Finding Patterns with a Rotten Core: Data Mining for Crime Series with Core Sets (KDD Workshop Version)

Cynthia Rudin and Tong Wang
Massachusetts Institute of Technology
Cambridge, MA, US
{tongwang, rudin}@mit.edu

Daniel Wagner and Rich Sevieri
Cambridge Police Department
Cambridge, MA, US
{dwagner, rsevieri} @cambridgepolice.org

ABSTRACT

One of the most challenging problems facing crime analysts is that of identifying "crime series" which are sets of crimes committed by the same individual or group. Detecting series' of crime can be an important step in predictive policing, as knowledge of an ongoing pattern can be of paramount importance towards stopping it. Currently, crime analysts detect patterns manually; our goal is to assist them by providing automated tools for discovering crime series from within a database of crimes. Our approach relies on a key hypothesis that each crime series possesses a *core* of crimes that are similar to each other, which can be used to characterize the modus operandi (M.O.) of the criminal. We find core sets of crime using an integer linear programming approach, and then construct the rest of the crime series by merging core sets to form the full crime series. To judge whether a crime series is indeed a core, we consider both *pattern-general similarity*, which can be learned from past crime series, and *pattern-specific similarity*, which is specific to the M.O. of the series and cannot be learned. We learn a similarity graph on the set of crimes to form the pattern-general similarity.

1. INTRODUCTION

The foundation of *predictive policing* is that if we are able to predict crime, we can takes steps to prevent it. Empirical approaches to predictive policing have recently been adopted by many law enforcement agencies, and the National Institute of Justice has recently launched an initiative in support of predictive policing [11]. One of the most important problems in predictive policing is that of "crime series detection," or the detection of a set of crimes committed by the same individual or group. If a crime analyst can identify an ongoing crime series, it is possible to apprehend the suspect(s), and possibly, actions can be taken to stop this pattern. Criminals follow certain modus operandi (M.O.) that characterizes their crime series; for instance, some criminals operate exclusively during the day, others work at night, some criminals target apartments for housebreaks, while others target single family houses. It is a crime analysts' job to find the characteristics of M.O. in order to identify crime series, and they currently do this manually [5]. If we can use automated tools to detect crime series and identify the M.O., it could potentially assist crime analysts to locate ongoing series, to assign correct attribution for past crimes, and to prevent further crime. There is evidence that place-based approaches, where police target specific locations and times, result in crime reduction and safer neighborhoods [14, 19, 20]. Research on crime pattern detection can lead directly to place-based approaches, and any advance in this area could lead directly to reductions in crime. (In contrast, it is well known that random preventive patrol does not deter crime [6, 13].) Despite its critical importance for public safety, there are few academic works on this problem (e.g., [1, 2, 8, 10]).

Without the capability of automatically detecting specific series of crime, it is possible that crime series may take much longer to identify, or may never be identified. This is especially problematic for certain types of crime - for instance for housebreaks (burglaries) there is often no suspect information. Nationwide only $\approx 14\%$ of housebreaks are solved [21]. In our work, we aim directly at identifying series' in an automated way.

2. BRIEF OVERVIEW OF METHOD

Crime series detection can be viewed as a type of subspace clustering problem where the M.O. defines the set of relevant features for each cluster. We cannot determine in advance what the exact M.O. of an undetected crime series will be. Generally speaking, we need to reveal groups of objects that are similar on an unknown subset of their features. We say that such sets exhibit a *pattern-specific* similarity. The pattern-specific similarity cannot be learned since it may be true for only a small number of crimes.

Though we cannot characterize exact M.O.'s before we see them, we can characterize the type of M.O.'s we expect generally - for instance, crimes in a series are often close in time and space so these two features should have higher weights. We define a *pattern-general* similarity that encodes which factors are generally common to most crime series. It is learned from past crime series; for instance, time and space have high pattern-general weights. The pattern-general similarity induces a *similarity graph* where pairs crimes are connected if they are similar in the pattern-general sense.

The main hypothesis in this work is that most crime patterns have at least one *core* of crimes that exemplify the M.O. of the series. If we can locate all small cores of crime, we should thus have located parts of most crime series'. This hypothesis is based on the intuition of analysts, and has the dual purpose of assisting with computation - we can indeed consider all possible small subsets to calculate whether they are plausible core sets. The core sets are found using an integer linear program, which specifies that core sets must have pattern-specific and pattern-general similarity.

Once the core sets are found, we construct the full crime series by merging overlapping core sets. The merging algorithm is a breadth-first search algorithm. We proved in the longer version of this work [18] that merging core sets preserves desired properties.

This method was tested on the full housebreak database from the Cambridge Police Department containing information from thousands of crimes from over a decade. It was able to provide new insights into true patterns of crime committed in Cambridge.

Our method thus consists of three steps:

- Learn the similarity graph from past crime patterns. This involves solving a mixed-integer linear program. Once this is created, it encodes the pattern-general information. The result is a graph where each node is a crime. An edge exists between crimes if they are similar in a pattern-general sense.
- Find core sets. This involves locating small subsets of crimes that are well-connected in the similarity graph, and which are similar in at least d different ways, where the d factors characterize the M.O. of the core set. In practice d ranges from 5 to 7 factors, so that crimes in a core set are similar in 5 to 7 different ways. This involves solving an integer linear program.
- Merge core sets. We merge core sets that overlap and share several defining features. This involves a breadth-first search technique. These steps are discussed in depth in a longer version [18].

Related Work

The problem we consider is a clustering problem, but where the similarity measure between objects is supervised. Our work relates to various subfields of clustering (see for instance [7]), including pattern-based clustering [12,16] which is a semi-supervised approach (unlike ours - we do not use test data at training time), subspace clustering (e.g., [3,15]) which detects all clusters in all subspaces, and work at the intersection of dense subgraph mining and pattern mining in feature graphs that is somewhat similar to ours [4,9]. Most previous work on crime series detection [8,10] have pattern-general weights that come from experts (like our baselines) and are not learned from data (with [1] as an exception), and do not consider pattern-specific aspects, and thus cannot capture specific M.O.'s of crimes. Some past approaches have flaws that require heuristic post-processing to handle [2]. Our previous attempt at solving this was an iterative approach that did not consider core sets [17].

3. DATA

Our data set was provided by the Crime Analysis Unit of the Cambridge Police Department in MA, USA. It has 7,067 housebreaks that happened in Cambridge between 1997 and 2011, containing 51 hand-curated patterns contained within the 4,864 crimes between 1997 and 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 took the 51 hand-curated patterns, and divided them randomly into four subsets (folds) with sizes 12 or 13 patterns each. We used 3 of the 4 folds to learn the pattern general weights and tested on the remaining fold for the experimental evaluation discussed below.

4. PERFORMANCE EVALUATION

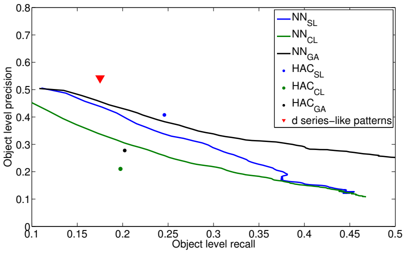

Figure 1: Average object-level precision and recall

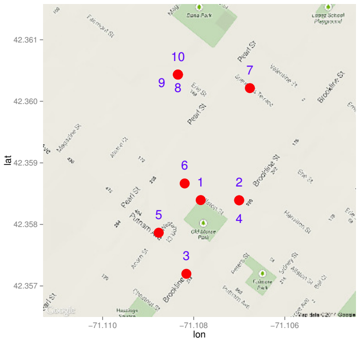

Figure 2: The locations of crimes in the series.

For each pattern we discovered, we determined how close it is to one of the real patterns. For each discovered pattern, we assigned a *dominating pattern* to be the real pattern possessing the most crimes that overlap with our discovered pattern. The definitions of object-level precision and recall for a d-series-like pattern $G = (V, E)$ are as follows:

$$\text{O-Precision(G)} = \frac{\sum_{\ell=1}^{|V|} \mathbf{1}(\ell \in V_{\text{dominating pattern}})}{|V|} \quad (1)$$

$$\text{O-Recall(G)} = \frac{\sum_{\ell=1}^{|V|} \mathbf{1}(\ell \in V_{\text{dominating pattern}})}{|V_{\text{dominating pattern}}|} \quad (2)$$

where $|V_{\text{dominating pattern}}|$ is the number of crimes in the dominating real pattern and the sum is taken over crimes in the pattern we found. We evaluated the average object-level precision and recall for all the patterns and over all the test folds, plotted as a point on Figure 1. On this figure, we also evaluated the quality of several baselines including Hierarchical Agglomerative Clustering and an Iterative Nearest Neighbor approach. For the same level of recall, the precision attained by our method was quite a bit higher than baselines. The extended version of this work includes several other quantitative evaluations [18].

5. CASE STUDY

We performed a blind test, where we aimed to detect crime patterns between 2007 to 2012 for which we do not have pattern data. One particularly interesting crime series includes 10 crimes from November 2006 to March 2007. Figure 2 shows geographically where these crimes were located. Table 1 provides details about crimes within the series.

First consider the similarity graph. The subgraph containing the 10 crimes is diagrammed in Figure 3. Crimes 1 to 5

Table 1: Details of crimes in the case study.

NO	Date	Loc of entry	Mns of entry	Premises	Rans	Resid	Time of day	Day	Suspect	Victim
1	11/8/06	Basement Door	Unknown	Unknown	No	Not in	10:45-15:00	Wed	null	1 F
2	11/8/06	Front door	Pried	Unknown	No	Not in	8:00-18:30	Wed	null	1 M
3	11/16/06	Front door	Shoved/Forced	Unknown	No	Not in	9:00-17:00	Thur	null	1 M
4	12/7/06	Front door	Pried	Unknown	No	Not in	9:00-17:00	Thur	null	1 F
5	12/22/06	Front door	Pried	Unknown	No	In	11:48	Fri	null	1 M
6	2/1/07	Front door	Shoved/Forced	Unknown	No	In	14:45	Thur	3 Males	1 F
7	2/15/07	Front door	Unknown	Aptment	No	In	12:00-13:30	Thur	null	2 F
8	3/5/07	Front door	Shoved/Forced	Aptment	No	Not in	12:22-14:56	Mon	null	1 M
9	3/5/07	Front door	Broke	Aptment	No	Not in	12:22-14:56	Mon	null	1 F & 1 M
10	3/8/07	Front door	Pried	Aptment	No	Not in	12:50-13:30	Thur	null	1 M

Table 2: Core sets and their defining features for the case study

Core Sets	Crimes			Geo Loc	Days apart	Loc of entry	Mns of entry	Premises	Rans	Resid	Time of day	Day	Suspect	Victim
1	1	2	3	✓	✓	✓	O	O	✓	O	✓	✓	O	O
2	1	2	4	✓	✓	✓	O	O	✓	O	✓	✓	O	O
3	1	2	5	✓	✓	✓	O	O	✓	O	✓	✓	O	O
4	1	3	4	✓	✓	✓	O	O	✓	O	✓	✓	O	O
5	1	3	5	✓	✓	✓	O	O	✓	O	✓	✓	O	O
6	1	4	5	✓	✓	✓	O	O	✓	O	✓	✓	O	O
7	1	5	6	✓	✓	✓	O	O	✓	O	✓	✓	O	O
8	2	3	4	✓	✓	✓	O	O	✓	O	✓	✓	O	O
9	2	3	5	✓	✓	✓	O	O	✓	O	✓	✓	O	✓
10	2	4	5	✓	✓	✓	✓	O	✓	O	✓	✓	O	O
11	2	4	6	✓	✓	✓	O	O	✓	O	✓	✓	O	O
12	2	5	6	✓	✓	✓	O	O	✓	O	✓	✓	O	O
13	3	4	5	✓	✓	✓	O	O	✓	O	✓	✓	O	O
14	3	5	6	✓	✓	✓	O	O	✓	O	✓	✓	O	O
15	4	5	6	✓	✓	✓	O	O	✓	O	✓	✓	O	O
16	5	6	7	O	✓	✓	O	O	✓	✓	✓	✓	O	O
17	6	8	9	✓	✓	✓	O	O	✓	O	✓	✓	O	O
18	6	8	10	✓	✓	✓	O	O	✓	O	✓	✓	O	O
19	6	9	10	✓	✓	✓	O	O	✓	O	✓	✓	O	O
20	7	8	9	✓	✓	✓	O	✓	✓	O	✓	✓	O	O
21	7	8	10	✓	✓	✓	O	✓	✓	O	✓	✓	O	O
22	7	9	10	✓	✓	✓	O	✓	✓	O	✓	✓	O	O
23	8	9	10	✓	✓	✓	O	✓	✓	O	✓	✓	O	O

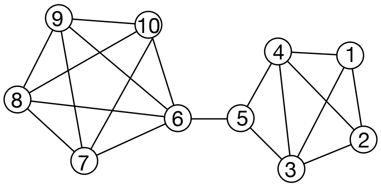

Figure 3: Similarity graph for crime series

are well connected, and crimes 6 to 10 are well connected. From only the similarity graph, the two subsets do not seem very related except for a single edge between crimes $\{5, 6\}$; however, this is only the pattern-general part of the story.

Our core set finder discovered core sets of size 3 with at least 6 defining features, shown in Table 2, where a check mark means the feature is a defining feature. These core sets show how the crimes are similar to each other in a pattern-specific way. All merged core sets included several features: geographic location, days apart, location of entry, the ransacked indicator, time of day, and day of the week.

As these data were reconsidered by crime analysts, we found out that when these crimes were analyzed back in 2006-2007, they were viewed as two unrelated patterns, one at the end of 2006, crimes 1 to 5, and one at the beginning of 2007, crimes 6 to 10. The connection between these two subsets of crime is very subtle and there is over a month gap between the two patterns, so it did not occur to the crime analysts to link them. The crime analysts' intuition agrees completely with the similarity graph, as the two subsets are weakly connected only by one edge; however, recall that this only describes the pattern-general part of the story. On examination of the core sets, not only are they are correlated in 6 features, but six of the core sets (core set indices 7, 11, 12, 14, 15, 16) contain crimes from both of the subsets, which is strong evidence that the two subsets should be merged. It is interesting that the core set consisting of crimes 5, 6, and 7 spanned the two subsets, where these crimes share the unusual feature that residents were present during the break-in. What may have happened is that the criminals left in December after crimes 1-5, and returned in February to continue housebreak number 6 in the same area where they were previously successful (in the lower part of the map); however they were witnessed committing crime 6, as indicated by the suspect information in crime 6, listed as "3 males"). They moved north just after that to commit crimes 7-10.

After revising this crime series, analysts now believe that these 10 crimes actually constitute a single series, and that the suspect information from crime 6 can be carried through to all the crimes in the discovered series.

This is a good example to show how crime patterns can be composed of core sets, and exhibit similarity both in a pattern-general way and pattern-specific way. It shows how we can use both aspects to mine patterns of crime. It is a subtle pattern spread over a period of time that an analyst could easily overlook.

Support Acknowledgement. MIT Lincoln Laboratory.

6. REFERENCES

[1] D. E. Brown and S. Hagen. Data association methods with applications to law enforcement. *Decision Support Systems*, 34(4):369–378, 2003.

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

[3] C. Domeniconi, D. Papadopoulos, D. Gunopulos, and S. Ma. Subspace clustering of high dimensional data. In *Proc. SIAM Conference on Data Mining*, 2004.

[4] S. Günnemann, B. Boden, and T. Seidl. DB-CSC: a density-based approach for subspace clustering in graphs with feature vectors. In *Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pages 565–580. Springer, 2011.

[5] S. L. Gwinn, C. Bruce, J. P. Cooper, and S. Hick. Exploring crime analysis. Readings on essential skills, Second edition. Published by BookSurge, LLC, 2008.

[6] G. L. Kelling, A. M. Pate, D. Dieckman, and C. Brown. The Kansas City preventive patrol experiment. *Technical report. Washington, DC: Police Foundation.*, 1974.

[7] H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data*, 3(1):1, 2009.

[8] S. Lin and D. E. Brown. An outlier-based data association method for linking criminal incidents. In *Proc. SIAM Conference on Data Mining*, 2003.

[9] F. Moser, R. Colak, A. Rafiey, and M. Ester. Mining cohesive patterns from graphs with feature vectors. In *Proc. SIAM Conference on Data Mining*, volume 9, pages 593–604, 2009.

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

[11] B. Pearsall. Predictive policing: The future of law enforcement? *National Institute of Justice Journal*, 266:16–19, 2010.

[12] J. Pei, X. Zhang, M. Cho, H. Wang, and P. S. Yu. Maple: A fast algorithm for maximal pattern-based clustering. In *Proc. International Conference on Data Mining*, pages 259–266. IEEE, 2003.

[13] L. Sherman and J. Eck. Policing for crime prevention. *Evidence-Based Crime Prevention*, 2002.

[14] L. W. Sherman, P. R. Garlin, and M. E. Buerger. Hot spots predatory crime: routine activities and the criminology of place. *Criminology*, 27(1):27–56, 1989.

[15] R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011.

[16] H. Wang, W. Wang, J. Yang, and P. S. Yu. Clustering by pattern similarity in large data sets. In *Proc.ACM SIGMOD*, pages 394–405, 2002.

[17] T. Wang, C. Rudin, D. Wagner, and R. Sevieri. Learning to detect patterns of crime. In *Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2013.

[18] T. Wang, C. Rudin, D. Wagner, and R. Sevieri. Finding patterns with a rotten core: Data mining for crime series with core sets, 2014. Submitted.

[19] D. Weisburd. Bringing social context back into the equation. *Criminology and Public Policy*, 11(2):317–326, 2012.

[20] D. Weisburd and L. G. Mazerolle. Crime and disorder in drug hot spots: Implications for theory and practice in policing. *Police Quarterly*, 3(3):331–349, 2000.

[21] D. L. Weisel. Burglary of single-family houses. Problem-Oriented Guides for Police Series, No.18, 2002.