns can be dynamic.* Sometimes the M.O. shifts during a pattern. For instance, early on, the criminal uses bodily force to open the doors as a means of entry, whereas later in the pattern, he uses tools to pry the door open. Methods that consider an M.O. as stationary (e.g., clustering) would not naturally capture these dynamics.

Series Finder for Pattern Detection

We use $\mathcal{P} = \{C_1, C_2, \ldots, C_{|\mathcal{P}|}\}$ to denote a true pattern of crime, where each of the C_i's represents a crime. Only a seed of a few crimes from $\mathcal{P}$ are known. Series Finder uses the seed to grow a set of discovered crimes $\hat{\mathcal{P}}$, in hopes that $\hat{\mathcal{P}}$ will eventually be similar to the underlying (and unknown) set $\mathcal{P}$. Specifically, it selects crimes from a candidate crimes set $C_{\hat{\mathcal{P}}}$ to add to $\hat{\mathcal{P}}$ sequentially. In practice, $C_{\hat{\mathcal{P}}}$ is usually a set of crimes occurring in the same year as $\mathcal{P}$. We need several definitions, stated below.

Crime-crime similarity The pairwise similarity γ measures how similar crimes C_i and C_k are in a pattern set $\hat{\mathcal{P}}$. We model it in the following form: $\gamma_{\hat{\mathcal{P}}}(C_i, C_k) = \sum_{j=1}^{J} \lambda_j \eta_{\hat{\mathcal{P}},j} s_j(C_i, C_k)$, where $\{\lambda_j\}_j$ are "pattern-general" weights, and $\{\eta_{\hat{\mathcal{P}},j}\}_j$ are "pattern-specific" weights. We form J similarity measures between crimes, and s_j is the similarity measure for the j^{th} attribute. Two crimes have a high γ if they are similar along attributes that are important specifically to that crime pattern, and generally to all patterns.

The *pattern-specific weights* $\eta_{\hat{\mathcal{P}},j}$ capture characteristics common to most or all crimes within a specific pattern. These weights are defined as:

$$\eta_{\hat{\mathcal{P}},j} := \frac{1}{\Gamma_{\hat{\mathcal{P}}}} \frac{1}{|\hat{\mathcal{P}}|(|\hat{\mathcal{P}}|-1)/2} \sum_{i=1}^{|\hat{\mathcal{P}}|} \sum_{k=1}^{|\hat{\mathcal{P}}|} s_j(C_i, C_k),$$

where $\Gamma_{\hat{\mathcal{P}}}$ is the normalizing factor $\Gamma_{\hat{\mathcal{P}}} = \sum_{j=1}^{J} \eta_{\hat{\mathcal{P}},j}$.

The *pattern-general weights* λ_j are learned from all patterns, using a coordinate-based optimization algorithm (not described here) that optimizes a balance of precision and recall. They consider the general importance of each attribute.

Pattern-crime similarity Pattern-crime similarity S measures whether crime i is similar enough to set $\hat{\mathcal{P}}$ that it should be potentially included in $\hat{\mathcal{P}}$. The pattern-crime similarity incorporates the dynamics in M.O. discussed in the introduction. The dynamic element is controlled by a parameter

d, called the *degree of dynamics*. The pattern-crime similarity is defined as follows for pattern $\hat{\mathcal{P}}$ and crime $C \in C_{\hat{\mathcal{P}}}$: $S(\hat{\mathcal{P}}, C) := \left(\frac{1}{|\hat{\mathcal{P}}|} \sum_{n=1}^{|\hat{\mathcal{P}}|} \gamma_{\hat{\mathcal{P}}}(C, C_n)^d \right)^{(1/d)}$ where $d \geqslant 1$. This is a soft-max, that is, an ℓ_d norm over $C_n \in \hat{\mathcal{P}}$. Use of the soft-max allows the pattern $\hat{\mathcal{P}}$ to evolve: crime i needs only be very similar to a few crimes in $\hat{\mathcal{P}}$ to be considered for inclusion in $\hat{\mathcal{P}}$ when the degree of dynamics d is large. On the contrary, if d is small, this forces patterns to be very stable and stationary, as new crimes would need to be similar to most or all of the crimes already in $\hat{\mathcal{P}}$ to be included. For our purpose, d is chosen appropriately to balance between including the dynamics (d large), and stability and compactness of the pattern (d small).

Series Finder algorithm Starting with the seed, crimes are added iteratively from $C_{\hat{\mathcal{P}}}$ to $\hat{\mathcal{P}}$. At each iteration, the candidate crime with the highest pattern-crime similarity to $\hat{\mathcal{P}}$ is tentatively added to $\hat{\mathcal{P}}$. Then $\hat{\mathcal{P}}$'s cohesion is evaluated, which measures the cohesiveness of $\hat{\mathcal{P}}$ as a pattern of crime: Cohesion$(\hat{\mathcal{P}}) = \frac{1}{|\hat{\mathcal{P}}|} \sum_{C_n \in \hat{\mathcal{P}}} S(\hat{\mathcal{P}} \backslash \{C_n\}, C_n)$. While the cohesion is above a threshold, we will continue to grow $\hat{\mathcal{P}}$. Here is the formal algorithm:

Initialization: $\hat{\mathcal{P}} \leftarrow \{\text{Seed crimes}\}$
repeat
    $C_{\text{tentative}} \in \arg\max_{C \in (C_{\hat{\mathcal{P}}} \backslash \hat{\mathcal{P}})} S(\hat{\mathcal{P}}, C)$
    $\hat{\mathcal{P}} \leftarrow \hat{\mathcal{P}} \cup \{C_{\text{tentative}}\}$
    Update: $\eta_{\hat{\mathcal{P}},j}$ for $j \in \{1, 2, \dots J\}$, and Cohesion$(\hat{\mathcal{P}})$
until Cohesion$(\hat{\mathcal{P}}) <$ threshold
$\hat{\mathcal{P}}^{\text{final}} := \hat{\mathcal{P}} \backslash C_{\text{tentative}}$
return $\hat{\mathcal{P}}^{\text{final}}$

Experiments

We use a dataset of 4855 housebreaks in Cambridge MA, USA, between 1997 to 2006. Crime attributes include geographic location, date, day of week, time frame, location of entry, means of entry, an indicator for "ransacked," type of premise, an indicator for whether residents were present, and suspect and victim information. We also have 51 patterns that were hand-labeled by crime analysts. We developed 13 pairwise crime metrics s_j, not discussed here.

The evaluation metrics we use are standard quality measures in information retrieval, namely the *average precision* and *reciprocal rank*. These evaluation measures consider the rank order in which the returned crimes are found.

We compare with hierarchical agglomerative clustering and an incremental nearest neighbor approach as competing baseline methods. For all methods, the seed was chosen to be the first two crimes, chronologically, recorded by the Cambridge Police Department for each pattern.

Hierarchical agglomerative clustering begins with every crime as a singleton cluster. At each step, the most similar (according to the similarity criterion) two clusters are merged into a single cluster, producing one less cluster at the next level. The *incremental nearest neighbor* approach begins with the seed set. At each step, the nearest neighbor (according to the similarity criterion) of the set is added to the pattern set, until the nearest neighbor is no longer sufficiently similar. Each model is used with three different criteria for cluster-cluster or cluster-crime similarity: *Single Linkage* (SL), which considers the most similar pair of crimes; *Complete Linkage* (CL), which considers the most dissimilar pair of crimes, and *Group Average* (GA), which uses the averaged pairwise similarity (Hastie et al. 2005).

Each model was trained on 35 patterns and tested on 16 patterns. The average precisions and reciprocal ranks are plotted in Figure 1. Each boxplot contains 16 average precisions/reciprocal ranks computed for the 16 patterns. This figure shows a substantial advantage of Series Finder in terms of both metrics. Note that the weights used by the competing models are provided by detectives based on their experience, while the weights of Series Finder are learned from data.

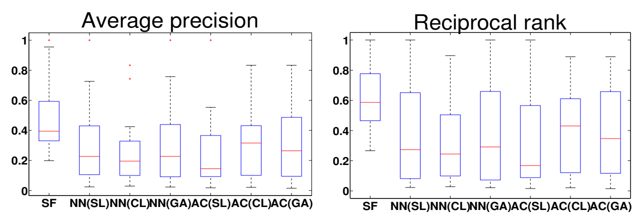

Figure 1: Average precisions and reciprocal ranks for all models. SF represents "Series Finder", NN represents "Incremental nearest neighbor classification" and AC represents "Agglomerative clustering".

Related Work

Many classic data mining techniques have been successful for crime analysis, such as association rule mining (Ng et al. 2007; Buczak and Gifford 2010), classification (Wang, Chen, and Atabakhsh 2004), clustering, and pattern detection (Nath 2006). For a general overview, readers can refer to (Chen et al. 2004). Domains related to our work include finding hot-spots, which are localized high-crime-density areas, e.g. see (Eck et al. 2005), and near repeats, which are localized in time and space (Ratcliffe and Rengert 2008). Much of the work from the UCLA group behind the PREDPOL software package has focused on predicting hotspots (Short et al. 2008; Cantrell, Cosner, and Manásevich 2012; Mohler et al. 2011) and near repeats (Short et al. 2009). A previous work on serial pattern detection is that of (Dahbur and Muscarello 2003), which uses a cascaded network of Kohonen neural networks followed by heuristic processing of the network outputs. However, feature grouping in the first step makes an implicit assumption that features manually selected to group together have the same importance, which our work shows is not necessarily the case. The work of (Nath 2006) uses a weighting of attributes provided by detectives, similar to our baseline comparison methods.

References

Buczak, A. L., and Gifford, C. M. 2010. Fuzzy association rule mining for community crime pattern discovery. In *ACM SIGKDD Workshop on Intelligence and Security Informatics*.

Cantrell, R. S.; Cosner, C.; and Manásevich, R. 2012. Global bifurcation of solutions for crime modeling equations. *SIAM Journal on Mathematical Analysis* 44(3):1340–1358.

Chen, H.; Chung, W.; Xu, J.; Wang, G.; Qin, Y.; and Chau, M. 2004. Crime data mining: a general framework and some examples. *Computer* 37(4):50–56.

Dahbur, K., and Muscarello, T. 2003. Classification system for serial criminal patterns. *Artificial Intelligence and Law* 11(4):251–269.

Eck, J.; Chainey, S.; Cameron, J.; and Wilson, R. 2005. Mapping crime: Understanding hotspots. Technical report, National Institute of Justice, NIJ Special Report.

Hastie, T.; Tibshirani, R.; Friedman, J.; and Franklin, J. 2005. *The elements of statistical learning: data mining, inference and prediction*. Springer.

Mohler, G. O.; Short, M. B.; Brantingham, P. J.; Schoenberg, F. P.; and Tita, G. E. 2011. Self-exciting point process modeling of crime. *Journal of the American Statistical Association* 106(493).

Nath, S. V. 2006. Crime pattern detection using data mining. In *Web Intelligence and Intelligent Agent Technology Workshops*, 41–44.

Ng, V.; Chan, S.; Lau, D.; and Ying, C. M. 2007. Incremental mining for temporal association rules for crime pattern discoveries. In *Proc. of the 18th Australasian Database Conference*, volume 63, 123–132.

Ratcliffe, J. H., and Rengert, G. F. 2008. Near-repeat patterns in Philadelphia shootings. *Security Journal* 21(1):58–76.

Short, M. B.; D'Orsogna, M. R.; Pasour, V. B.; Tita, G. E.; Brantingham, P. J.; Bertozzi, A. L.; and Chayes, L. B. 2008. A statistical model of criminal behavior. *Mathematical Models and Methods in Applied Sciences* 18:1249–1267.

Short, M. B.; D'Orsogna, M.; Brantingham, P.; and Tita, G. 2009. Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology* 25(3):325–339.

Wang, G.; Chen, H.; and Atabakhsh, H. 2004. Automatically detecting deceptive criminal identities. *Communications of the ACM* 47(3):70–76.