Learning to Detect Patterns of Crime

Tong Wang[1], Cynthia Rudin[1], Daniel Wagner[2], and Rich Sevieri[2]

[1] Massachusetts Institute of Technology, Cambridge, MA 02139, USA
{tongwang,rudin}@mit.edu
[2] Cambridge Police Department, Cambridge, MA 02139, USA
{dwagner,rsevieri}@cambridgepolice.org

Abstract. Our goal is to automatically detect patterns of crime. Among a large set of crimes that happen every year in a major city, it is challenging, time-consuming, and labor-intensive for crime analysts to determine which ones may have been committed by the same individual(s). If automated, data-driven tools for crime pattern detection are made available to assist analysts, these tools could help police to better understand patterns of crime, leading to more precise attribution of past crimes, and the apprehension of suspects. 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, and has had promising results on a decade's worth of crime pattern data collected by the Crime Analysis Unit of the Cambridge Police Department.

Keywords: Pattern detection, crime data mining, predictive policing.

1 Introduction

The goal of crime data mining is to understand patterns in criminal behavior in order to predict crime, anticipate criminal activity and prevent it (e.g., see [1]). There is a recent movement in law enforcement towards more empirical, data driven approaches to *predictive policing*, and the National Institute of Justice has recently launched an initiative in support of predictive policing [2]. However, even with new data-driven approaches to crime prediction, the fundamental job of crime analysts still remains difficult and often manual; *specific* patterns of crime are not necessarily easy to find by way of automated tools, whereas larger-scale density-based trends comprised mainly of background crime levels are much easier for data-driven approaches and software to estimate. The most frequent (and most successful) method to identify specific crime patterns involves the review of crime reports each day and the comparison of those reports to past crimes [3], even though this process can be extraordinarily time-consuming. In making these comparisons, an analyst looks for enough commonalities between a past crime and a present crime to suggest a pattern. Even though automated detection of specific crime patterns can be a much more difficult problem than

H. Blockeel et al. (Eds.): ECML PKDD 2013, Part III, LNAI 8190, pp. 515–530, 2013.

estimating background crime levels, tools for solving this problem could be extremely valuable in assisting crime analysts, and could directly lead to actionable preventative measures. Locating these patterns automatically is a challenge that machine learning tools and data mining analysis may be able to handle in a way that directly complements the work of human crime analysts.

In this work, we take a machine learning approach to the problem of detecting specific patterns of crime that are committed by the same offender or group. Our learning algorithm processes information similarly to how crime analysts process information instinctively: the algorithm 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. The M.O. is the set of habits that the offender follows, and is a type of motif used to characterize the pattern. 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.* Criminals are somewhat self-consistent in the way they commit crimes. However, different criminals can have very different M.O.'s. Consider the problem of predicting housebreaks (break-ins): Some offenders operate during weekdays while the residents are at work; some operate stealthily at night, while the residents are sleeping. Some offenders favor large apartment buildings, where they can break into multiple units in one day; others favor single-family houses, where they might be able to steal more valuable items. Different combinations of crime attributes 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, similarity in time and space are often important to any pattern and should generally by weighted highly. Our method incorporates both general trends in M.O. and also pattern-specific trends.
- *Patterns can be dynamic.* Sometimes the M.O. shifts during a pattern. For instance, a novice burglar might initially use bodily force to open a door. As he gains experience, he might bring a tool with him to pry the door open. Occasionally, offenders switch entirely from one neighborhood to another. Methods that consider an M.O. as stationary would not naturally be able to capture these dynamics.

2 Background and Related Work

In this work, we define a "pattern" as a series of crimes committed by the same offender or group of offenders. This is different from a "hotspot" which is a spatially localized area where many crimes occur, whether or not they are committed by the same offender. It is also different than a "near-repeat" effect which is localized in time and space, and does not require the crimes to be committed by the same offender. To identify true patterns, one would need to consider information beyond simply time and space, but also other features of the crimes, such as the type of premise and means of entry.