Pattern Recognition in Large Police Crime Data Sets

Timothy O'Shea, PhD
University of South Alabama
Political Science and Criminal Justice Department
Mobile, Alabama

Thomas Muscarello, PhD
DePaul University
Computer Sciences Department
Chicago, Illinois

A key responsibility of the police function is to identify and address problems associated with crime and disorder. Since police departments began to keep records of criminal incidents, they have attempted to put this information to strategic and tactical use. Pattern discrimination within the crime data set has always been the heart of crime analysis. By pattern, we refer to its customary usage in police parlance. That is, a pattern refers to an individual or group of individuals who are characterized by the fact that they commit a series of criminal offenses over an extended period of time. Additionally, those individuals choose a particular crime category(e.g., rape, robbery, burglary, arson, etc.) and habitually execute the same method of operation through the series of separate incidents.

A fundamental feature of the community policing model, a reorganization effort strongly supported by academics and practitioners alike, is the accurate and comprehensive identification of problems. Crimes attributed to the career criminal stand out as a particularly relevant problem area since the empirical evidence suggests that a rather small subset of the universe of offenders is responsible for a rather large subset of the universe of criminal offenses. If this is true, then a high return on resource deployment would result from the identification and apprehension of career criminals.

There are no automated methods of discriminating patterned incidents in police data sets. This may not be a critical area of concern for smaller police agencies since patterned incidents tend to stand out; it is, however, somewhat more difficult in large urban departments to discover patterns because of the large data sets. For officers in these large departments to cull out patterns[1] they first must manually scan hard copies of case reports and memorize the values of

the reported characteristics of each incident(e.g., the physical characteristics of the offender, the location of the incident, the type of incident, weapon used, property taken, time of the incident, vehicle used, etc.). All of the characteristics of all the cases in the set of incidents must be stored in memory and then the characteristics of each case must be compared against all of the other cases. The number of robberies in a single detective administrative jurisdiction in Chicago for a several month span may number in the thousands. With each case report containing approximately 35 incident characteristics, the detective would have to hold 30-40 thousand values in memory and compare them all to discover the matched combinations of characteristics. Clearly, the information processing limitations of humans make proactive identification of patterns a virtual impossibility.

Focus Groups

Three focus groups were conducted to discover the pattern recognition methods of experienced Chicago Police Department robbery detectives. According to the participants, the pattern descriptions were ever-changing. That is, the features of a pattern were rarely ever the same, but changed in the combinations of characteristics and their values over time. These officers recognized that effective and comprehensive identification of patterns in the data under these conditions was unlikely and they responded by constructing a number of heuristics, or decision shortcuts, to overcome the deficiency. One method employed was a variation of fuzzy logic, in which characteristics of the incidents were redefined according to a more tactically useful standard. For example, when noting physical characteristics of offenders, a detective

would categorize the characteristics as very small, very large, or average in place of the reported values recorded on the case report. Since the normal distribution of offenders would likely consist of relatively few very large or very small offenders, this class of cases was examined first. The set of cases to be considered was thereby substantially reduced and made, for analytic purposes, more manageable.

In addition to the problem associated with the volume of cases to be analyzed, organizational structures were found to effect the pattern recognition capacity of the detectives. In 1980, the detective division was reorganized from a specialized(homicide/sex, robbery, burglary, and general assignment) to a generalized structure(violent and property crimes). Focus group participants maintained that when the detective division was specialized the corresponding institutional arrangements facilitated pattern recognition skills. For example, the robbery unit's social interaction patterns fostered ongoing discussions of active cases and thereby on occasion brought patterns to light. Also, when a single crime category, e.g., robbery, was the sole concern of the unit commander, they were able to focus their energy on creating and enforcing incentives to insure maximum productivity. For a robbery commander that meant identifying patterns, which translated into multiple case clearances, a measure of productivity for the commander.

The evidence drawn from the focus groups led to a compelling conclusion. Only the most obvious patterns are likely to be discovered by detectives, irregardless of the development of heuristics or the modification of organizational structures. Human information processing capacity is simply inadequate to analyze the quantity of data found in large urban crime data sets. Any meaningful improvement in the identification of career criminals will come only by way of statistically based pattern recognition tools. The objective of this research was to explore two

pattern recognition methods, artificial neural networks and cluster analysis, and ultimately develop a tool to proactively discriminate patterns.

Artificial Neural Networks

Neural networks are systems that seek via hardware or software to simulate the architecture and working of the human brain. Such systems are not procedural or rule based. They simulate a large number of neurons on various levels, all interconnected. Neural Networks are adept at finding patterns in historical data which can be used for predictive or forecasting purposes.

Neural nets must be trained to identify patterns before they can be used to recognize or categorize unknown patterns. A neural net is trained by presentation of a sufficient number of sample or training data patterns, such that the net converges on a weighted internal pattern which is activated for each exemplar group. A net begins in a quiescent state. As each pattern is presented it propagates through the net. Dependent on the learning algorithm used, the weights are adjusted as the net learns. Weights are strengthened when the net chooses a correct output and weakened when an incorrect output is chosen. When enough training patterns have been presented and verified, the net will converge to a weighted pattern.

For crime analytic purposes, one of the most important findings of this research concerns the two basic modes of learning associated with the development of neural networks: supervised and unsupervised. Supervised learning requires a "teacher" in the form of a training set of known patterns, or a human to grade performance. Using the supervised learning scheme

the actual output of a neural net is compared to a known desired output. On each iteration nodal connection weights are reset to minimize error. This is essentially what was described above. For example, say you want to predict which individuals are likely to default on a loan. You have a file cabinet full of data about individuals who have defaulted in the past. You use this data to teach the net what to look for. After the net has been trained, it can be asked to analyze a set of loan prospects to determine whether or not any fit the defaulter profile.

Unsupervised learning is sometimes called "self-supervised learning." Here networks use no external influences to set and adjust weights. Instead there is an internal performance monitor. Sometimes this is done by clusters of neurons working together in the network. Return to the loan defaulter problem. We assumed that the characteristics of loan defaulters were consistent from one defaulter to another. Furthermore, we assumed that there was a recurring configuration of characteristics over time that formed the profile of a loan defaulter. But what if over time, due to say, social or economic conditions, the configuration of characteristics of the loan defaulter changed. The unsupervised learning mode would detect this change in the configuration of characteristics and adjust the neural net accordingly. At the present state of the art, unsupervised learning is not well understood and is the subject of research.

Supervised learning procedures have achieved a reputation for producing good results in practical applications and are the most commonly used learning algorithm. Supervised learning was the learning mode used in this research. The obstacle found with the police data is that it is fluid and dynamic. The configuration of characteristics change over time. We may find today a pattern consisting of 20 incidents in which a male, white, 25 years old, 6 feet tall, red hair, and a tattoo of old ironsides on his right forearm is robbing old ladies downtown while armed with a

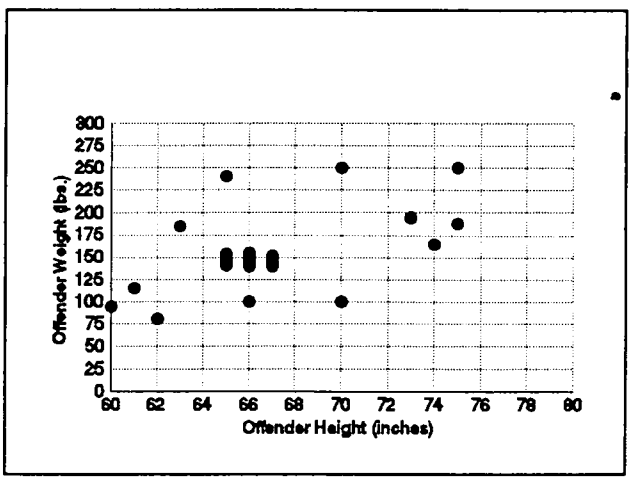
red, white, and blue battle axe. The offender undergoes a religious experience and chooses to take up missionary work in New Guinea. This configuration of characteristics ceases to appear in the data set, yet we can be sure that a different configuration will take its place. The configurations of characteristics will change over time in police data, patterns are not stable, and therefore virtually require an unsupervised methodology.

Cluster Analysis

The failure of neural nets to solve our pattern recognition problem led us to cluster analysis, a generic name for a range of formal, multivariate statistical procedures. The method seeks to group cases according to similarities of their defining characteristics. It has been used extensively in the fields of biology, anthropology, psychology, and political science.

Similarity has taken on varied meanings. Similarity, for our purposes, is associated with the concept of metrics. This describes similarity as the distance between objects in a Euclidean space. Objects are said to be similar or dissimilar by the distance which separates them in an n-dimensional space. The number of dimensions is determined by the number of variables used to describe the cases under consideration, which in the case of police data are the characteristics of the reported incident(e.g., the height,

Exhibit 1



weight, race, sex, age of the offender; the type of crime; the location of the crime; the weapon used; the proceeds taken; the vehicle used; the age, sex, race of the victim, etc.). Exhibit 1 illustrates a two-dimensional space, in this case the height and weight of 25 offenders. By plotting these characteristics of the individuals, we can easily see a cluster around the five foot six inch, 140 pound intersection of the X,Y axis. The data points that cluster around this point are said to be similar. Imagine adding a third, fourth, fifth, etc. dimension to create a multi-dimensional space, these dimensions representing the additional relevant characteristics of the offense under consideration.

Selection of the appropriate variables to describe the case is critical. A number of variables may characterize a case. The analyst's decision is to choose those dimensions that in combination will permit clusters that are similar according to a theory-based standard. Another variable selection decision concerns the weighting of variables. Which, if any, variables are of greater relevance in describing the case, and to what degree. In our case, a nineteen-dimension Euclidean space was constructed. Values were assigned to each variable and nearest neighbor clustering method was utilized to determine the relative spatial positioning of cases selected and subsequently group fixed number of similar cases.

Interface

There are a number of problems that act against an easy, efficient, quick implementation of such a system on a department wide basis. As the system has been developed it can be installed to one machine and updated with relative ease. As the number of machines increases

and their geographic area spreads the task becomes more complicated. This is especially true because of the relatively large size of the data file used(some in excess of 70 megabytes).

The analysis software requires large amounts of Random Access Memory to operate efficiently and quickly. There is a great deal of comparison and computation involved in clustering nearest neighbor. It is still relatively uncommon for IBM PC desktop computers to be configured with 64 or 128 Meg of RAM.

The operating system and basic architecture of personal computers is not well suited for running the type of program developed under this project. High speed architecture of the type available in mini computer/workstations is more desirable for massive number crunching. These workstations also usually run some version of Unix operating system which is better suited to programs of this type.

To make optimum use of the departmental data the systems should be networked to access data from the central data base. The personal computers used by case analysts should not be used as surrogate central data repositories. Most offices in the field do not have the network capability to make fast connection for downloading data. This can be a timely proposition using modems and phone lines.

Individual case analysts, particularly at remote sites, should have an easy to use program. They should not have to be responsible for maintaining central data stores, updating end-user programs, maintaining networks, or transmitting and maintaining large files. This task is better suited for information systems personnel, in this case headquarters crime analysis staff.

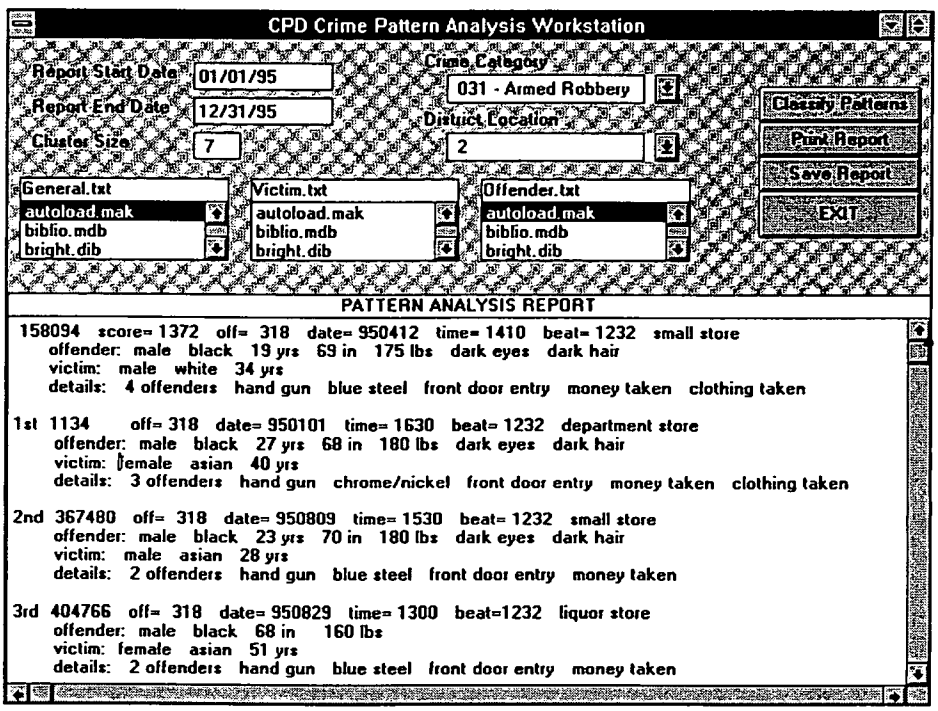
End users of the system should find it easy to use, easy to learn, and easy to remember. The system should also be transparent to end users. It should not matter to them where the data

resides and where the processing is done, so long as they get their results delivered to their display.

For these reasons we chose to move the delivery platform of the system to an environment which will support the needs enumerated above. This involves the use of the World Wide Web(WWW or Web) as the network of delivery channel for analytic reports that would be processed on a powerful high-speed Unix based server system. Users would access the system capabilities via a simple, cheap browser as user interface.

The user interface is graphic, and was built with tools that allow great flexibility in choice of hardware platforms. The most innovate aspect of the user interface is that it runs in a state-maintaining Web browser. In addition, the interface runs under windows as well as X-windows, and on any of the many machines that support Java or Tcl/Tk(two of the main languages in use in the development of applications embedded in Web pages).

Exhibit 2



Connectivity between the layers, and with the rest of the world, will be built on the Internet and WWW. Data bases,

applications, and decision support tools can be located on net nodes accessible via Internet connection. Exhibit 2 illustrates the end user GUI screen.

Conclusions

The findings of this research point to several conclusions. The first concerns the use of artificial intelligence in an effort to discriminate crime patterns. As we defined patterns in this research the use of artificial intelligence, i.e., neural networks, was found to be unsuitable. We caution, however, we are not suggesting that artificial intelligence has no place in the array of pattern recognition problem areas in law enforcement. We simply maintain that given the current state of the art, the neural network is not a solution to discriminating career criminal clusters in large police data sets. In those cases when the assumptions required for a supervised learning application are met, that is, when the configuration of characteristics that define the pattern are stable over time, neural networks are likely to be a useful tool. For example, the Chicago Police Department currently is exploring the use of neural networks as an early warning system to identify officers likely to engage in abuse of power behaviors. The profile that describes an officer likely to engage in this type of behavior would appear to conform to the stability assumption.

Second, this examination points to a variety of issues related to the nature of police data. Because much of the data collected by the police is missing, inaccurately described(by victims or witnesses), inaccurately entered into the data base system, or fuzzy(height of 66-76 inches) future systems development must address this problem. This will require the following:

*1) **Redesign of the case report form and data base**.* This will allow the capture and/or storage of data which will increase the usefulness of the categorization system. The narrative section of the report could possibly be coded to capture additional useful information. Additional data could be added which could help in the pattern classification process(e.g., geographic coordinates of the location of the occurrence could be used. At this time the coordinates of the beat in which the crime occurred are available).

2) ***Fuzziness in data must be addressed.*** In many cases the data values obtained from victims are described in fuzzy terms. As an example, offender age or height are frequently given as ranges. It is highly likely that these entries begin as simple qualitative verbal statements made by a victim, e.g., "he was tall, not very old, maybe 20 or 30." Even when a victim/witness provides a fixed value, e.g., "he was six feet tall," in reality this value falls within some range. Future research should examine available data in order to establish the range for any given value, quantitative or qualitative. A single "fuzzy" value could then be assigned to that characteristic. Although the analytic tool's capacity to arrive at precise patterns is reduced, in reality the value was not precise in the first place. Exploration of this question could lead to more useful victim/witness information for analytic purposes.

3) ***Preprocessing of data must be expanded.*** The data should be examined and manipulated to clean up errors and noise. Again utilizing fuzzy set theoretic logic, data elements should be looked at in conjunction. This process would address some data problem areas in ways that could clear up the uncertainties and fuzziness in victim and witness descriptions. For instance, the elements of offender hair color, skin color, eye color, race could all be used to come up with a composite "complexion/color" value. Likewise, offender height and offender weight taken in

conjunction could give a size measure. It may be more accurate to then map data values to fuzzy values such as dark or light, small or large.

Future work in this area must begin to examine the choice of data collected and its relation to the analytic function. There has been some discussion among crime analysts concerning the variety of data that should be collected in a case report. Some have suggested that in order to analyze and arrive at precise patterns that can be attributed to a single career criminal, we must collect more details about the incident. The question is how much is sufficient? In this project we decided that 19 characteristics best defined the incident, yet only a lengthy assessment of the pattern output from the analytic tool will tell us the degree of precision we have reached. It is highly likely, in our opinion, that the variables we have chosen do not represent the optimum description of the incident for analytic purposes. The answer to this empirical question will be a key factor in the success of precise pattern discrimination.

Finally, we suggest that police departments have for some time chosen to resist outside involvement in the analysis of police data. Police officers, somewhat familiar with statistical methods or computers have been targeted by their superiors to develop crime analysis routines. By and large, police officers and civilian crime analysts are ill-equipped to develop the sophisticated statistical tools necessary to analyze the data in ways likely to produce useful strategic and tactical outcomes. Police agencies must accept that technical assistance is needed; however, at the same time the police organization must realize that while the consultant may, and should, be knowledgeable about the methods to be applied to the problem, they are apt not to understand the strategic and tactical nature of the problem. The effort to produce an effective analytic tool is clearly a collaboration, in which the consultant provides the technical wherewithal

to address the problem and the police officer provides the experience to frame the problem.

Notes

1. I refer here to patterns in which the detective has no information about a set of characteristics that form the pattern. This is contrasted with characteristics that define the pattern, for example when an offender is arrested for robbery. We now know the characteristics and can query the database to pull out similar cases.