USING TECHNOLOGY TO ENHANCE POLICE PROBLEM SOLVING

Final Report submitted pursuant to NIJ grant #95-IJ-CX-0082.

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

Thomas Muscarello
DePaul University
Computer Sciences Department
Chicago, Illinois

TABLE OF CONTENTS

EXECUTIVE SUMMARY

- Automated methods of proactive pattern recognition are not currently available to police crime analysts. Reactive methods exist in the form of data base queries and geographical information systems, however these methods are inadequate for proactive pattern recognition.

- Detectives in large urban police departments are unlikely to proactively discriminate crime patterns due to human information processing constraints vis-a-vis their large crime data sets. The volume of cases reported and the number of characteristics of each incident captured make it virtually impossible for humans to match and compare the hundreds of thousands variable values over a relatively short time span for even a single crime category, e.g., robberies.

- In order to overcome their information processing deficiencies and thereby identify at least some portion of patterned incidents in the data, detectives construct a variety of heuristics, or decision shortcuts. For example, since vehicle descriptions are rarely reported in robbery cases, some detectives will look for cases with a vehicle described. The detective will then search only this variable in other cases. Given there will be very few cases with this variable described, the search for a patterned incident where the offender used a particular vehicle will be relatively manageable.

- Heuristic construction is related to a variety of organizational structures. The institutional arrangements of an organization will impact the incentive and disincentives that encourage or discourage detectives to construct heuristics. In particular, specialization of the detective function will likely serve to facilitate pattern recognition heuristics.

- This project attempted to create an automated pattern recognition tool to classify patterns of like criminal cases using two classification techniques. First a standard Neural Network approach was attempted. Employing supervised learning with undifferentiated crime patterns resulted in suboptimal performance of the neural net. There were no visible patterns to be used as training patterns. The Back Propagation Network architecture is insufficient to the task of differentiating patterns of activity given the data available. Unless the data base is changed to include a more comprehensive collection of variables, there is little chance that patterns can be abstracted for use in training a net.

- Each time that a new case pattern is entered ~~into~~ the net, it could fit numerous categories or patterns to a greater or lesser extent. This is a one to many mapping. This problem of conflict is not handled well by Back Propagation networks. One possible neural net solution ~~to~~ this problem may be the use of ~~the~~ Elman network which adds a number of "context nodes" to the input pattern layer. may allow differentiation of patterns occurring at different times. ~~In effect~~ Our system added context by preprocessing the input data using some simple heuristics and then feeding the data through a nearest neighbors pattern recognition component. The output from this component was far more successful at grouping potentially related patterns of criminal activity than was the neural net.

- The nearest neighbors technique was shown to exhibit great promise in clustering like cases. 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. Initial cluster validation indicates that the output is tactically useful to detectives in a first cut at the data, leading ultimately to pattern classification.

I. INTRODUCTION

STATEMENT OF THE PROBLEM

NIJ's 1994 National Assessment Program reported, "Overall, needs were greater in information systems then in any other area explored in the survey... In every application area but one (tracking the dates of hearings, noted by prosecutors), a majority of respondents said their systems needed improvement... Among police chiefs and sheriffs, for example, applications of greatest interest were systems to support problem solving..." (McEwen, 1995)

Anyone who has been remotely exposed to the criminal justice literature, for that matter the popular media, is aware of the remarkable diffusion of community policing in law enforcement. Although widely embraced by practitioners, there is considerable confusion over a consistent definition of community policing. The practice has taken on as many shapes and forms as there are agencies that have implemented it; nevertheless adoption of community policing has come to be synonymous with "progressive." The danger, as Goldstein (Rosenbaum, 1994) points out, is that policy makers embrace this politically appealing approach to policing without intellectually investing sufficiently in understanding the principles that drive it. This project seeks to focus on one such principle, one that we argue represents a foundation of the policy, and explore methods to structure innovative technology-based responses to facilitate its fulfillment.

Few would argue that the success of community policing rests on its ability to develop a partnership between the police agency and the community it serves. Goldstein (1990) offers a method to achieve this end and in so doing provides the core of the community policing philosophy. Goldstein suggests, quite reasonably, that police departments must identify, analyze the nature of, and specify alternative responses to problems. Police agencies must proactively rather than reactively address problems of crime and incivility. The effectiveness of this process rests on its ability to master each successive stage serially, i.e., each stage provides an underpinning for the succeeding ones. We propose, in this project, to focus our attention on what we contend is a gross deficiency in current police practices, i.e., problem identification, the initial stage of the Goldstein model.

The safety of a community rests largely on the ability of the police to identify and apprehend individuals who would put the public at risk. This has always been the case and we suspect this expectation will continue. Reformers have taken the position that this end is better served when police and the community form collaborative relationships. The success of this collaboration is grounded in the capacity of the police agency to accurately analyze data and transform it into useful tactical information.

We maintain that mid to large-sized police departments are profoundly impaired in their ability to analyze their large data sets in ways that sufficiently recognize patterns. We define pattern here in 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, we are interested in those individuals who choose a particular crime category (e.g., rape, robbery, burglary, arson, etc.) and furthermore habitually exhibit the same method of operation through the series of separate incidents. This type of individual is commonly referred to as the serial offender or career criminal (Tracy, Wolfgang and Figlio, 1990; Alan and Wolfgang, 1989; Waldo, 1990; Blumstein et al., 1986).

The accurate and comprehensive identification of problems is fundamental to the problem-oriented policing model. Identification of the career criminal stands out as a particularly relevant problem area. In addition to the intuitive sense of working police officers, the empirical evidence suggests the importance in identifying clusters of incidents associated with the career criminal. The Rand study (Petersilia and Honig, 1980; Rolph, Chaiken, and Houchens, 1981; Chaiken and Chaiken, 1982; Greenwood, 1981), a widely cited research effort in the career criminal literature, found that a rather small subset of the universe of offenders is responsible for a rather large subset of the universe of criminal offenses. Several relevant policy implications ensue from the Rand study findings:

1. Targeting the subset of career criminals would represent a significant improvement in the efficient tactical allocation of police resources.
2. Identification and apprehension of the career criminal would significantly reduce the frequency of offenses, more so than the non-career criminal.
3. Community awareness of the details of a career criminal pattern would improve the likelihood of identification and apprehension through a proactive collaboration between the community and police.
4. Community awareness of the details of a career criminal pattern would improve the likelihood that community members could better protect themselves from being victimized.

In an informal survey we contacted officials of federal, state, and local law enforcement agencies throughout the nation. The objective was to determine if any automated systems existed to proactively identify crime patterns in large crime data sets. We found none and are confident that none exist. ([illegible] done) This is not to say that police departments ignore the career criminal problem and make no effort to discover them. Quite the contrary, most departments devote considerable energy to this end. A look at prevailing methodologies should illustrate this point.

Every department provides officers an opportunity to view data collected in case reports. Often this data is found on clipboards where a hard copy of the original case report is retained. In a large department, this clipboard will contain several hundred cases, particularly in crime categories such as robbery and burglary. The officers look for patterns by scanning the reports for matches in the values of the features of the incident, e.g., physical characteristics of the offender, the location of incident, the type of location, victim characteristics, property taken, words spoken by the offender, weapon used, point of entry/exit, etc., i.e., data collected by officers in case reports and eventually stored in large database files. This might be described as the classical pattern recognition technique. Police officers have been employing this method since departments began collecting data in report form. It is obviously extremely inefficient and ineffective, given the limits of human information processing.

A second method derives from the capability through PC and mainframe database programs to query large data sets to match arrested offenders with incidents possessing similar characteristics. This in some cases will uncover a pattern, albeit reactively. Yet we would not expect to have high success rates identifying career criminals using this method. Very few arrests are the product of a proactively targeted pattern. The arrested offender, therefore, is a virtual random selection from the universal set of offenders. One need not have a sophisticated knowledge of statistics to appreciate the rather low probability that a non-targeted arrestee will be a career criminal, in light of the relatively small number of offenders in the career criminal subset.

Another method of pattern recognition used by many departments and currently quite popular in the police community is the computer-generated crime map (Maltz, 1991; Block and Dabdoub, 1993). Relationships can be explored through the mapped display of data in the hope of identifying patterns. Countless anecdotal evidence points to the successes of this method. While this approach is without question useful, it is limited. Often location variables are sufficient conditions to recognize the activities of a career criminal; however, they are not necessary. Patterns may not be visible in the map. The pattern may disperse over a large area, e.g., the robbery offender who targets fast food restaurants. The pattern may be masked in an already dense area, e.g., an area of the city where many street robberies occur, committed by numerous individuals. In other words, patterns that take shape by similarities of features of the incidents not related to location. Commonality may be found, for example, in combinations of features such as the type of victim selected, the weapon used, the type of proceeds taken. The offender, while armed with a nickel- plated sawed-off shotgun with a, robs elderly female victims cashing their social security checks, committing these offenses in all parts of the west side of the city. In short, instances in which location is not a constituent characteristic of the career criminal pattern.

The problem oriented policing model is clear on process. We understand from the literature the stages in problem solving, i.e., a systematic generalized method of attacking problems. Implicit in the process is the need for problem solvers to possess a reasonable degree of analytic skills. Problem oriented policing models are conspicuously silent on specific analytic methods. We have seen that the methods currently available are insufficient to satisfy the police tactical needs. It should be readily apparent that given the current methodologies and the limitations of human information processing, only the most obvious patterns will be discovered. Substantial numbers of patterns go unrecognized.

This translates to a breakdown in the primary stages of problem solving according to a Goldstein-based model. The task is to identify potential methods and apply these methods to our research problem. The ultimate goal is to overcome the information processing limitations that are nowhere more manifest than our efforts to discriminate patterns in large data sets.

METHODOLOGICAL BACKGROUND

We chose to apply two pattern recognition methods in an effort to solve our problem. Both hold out promise of discriminating patterns of the type found in police data sets.

The first is artificial intelligence. The development of advanced software methods and techniques in the fields of artificial intelligence and expert systems has had a great effect on the field of computing. Artificial intelligence techniques, tools, and methodologies have greatly changed the ways in which the computer is used to acquire and store facts and knowledge. The way in which that knowledge is manipulated has also been altered. Initially artificial intelligence and knowledge-based expert system techniques were used in research environments to establish the foundations of this new science and to develop the tools and environments needed to allow the transfer of this approach to practical applications. Two of the most widely used artificial intelligence technologies in business and industry are knowledge-based expert systems and artificial neural networks.

We proposed building a system which employs the most commonly used artificial intelligence programming technologies to complement each other. Rule based expert systems are the most frequently fielded type of artificial intelligence systems, closely followed by neural networks. Each has strengths and weaknesses. When used together the systems complement each other well.

Rule based expert systems are good at implementing expert reasoning ability by encoding heuristics or rules for problem solving. This procedural knowledge is available for use by the system and is available to the system user in the form of explanations. A neural net excels at pattern recognition tasks.

The trend in artificial intelligence development methodology is to use a number of tools in combination (Newquist, 1990). This allows the synthesized system to combine the best aspects and outcomes of logic and reasoning (expert systems) with pattern matching, recognition, and pattern analysis (artificial neural networks).

A knowledge-based expert system is a computer system which incorporates the knowledge of a human expert (or experts) in some domain. It seeks to emulate the human problem solving capability. Such systems use IF-THEN (or procedural) rules to represent the expert's discovery process. Knowledge-based systems consist of a knowledge base containing these rules, an inference engine which determines the order in which rules are used, and a store of data or facts. These last are frequently obtained by interactive sessions with the user or may be obtained by data base retrieval (Schacter and Heckerman, 1987).

Artificial neural networks are systems that seek via hardware or software to simulate the architecture and workings of the human brain (Rummelhart and McClelland, 1986; Chester, 1993).

Such systems are 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 (Pao, 1989). They learn on a trial and error basis, creating webs of connection strengths. These strengths are weakened by the system if it chooses wrong and strengthened if it chooses right. After it is trained a net can be used to examine data and make recommendations.

While not as common as expert systems, neural networks are frequently used in banking and in the financial instruments industry. They are also commonly used in flight and weapons systems for recognition of features and targets. Other applications include fault diagnosis, medical diagnosis, feature recognition, and robotic control systems (Zurada, 1992).

We chose cluster analysis as the second method to explore. Cluster analysis is the 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 (Romesburg, 1984).

Similarity has taken on varied meanings (Everitt, 1980; Sneath and Sokal, 1973; Clifford and Stephenson, 1975). 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 separate them in an n-dimension Euclidean space. The number of dimensions is determined by the number of variables used to describe the cases 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 decision point regarding variable selection concerns the weighting of variables (Williams, 1971). Which, if any, variables are of greater relevance in describing the case, and to what degree?

Aldenderfer and Blashfield (1984) note several cautions about cluster analysis. They suggest that cluster analysis is relatively simple and unsupported by a broad base of statistical reasoning, in comparison to factor analysis. Secondly, various clustering methods can result in different cluster configurations with the same data. Finally, clustering will impose clusters on the data analyzed. This is an extension of the previous caution. That is, the analytic tool will find groups of similar cases, whatever the data set. Given these cautions, particularly the second and third, it is imperative that the outcomes be validated.

GOALS AND OBJECTIVES

This project proposed to design an analytic tool for police officers with the end goal of identifying patterns that point to clusters of incidents attributable to career criminals. We intended to achieve this goal through the attainment of the following objectives:

1. Focus the investigation on a single crime domain, robbery. Our aim here was to develop an understanding of the fundamentals of pattern formation within a manageable area of criminal behavior. This would provide a beginning point to proceed later to additional domains.
2. Develop an empirically-based understanding of the structure of criminal patterns through a systematic assessment of the decision processes of experts, i.e., police officers.
3. Develop an empirically-based understanding of the structure of criminal patterns through a systematic investigation of data derived from identified patterns.
4. Develop a customized software application utilizing a combination of artificial intelligence methods, i.e., expert systems and artificial neural networks.

5. Develop a customized software application utilizing cluster analysis.
6. Evaluate the existing police database, with respect to collection and storage routines, in regards to its applicability to the methods designed pursuant to this project.

METHODOLOGICAL APPROACH

Stage 1: Focus Groups. The developers of the neural network and the cluster analysis will require knowledge of the structure of patterns in the police data set. To this end we conducted 3 focus groups. The participants in the focus groups were robbery detectives of the Chicago Police Department. Each group consisted of 6-8 robbery detectives, all having extensive experience in the investigation of this offense. Generally, the questioning route concentrated on the decision process employed by veteran detectives in determining which features of a robbery incident structure a pattern. The proceedings of the focus groups were audio-taped and transcribed. The transcriptions were coded and analyzed. A report was prepared describing the patterns and themes discovered in the analysis of the data. The fundamental objective of this task was to elicit the structure of pattern development in robbery offenses according to the domain experts.

Stage 2: Solved Pattern Analysis. To further explore the structure of patterns we conducted the following document analysis: The Chicago Police Department issues pattern alerts, which emanate from Detective Division Headquarters. These patterns are discovered by either crime analysts at Detective Division Headquarters or any of the five outlining ? Detective Areas. We examined identified robbery patterns for the previous 2 years, coded them according to the features that warranted pattern classification, and attempted to empirically establish any patterns that arise among the crime patterns examined. In combination with the focus group findings, we expected this to further clarify the structure of patterns, i.e., the features of a robbery incident that characterize serial offender clusters. The primary target for these findings was to provide a "training set" for the neural network developer.

Stage 3: Neural Network Development. A plan for a neural network for application to crime pattern recognition of career criminals was proposed. There were 10 tasks that composed the development of the neural network software: 1) Obtain the data. It consisted of data sets downloaded from the Chicago Police Department mainframe files and provided to the neural network developer. The domain of interest was robberies and the variables were determined after consideration of the focus group and document analysis. It was expected that data requests would continue throughout the neural network development stage; 2) Randomly select 30% of the data and put in reserve for future testing; 3) Review inputs and outputs for feasibility neural network #1; 4) Feed neural net number 1 to a genetic algorithm; 5) The genetic algorithm will dictate the neural network architecture; 6) Train feasibility neural network number 1; 7) Upon convergence of neural network number 1 conduct a Pareto analysis to determine the ranking of the inputs from "least significant to most significant"; 8) Repeat this feasibility phase at least twice or more depending on the scoring on final tests; 9) Upon a successful tested prototype at convergence with no minima traps occurring design and train final neural network; 10) Complete documentation.

Stage 4: Expert System Development. There were 13 tasks that composed the development of the expert system: 1) Assess AI technologies. There are numerous advanced technologies, more specifically those related to AI techniques. This task involved an examination of the technologies in general as well as the hardware and software tools currently available; 2) Research similar systems. Advanced computing systems in law enforcement are relatively rare. AI investigative systems developed for use by agencies and industry were to be examined; 3) Select software tools. Once the sub-domain of reference is formalized and the requirements set, the software development environment was to be determined and appropriate tools obtained for implementation of the system; 4) Obtain mainframe data; 5) Preliminary data analysis; 6) Formalize knowledge. This involved the elicitation of the knowledge used by expert investigators in the working of their cases. In addition to focus group and document analysis we intended to rely on verbal protocols, observation, and written manuals or guidelines; 7) Design knowledge representation. Determine which type of knowledge representation technique to use; 8) Code rules. Establish, code, and test rules for the expert system.; 9) Review and refine rules. Review rules with experts. Try some test cases. See if rules are correct. This is an iterative process; 10) Integrate expert system with neural net and existing systems; 11) Test on data base. Formal tests will be conducted with trainee police officers. Run rule systems against real records and examine output; 12) Prepare final reports. This report will incorporate the findings of the previous stages; 13) Prepare documentation. User documentation, training procedures, and maintenance procedures will be prepared. These task were not performed as the neural network development failed.

Stage 5: Cluster Analysis System Development. There were 6 tasks that composed the cluster analysis development: 1) Review the focus group results and identified pattern data; 2) Receive data sets from the Chicago Police Department which consisted of all crime data for the calender year 1995; 3) Analyze the data and write necessary code to prepare the data for the cluster analysis program; 4) Determine the relevant variables and their weightings for the model; 5) Analyze the data with nearest neighbor cluster tool; 6) Validate the output.

Stage 6: Intelligent System Development (neural network and/or cluster analysis). There were 5 tasks that comprised the development of the intelligent system: 1) Write specific interface and requirements; 2) Determine methods of user access; 3) Design system, finalize embedment language (C++ or VBX); 4) Code system; 5) Write test criteria for final acceptance of complete Intelligent System.

The remainder of this report describes in detail the 6 stages discussed above. Chapter 2 describes the method and findings of the focus groups. Chapter 3 discusses the efforts to develop a pattern discriminating tool through artificial intelligence methods. In addition, we seek to explain how our findings suggest that this method is likely inappropriate for use in solving police type pattern recognition problems. Chapter 4 documents the cluster analysis efforts and the development of the interface. Chapter 5 proposes several recommendations for improving police data collection and analysis and suggestions for future efforts.

II. FOCUS GROUP

INTRODUCTION

Over a period of approximately thirty years, researchers have begun to develop, albeit tentatively, a picture of the operations of criminal investigators. The empirical evidence has led to interesting, and conflicting, conclusions.

Isaacs(1967), Greenwood(1970), Greenwood et al.(1977), Eck(1979), and Gaines et al.(1983) tell us that the efforts of detectives have little to do with the clearance of a case. Eck(1992) calls this position the circumstance-result hypothesis, which maintains that detectives have little control over the factors that lead to case closure, e.g., existence of a good eye witness, whether the offender and victim knew each other, the willingness of witnesses and/or victims to cooperate, etc. Information that is most useful comes from the victim and is collected in the original case report. The great majority of arrests are made in close proximity to the scene of the offense and made generally by a patrol officer. This has led some to question the value of specialized detective units.

Eck(1983) proposed that cases could be categorized into three types. The first were unsolvable. The second type were those cases in which an offender was named and the task was simply to locate and make the arrest. The final category consisted of cases that, with some measure of investigatory expertise, could be solved. Eck concluded that effective use of a variety of information, including case report data, could bring solvable cases to a successful conclusion. Eck(1992) called this the effort-result hypothesis.

Case report data, traditionally, have served to describe the nature and extent of crime as well as to identify crime patterns. For purposes of this investigation, we are concerned with pattern identification, i.e., pattern in the police sense as discussed previously. The type of case report data collected by both large and small departments is very similar. The officer taking a report from a victim collects approximately thirty characteristics of the incident. These include the address of incident, the date and time, the type of location (e.g., the street), characteristics of the victim (e.g., sex, race, age), characteristics of the offender (e.g., sex, race, age, and various physical and clothing descriptors), the vehicle used by the offender, the weapon used by the offender, the proceeds taken, etc.

The Rand study, as noted above, found that a small subset of the universe of offenders is responsible for a rather large subset of the universe of criminal offenses. Their findings led them to conclude that approximately 15% of the universe of robbery offenders were responsible for approximately 75% of the universe of robbery offenses. Assuming that offenders behave habitually, that is persist in committing the same type crime(e.g., robbery), with the same method of operation, we should expect to see patterns develop in the police data set, i.e., data contained in the original case report.

In order to discover a pattern in the data, the detective looks for combinations of similar characteristics over a series of incidents. For example, elderly females are robbed of their social security checks on the northwest side of the city, outside currency exchanges in mid-afternoon, by a male white offender approximately 6 feet tall and 250 pounds, who displays a nickel plated sawed-off shotgun. Over a period of several months, a detective discovers approximately 15 incidents with these matching characteristics. He is relatively certain that all these incidents are being committed by the same individual, a career criminal.

In a small suburban or rural police department the task of identifying a pattern of like incidents is manageable. Given the relatively small volume of cases, patterns tend to stand out. For officers in a large urban department, the problem is somewhat more complex. At present, there are no automated methods of identifying patterned incidents. Each detective must manually scan hard copies of case reports on a board containing hundreds of reports of incidents. The detective must somehow keep stored in memory the thirty or so characteristics of each case report. He/she must then compare all characteristics of all cases against all other cases, looking for combinations of characteristics that are sufficiently similar to warrant classifying those reports a pattern. In Chicago, over a period of one year, approximately thirty thousand robberies are reported. Detectives seeking patterns in this sea of data are faced with a formidable task. In 1995, less than 1 percent of all robberies were classified as pattern incidents in Chicago.

The purpose of this research was to discover how expert detectives managed to identify patterns. As a product of this exploration, we hoped to satisfy five objectives: The first and most straightforward was to discover the methods used by detectives to identify patterns in the data, in this case robbery data. The second was to establish the relative value of characteristics of offenses captured in case reports. Next we intended to identify the weaknesses and strengths in the data collected. Fourth, we sought to determine the information capacity of detectives in utilizing police records in the solution of crime. Finally, we endeavored to identify the structural features of the organization that impacted data analysis.

METHOD

In order to explore this question we chose to employ a focus group method. Focus groups are particularly efficient tools for mapping heuristics that individuals use to make decisions. We identified eighteen robbery detectives, each with a minimum of twenty years of service, and each consistently rated in the upper 10% of his unit on department performance ratings. In addition, robbery detectives from the five detective areas were polled informally and asked to identify peers believed to be the most effective in pattern recognition skills.

Three focus groups were held. The first panel served as a pre-test for the questioning route and tasks we asked the detectives to perform. As a result of this dry run the questioning route and tasks were refined and subsequently administered to the second and third focus group. The first group had seven participants, the second group five participants, and the third six. All the participants were male, white, and ranged in age from 45 to 52 years of age. All were currently assigned to the detective division in various capacities and further had a minimum of twenty years experience as a detective. All had been formerly assigned to specialized robbery units throughout the city.

The second and third focus groups were audio recorded and the tapes transcribed. Each of the three focus groups lasted approximately two hours. The transcripts were then imported into a qualitative software program, where the responses of the participants were coded and analyzed.

A moderator administered the questioning route, which included two tasks for the detectives to perform, the first a sorting task and the second a pattern identification task. The object was to discover the priorities detective assign to both individual characteristics of an incident and combinations of those characteristics to identify patterns.

In the first task the detectives were given cards, each of which contained one characteristic of an incident (e.g., victim age, offender age, location of the incident, etc.). The participants were then asked to prioritize the characteristics from the most important to the least important, filling out a pre-printed form. After completing the sort task, each participant was asked to present his results and discuss his rationale. After all participants had completed this, the moderator opened a discussion to establish the reasoning for any differences discovered in the prioritization. The moderator made an effort to reach group consensus on the rankings, identifying several highly important characteristics, several characteristics considered useless, and the rest falling in a mid-range.

After sorting for single characteristics, the participants were asked to consider grouping characteristics to form MOs. The object here was to determine the experts notions of how patterns structure in general terms. Were there combinations of characteristics that appeared to be more useful in discriminating patterns than other combinations? Again, an effort was made by the moderator to arrive at some consensus among the participants.

The next section of the focus group was designed to establish the perceived quality of data collected from victims in the case reports. The moderator probed the participants to discover any perceived variance in the quality of data from one characteristic to another. In those instances when data were found to be consistently reported inaccurately (e.g., the physical characteristics of an offender), the moderator asked participants to estimate the range of these inaccuracies.

Finally, a stack of 21 hard copy case reports were given to the participants. Within the 21 cases there were 4 previously identified patterns. The 21 cases were mixed and the moderator asked the participants to find patterns within the pile. After completing the task, the participants discussed how they worked through their sorting process.

FINDINGS

Structure

Throughout the sessions, the participants spoke of a variety of structural features of the department that influenced pattern recognition behaviors. The institutional arrangements of the detective division were found to have a substantial influence on the data processing efforts of the respondents. We believe this sets the stage for specific pattern recognition behaviors, which we discuss in more detail later.

In Chicago, detectives are assigned to one of five geographically-based detective units(called areas) in the city. Each detective unit consists of four, and some five, geographically-based patrol units(called districts). Each detective area is divided into violent crimes and property crimes units. Robberies are the responsibility of the violent crimes unit.

Violent crimes detectives are generalists to the extent that they may be asked to investigate any violent crime incident, i.e., murder, robbery, and serious aggravated battery. Although in some detective areas there are detectives that informally specialize, for all practical purposes the detective division is designed according to a generalist strategy. The exception is sexual assault incidents, which are assigned to specialists. Case management sergeants assign cases to detectives for follow-up investigation.

In the violent crimes unit there are binders into which copies of all case reports(by crime category: robbery, criminal sexual assault, aggravated battery, and murder) are inserted. This provides a opportunity for detectives to become aware of criminal activity throughout the area. For robberies, a binder exists for each of the patrol districts in the detective area. The case reports are placed into the binder in reverse chronological order, the most recent cases at the top. All case reports are inserted and cover a period of roughly 3-4 months.

Each binder then contains approximately 3-4 hundred case reports and since there are 4-5 districts in each area, at any given time there are from 1200 to 2000 active case reports in all.

As noted above, field detectives find patterns by scanning the binders looking for case reports that are characterized by similarities between characteristics reported in the original case reports. This requires the detective to store in memory the characteristics for each case report and match those against all other case reports. In order to find patterns that might overflow into other binders it is necessary to expand the memory capacity to include those surrounding binders containing case reports from adjoining patrol districts.

Each area has a supervisor that is designated as a robbery coordinator. The duties of the robbery coordinator varies from one detective area to another, however is similar in one respect. This person is expected to maintain an awareness of the robbery conditions in the area. This entails review, albeit often cursory, of all robbery case reports. The primary purpose of this review is to maintain the modus operandi(MO) file, which is simply a collection of hard copy case reports classified and stored in a file cabinet. The type and quality of the MO file varies from one area to another. In all areas, however, the primary goal is to divide robbery case reports according to a crude classification scheme. Cases are classified geographically,(by patrol district of occurrence), by offender description(generally by race and sex), and by type of location of incident or victim(e.g., currency exchanges, cab drivers, banks, fast food restaurants, etc).

The MO file is more generally used in a reactive fashion. The file is routinely employed after an individual is arrested. An arresting detective will inspect the MO file to determine if an in-custody offender fits a category. When a "hit" is found, the MO file facilitates a more rapid response to multiple clearances. Victims of similar type incidents can be readily identified and asked to view the offender in line-ups. This is not to say that the MO file can and is not used proactively. On occasion a single category will stand out as a pattern; however, ordinarily the categories are too generic to be considered linked to a single individual or group of individuals.

The detective division also has a crime analysis unit, located at detective division headquarters. This unit serves both an administrative and analytic function. In its administrative capacity the analysis unit reviews patterns that originate in the districts or detectives areas, most coming from detective areas. The crime analysis unit determines whether the suspect pattern warrants a citywide notice, in other words, an official department designated crime pattern. In its analytic capacity, officers in the unit seek to unilaterally identify patterns, either that have been missed by patrol and detective units, or identified patterns that overlap jurisdictions.

Officers identify patterns in the same fashion as area detectives, i.e., manual review of hard copy case reports.

Incentives

The focus group participants pointed out several structural changes that have served over time to modify the incentive structure for detectives. Prior to 1980, the Chicago Police Department Detective Division assigned personnel to specialized units. Detectives were assigned to one of four positions: homicide/sex, robbery, burglary, and general assignments. Each of these specialized areas of investigation constituted a separate unit within the larger detective area. The area as a whole was supervised by a commander. Each of the specialized units had a commanding officer who reported to the area commander. Both command positions were considered high status ranks at the middle management level.

The performance of the area and unit commanders was measured by a variety of standards; however, clearly the most important were arrests and cases cleared. For this reason, particularly in the robbery units, multiple clearances were prized events. Commanding officer in the robbery units therefore created a number of formal and informal incentive structures to maximize this form of activity. Robbery detectives were required to submit multiple clearance reports. The division had the capacity to enforce this job requirement by their ability to demote, without cause, any detective. The introduction of a union has made demotion from the detective rank an extremely difficult control mechanism for management.

Participants further agreed that when the division was specialized identifying patterns was a task that resulted in a good deal of job satisfaction. Unit members that were proficient in the identification of patterns were considered by their peers as having an admirable talent, and one that often was useful in tying together multiple cases for multiple clearances, thereby improving the standing of any detective who could reap the benefits of a detected pattern.

Also the participants pointed out that the hierarchical structure has changed since the 1980 reorganization. There are no longer specialist commanding officers and sergeants. A violent crimes commanding officer is now responsible for homicide, sex, aggravated battery, and robbery investigations. There is simply less time to monitor the activities of subordinates. Robbery multiple clearances are no longer demanded as they were. The performance of superiors, according to the perceptions of the participants, is not tied to this form of activity as it once was. Consequently, the supervisor-generated incentives and disincentives have changed, particularly with respect to multiple clearances.

Most important, according to the detectives, specialization meant that detectives had the opportunity to become more familiar with robbery conditions since these were the only case reports they reviewed. Additionally, the social environment of the station house dictated that specialists tended to interact with members of their units rather than members of other units. That is, robbery detectives tended to socialize with other robbery detectives. Conversations often centered around crime conditions associated with their speciality. Because of these changes in institutional arrangements, detectives have become less immersed in the data relative to robbery incidents. According to one detective

> When we had robbery units guys would talk about who was out there working robberies. Who was in the joint. Who just got out. We would pretty much stick with guys from robbery. The homicide guys were off by themself. It was different then.

Another participant added

> It's hard now to find a pattern. I get burglary, homicide, robbery, phone harassment jobs. You name it, I get it. How can I keep track of patterns of robberies. It can't be done.

Also, these participants perceived that there has been an erosion of interest on the part of department command staff about the importance of some crime categories, particularly robbery and burglary. The emphasis, they believed, is currently on homicides and gang-related shootings. Multiple clearances are neither rewarded nor required to the extent that they were when the units were specialized. The comments of 2 detectives are illustrative:

> If they aren't dead then is doesn't matter.
>
> The department cares about what they think the public wants, and right now that's murder and gang-bangers.

Pattern recognition heuristics

Irregardless of these organizational changes, all the participants were in agreement that the identification of patterns in the data was extremely difficult under the best of circumstances. All, whether aware of it or not, had developed heuristics to manage this seemingly insurmountable chore.

There was general agreement among the participants that few individuals commit the great bulk of robberies. When asked to prioritize the characteristics that they considered the most important in identifying patterns, there was a good deal of agreement. Among the most important reported characteristics was the type of robbery, either armed or strong armed, the type of location, and the geographical location. Several comments illustrate this:

> Offenders stick to commercial targets or they are street robbers. When I'm sorting through the board looking at jobs I keep my eye out for guy's that are doing stores, liquor stores or gas stations, or whatever. It seems that somebody that is out doing a pattern quite often sticks to what they are successful robbing.
>
> My thinking is that if someone is doing

> robberies either doing store robberies or commercial robberies or maybe street robberies. They'll stick with one or the other.
>
> Yes, I look at the geographical location. They don't normally stray too far where they're doing the robberies.

These detectives further believed that physical description was also very important, consistently ranking at or near the top. Their primary concern was the race and sex of the offender. This they believed to be reliable case report information. However, other physical characteristics(e.g., height and weight), were considered less trustworthy. The participants were aware that people generally, not to mention someone under the stress associated with being the victim of a robbery, had difficulty in accurately describing the physical characteristics of others. In order to compensate for this deficiency, the detectives used a form of fuzzy logic. That is, they classified offenders as tall, short, fat, thin, old, or young.

At the other end of the spectrum were those characteristics that the participants felt were least useful in discriminating patterns in the data. There was uniform agreement that both witness information and the relationship between the victim and offender were of no use in developing a pattern.

> The witness, anything about the witness at all is useless...The witness is a witness, he just happens to be there.
>
> In a case of a street robbery or a reported street robbery it indicates that the people

> are acquainted, you're probably not dealing with a real robbery.

Several additional characteristics served as a tentative indicator that an incident is probably not part of some larger patter of incidents, although with somewhat less assurance than the above noted two. Detectives view incidents in which the victim had been drinking or cases in which the date reported is days after the actual incident with a degree of skepticism.

> That's one of the biggest ones I look for. If it occurred on 9 April and at 3:10 in the morning and you've been drinking out at a tavern, but he doesn't report it until the 11th or something like that. And that's an indication that you know, you better talk to him because maybe there's something here that's not right, you know? Maybe he lost the rent check and didn't want to tell his wife. Who knows.

Participants generally agreed when asked to prioritize characteristics of an incident at either of the extremes. There was also a good deal of consensus regarding the placement of characteristics in the mid-range. Several of the variables falling into this range included the victim, weapon, day of week and time, and proceeds taken. The rationale for placing characteristics in the mid-range was that they were neither always reliable, as in race and gender of an offender are, nor were they always useless, as in the witness and relationship data. These characteristics, rather, were placed in the mid-range because of the uncertainty of their value. This goes to the heart of pattern recognition operations. The point is that configurations of patterns are ever-changing. This mid-range set of characteristics will very often be the defining characteristic, or combination of characteristics, of a pattern, e.g., the proceeds taken, or the weapon used. Yet, in most instances, they are of little or no value. As one participant put it:

> I put proceeds in the middle. You're going to get what you can get. There are times where that becomes important. There are times if someone is asking "give me your watch" or "give me your starter jacket." These become important if there is a break in the normal routine of "give me everything that is in your pocket." That makes that guy stand out.
>
> It's a crime of opportunity. The guy might get hit on the head with a bottle or it might be a brick or whatever the guy picks up. But I think that the weapon might sometimes be important. And date of crime and time of crime. That could be important, but I put it in the middle. All these things might be very important.

Another characteristic that fell into this mid-range was the narrative portion of the case report. In this section of the case report the reporting officer describes the incident in a brief narrative fashion. The quality of the narrative section, according to the participants, varied from one report to another. It was considered useful to the participants only on occasion, and was often not looked at in the initial scan of case reports. This portion of the case report was often visited by the participants only after some other characteristic or combination of characteristics had been identified as a suspect pattern. The narrative was then primarily used as a confirming mechanism.

The description of the offender's vehicle fell into a rather special category and began to illustrate the short-cut methodology of detectives. The participants maintained that this characteristic is rarely reported. Yet when it does appear it may be a very useful variable in identifying a pattern.

> You don't see vehicle information too often. But when I do see a car on the report I take a real close look. I look at other reports and see if a similar car is on any other reports. It's easy to do since there aren't too many. If you see that yellow Chevy with a smashed in front end in a couple of case reports you can be pretty sure you're on to a pattern. That one used to work for me pretty good.

This notion of identifying short-cuts in the discrimination of patterns was not confined to targeting vehicles as a primary characteristic. The detectives agreed that as one scanned through the volume of case reports, the best one could hope for is to observe something obvious in the data, a characteristic in a report that was easily distinguished from the same characteristic in other case reports, or for that matter combinations of characteristics.

> There has to be something because of the volume of robberies. Something has to jump out at you as you are sorting through the cases.
>
> We have to begin work with what is the most obvious thing and the fastest. You head for the easiest, and most obvious thing.

Detectives stated that they may look for someone who is very tall, very short, very heavy. They may look for a vehicle as was already mentioned. They may look for cases in which the location is always a convenience store or gas station or close to a school. The victim may be all older ladies. The weapon may be unusual. The process then generally begins with an obvious characteristic that repeats over cases, and furthermore one that for some reason to the detective stands out from the rest. The comments of one detective typify this method:

> You start with one thing, say a guy in a red beret. What you want to do is find an additional case report where you think it's the same offender. So now you have two and you are going to expand on the clues. And then you keep expanding from there so there is immediate unusual characteristics, but you want to find two cases with the guy in the red beret. Now you have two to take with red berets and so now you extend the parameters. Now say you have guys with dark complexion. And so on.

The initial identification of the "obvious" characteristic in most cases began with the detectives paying particular attention to oddities in the physical characteristic of the offender, the type of location(usually a type of business), and the geographical location(usually the beat of occurrence). Some, however, seemed to have favorite characteristics that they were particularly attentive to. One example, noted above is the detective who focused on vehicles. Another detective remarked on his interest in gas station robbery crews, taverns, or convenience stores. The point made by the participants was that pattern characteristic configurations have a tendency to repeat themselves, albeit with different offenders. And more importantly, the identified pattern recognition begins with a characteristic or combination of characteristics that are extraordinary.

CONCLUSION

The most striking conclusion is that a comprehensive discrimination of robbery patterns by manual means is likely beyond the information processing capacity of an individual detective. This is not to say that patterns are not identified, for they are. The detectives that we interviewed maintained that patterns were culled from the large data sets. The question is, given the information processing constraints, what sorts of adaptive behaviors did these individuals exhibit?

The detectives first came to recognize that the large data set could be divided into types of cases that adhere to the principles of exclusivity and inclusivity. Rather than comprehensively analyzing the entire universe of cases, the set could be reduced to a smaller number of subsets. Some of the types were obvious, while others required a more sophisticated understanding of the nature of the criminal incident and the data.

For example, it was obvious to these detectives that one could type cases by sex and race of the offender. Both of these variables were first of all likely to be accurate as reported by the victim. Secondly, it is unlikely that offenders, by ability or inclination, will seek to change their racial or sexual appearance from one offense to another. The detectives have now reduced the number of case reports to analyze from one very large data set to approximately twelve(6 racial categories times 2 sex categories) mutually exclusive smaller ones.

The next step taken to reduce the processing costs was related to the work experience of the detectives. In this case each of the twelve primary categories could be further divided according to the type of robbery(armed or strong arm), the location(e.g., street, store, tavern, etc.), and the location of the incident(by beat). This categorization was possible because their work experience permitted a probabilistic view of the data. Their experiences convinced them that it is unlikely offenders will switch from armed to strong arm robbery, that they will likely choose a type of target(e.g., grocery store, person walking on the street, etc.) and stay with that target, and that offenders tend to remain within a relatively narrow geographical perimeter to commit their crimes. Believing these assumptions to be true, the detectives were able to increase the number of types and thereby decrease the case reports in each subset. The analytic task had been further simplified.

A combination of common sense and practical experience permit the detective to organize the data set in a way that adapts the analytic methods to the information constraints of those performing the analysis. The data set is thereby transformed from a very large, unwieldily one, to many smaller, manageable ones. To this point the task has been relatively simple, requiring straightforward sorting tasks only. More sophisticated information processing skills are required to enhance the precision of identified patterns. The ideal level of precision is that cluster of cases attributable to a single offender or group of offenders. In an effort to move toward that end detectives rely on several decision heuristics.

The most direct method of pattern recognition is to proactively identify a set of pattern "types" and then scan cases to match the pre-selected pattern configuration to the data reported in a case report, a straightforward data base query. Each new case found with characteristics matching the predetermined pattern configuration would be placed into the pattern. In order for this method to succeed one would have to identify this set of recurring patterns that persist over time. The detectives maintained that like combinations of characteristics did not repeat with any degree of regularity sufficient for this ideal predetermination method. This makes sense, given that there are approximately thirty-five reported characteristics to an incident and each criminal will vary in the method he or she chooses to carry out their criminal act, not to mention the variations of the physical characteristics of the offender.

While pre-selection in its purest form was not used, a clever variant was. Detectives, recognizing their information processing limitations, realizing a comprehensive scan of the data was impossible(even given the reduced data sets), endeavored to devise a method that adhered to the spirit of pre-selection. The first analytic pass at the reduced data sets effectively ignored characteristics within the "normal" range and focused on the tails of the distribution, those characteristics seen infrequently. The detectives looked for the "obvious:" some offender physical characteristic out of the norm, e.g., very tall, very short, very young, very old, very thin, very heavy, etc.; robberies in which the offender wore some unusual clothing; robberies in areas of the city where they were uncommon; robberies where a vehicle was described, etc. Using these methods will likely result in the identification of only the most glaring patterns. Those contained in the normal range will probably be missed. Moving away from those patterns that are found in the extremes requires a substantially more discriminating analysis of the data, and quite possibly a more comprehensive description of the incident.

The remarks of the detectives further indicated that information processing may well be a function of the organization's institutional arrangements. In the past the department has, either by design or inadvertently, addressed the information processing deficiency through a variety of institutional arrangements. The designation of specialized robbery units is a case in point. The existence of the specialized unit created an organizational culture that facilitated collective data analysis.

In the specialized units detectives were, as noted above, assigned to investigate only a single crime category(e.g., robbery, homicide, sex, etc.). This effectively reduced the information processing requirements of the members of the unit. Detectives were able to limit their attention to robbery cases only. Their first preliminary search was confined to a rather small subset of all robberies, i.e., those assigned to them for investigation. Upon locating a suspect pattern, their normal social conversations with other robbery detectives could ascertain if similar small clusters of cases assigned to other unit members expanded the suspect pattern. With the introduction of generalist detectives, this pattern of social contact and case assignment methodology vanished.

A second feature of the organizational culture concerned the incentive structure that encouraged the search for patterned incidents. The hierarchical structure under the specialized detective division consisted of a network of incentives and disincentives that encouraged pattern recognition behaviors.

These incentives and disincentives were a product of a clearly articulated division policy that emanated from command staff and filtered its way to mid-level managers. The clarity of policy was supported by a chain of command with a focused mission(e.g., robbery unit productivity measured by cases cleared), thereby securing the utmost benefits of specialization, at least with regard to pattern recognition behaviors in robbery units.

It is unlikely any time soon that the current generalist structure preference will change. The dominant view in police administration is to move away from the police officer as a specialist. Rather than being driven by incidents, the police officer is now asked to devote his or her attention to a particular geographical area. The officer is further expected to become familiar with "problems" associated with that area and in collaboration with the community devise solutions. The literature of community policing has urged this method and discouraged the perpetuation of specialized units.

This suggests that the benefits to be derived from specialization relative to pattern recognition are not likely to be realized. In any case, these findings imply it is doubtful that manipulation of the institutional arrangements would result in anything but the recognition of the most flagrant patterns. It is possible that reducing the geographical area of concentration for a police officer, as is recommended by the proponents of community policing, will reduce the information processing requirements of a detective and thereby permit more efficient pattern recognition. There will be fewer cases in each category, thereby simplifying the pattern discrimination objective; however, this will likely be offset by the addition of crime categories. In either case, an information overload is probable.

The application of sophisticated statistical methods is one approach to overcome this problem. Police departments recognize the need for this type of support. Very little, nevertheless, has been done to develop systems devoted to automated pattern recognition for tactical purposes.

While we may recognize recurring community crime problems, it is quite another matter to devise tactical solutions to address them. This demands an understanding of the problem from an analytic perspective. We may know a particular beat is plagued by strong arm robberies. It would be useful to break this series of robberies into its component parts to better understand the nature of the problem, particularly, how many individuals are committing these robberies? Answering this question, as we have seen, is beyond the capacity of police officers, especially in large urban departments with correspondingly large data sets. The manipulation of structural features of the organization may ameliorate the problem somewhat, but not to any meaningful degree. An automated pattern recognition tool is clearly indicated.

III. ADVANCED COMPUTER SYSTEMS AND LAW ENFORCEMENT

INTRODUCTION

Computer Systems are commonly used by most law enforcement agencies in the traditional ways, mainly in the data base and information retrieval arenas. The foremost use of computer assistance in law enforcement information management is in the area of forensic analysis. Computerized communication systems help to link these data stores for transmission of law enforcement data. Nixdorf Computer provides such a system to law enforcement agencies. The system include a Law Enforcement Management System (LEMIS), Paperless Information System Totally Online (PISTOL), Computer Aided Dispatch (CAD), and Computer Aided Complaint (CAC), as well as online booking.

who? why are they useful? more explanation

The use of advanced computer technology has made its way slowly into use in only a few areas. The most advanced systems in use are probably the forensics systems used by the IRS and FBI crime labs. These systems use computers to aid in the analysis and recognition of evidence such as documents, voice recordings, photos, and fingerprints.

The use of advanced computing technologies in investigation, however, has been slow. Few systems have been developed in this area (Newquist, 1990). The most prominent among these are:

1) DIANA, created by Bolt Beranek and Newman for use by the

Home Office in the United Kingdom. This system was created for the analysis and interpretation of intelligence data gathered by the police. This system uses both pattern recognition and knowledge based techniques to analyze existing data, assist analysts and prepare reports.

2) "Big Floyd", an expert system created by the FBI to assist agents in investigating labor racketeering cases.

3) PROFILER, an FBI system which creates personality profiles of likely perpetrators of violent crimes. The system examines data on the crime, method of commission, victim, location, and stored reports and presents a profile of the perpetrator.

A number of similar systems have been developed by other agencies, industries and companies for use in investigating crimes, chiefly financial fraud. (McGowan, 1991; Mena, 1994).

Over 60% of the top 100 U.S. insurance companies are using or developing expert system or Knowledge Based System capabilities (Shipilberg 1986, Leinweber 1988, Martin 1988). Most such systems are now being used in underwriting, sales, and auditing activities (Automation Review, 1988). Such systems are finding increasing use in the area of targeting fraudulent activities. Major credit card companies have also developed such systems (Rothi and Yen, 1990; Fryer, 1993). These systems and others are discussed in detail in the chapter on past research in this field.

Use of an investigative expert system is not meant to replace the experts in the field, nor to direct their decision making. Rather, such systems should provide decision tools to be used in the decision making process. The use of the proposed system is seen as an extender of the investigator's memory and a compendium and retriever of crime and investigative information and knowledge. The system could also be used as an aid in instruction of new investigators. It has been recognized that the use of computer models and systems, as supplements to the physician's clinical decision making abilities (diagnosis and treatment planning) have definite advantages (Keen & Scott-Morton, 1978). We feel that similar benefits may be obtained in developing such a system for use in the identification and investigation of criminal activities.

The list cited above includes knowledge based systems which were developed for interactive use by investigators in various domains. The development and use of AI systems in an enterprise requires technical and organizational change in the way that systems are developed and implemented (Liebowitz and DeSalvo, 1989). It also requires a cultural change in the way that a corporation perceives the uses and value of its systems (Frenzel, 1992). The problems of implementing such systems in large urban police departments are formidable. Technical problems are easily overshadowed by managerial, organizational, cultural, and political considerations. Although intelligent systems show promise in this domain they must be fielded in such a way that the cost of information provided is minimal. This low information cost may be more easily obtained by making the tools available to specialized crime analysts who can abstract and report on the findings made available by the intelligent systems.

NEURAL NETWORKS

Definition: Neural Networks are systems that seek via hardware or software to simulate the architecture and workings of the human brain (Rumelhart and McClelland, 1986). 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.

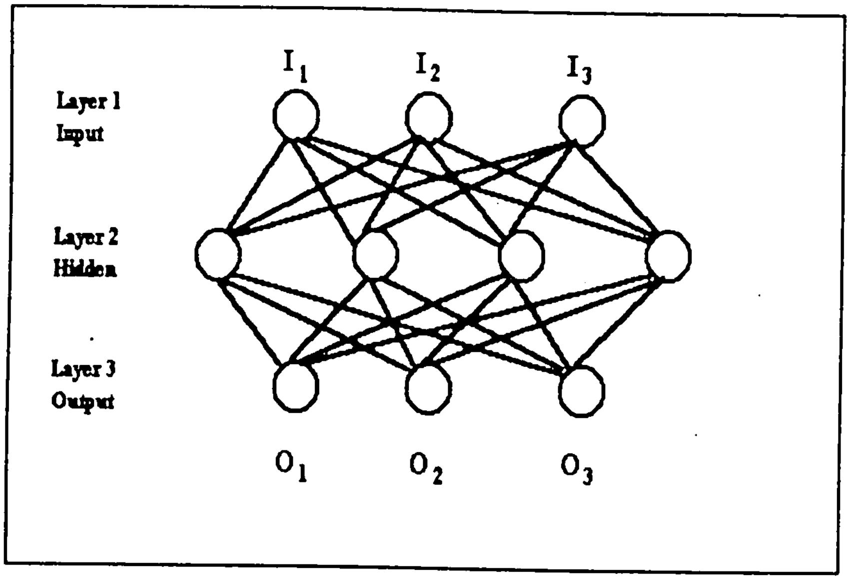
Architecture: Neural networks may be thought of as directed graphs in which processing elements (neurons) are nodes on the graph. Arcs on the graph represent the flow of activation through the net. Generally these are weighted by some algorithmic scheme. Nodes are grouped together in layers. Each node of a layer may have one or more input arcs and one or more output arcs. A node will use input values to compute an output. Nodes are usually connected via arcs only to nodes on layers above or below them. The number of layers, and nodes per layer will vary dependent on the type of problem and the particular learning algorithm used.

paradigms explored in this paper

The Neural nets operate by the presentation of an input pattern which is processed by the net. All nodes of any layer receive stimulation from the input pattern. The output of any node is based on the input strengths. Input strength above a certain level will activate a node to create an output signal. General processing occurs via the stimulation of internal nodes by the input node layer. There may be one or more layers of internal nodes. Finally, nodes in an output layer are stimulated to present an output pattern.

Figure 1 represents the structure of a simple multilayered neural net. Input data are fed in to the Input Layer. Processing proceeds through the Middle Layer and ouputs or categorizations are propagated through the Output Layer. Each node (represented as a circle) is a processing unit. Arcs connecting nodes have weights assigned to them by the use of certain computational algorithms. These weights change as the net is trained.

Figure 1. Simple Neural Net



TRAINING AND CONFIGURATION

Neural nets must be trained to identify patterns before they can be used to recognize or categorize unknown patterns. That is, the connection weights linking nodes across levels must be set. This is done by training. Humans can select the number of nodes and layers, number of inputs and outputs, and training patterns used. Patterns are usually presented to nets automatically. Humans also can present the patterns (or input data) and score the outputs. But, it is the neural net ~~itself~~ algorithm which adjusts connection weights during training.

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 weighting pattern.

This training of neural nets via connection weight adjustment is accomplished via the use of one of a set of learning algorithms. The most commonly used of these algorithms is the backpropagation algorithm. This is a general purpose learning algorithm. It learns by attempting to minimize the error of the output. Learning generally proceeds according to the following algorithm:

- LOOP Until trained
- Input pattern presented to Input Layer
- Pattern is propagated through Internal Layers
- Output pattern activated at Output Layer
- "Grading" of output (Correct? Incorrect?)
- Error at output is calculated (sum of errors of each node)

- Nodes modify input weights so as to minimize error
- END LOOP

This modification proceeds, or is propagated, backwards through the net, thus the name. The backpropagation net starts off in a randomized weighting configuration. Learning usually proceeds until the net can successfully give appropriate outputs to training inputs. At this point the connection weights are ~~set~~ fixed and ~~will~~ do not change in response to input patterns. The net may then be put into production by presenting new input patterns. Those patterns identical to training examples should elicit, as output, the pattern associated with the input pattern on training. When an new input pattern differs from training input patterns the net will produce an output resembling one of the training output patterns, usually the closest that it can match.

GENERAL SHORTCOMINGS TO THE USE OF NEURAL NETWORKS

System Development Methodology: One of the major shortcomings of Neural Networks is that there are no established guidelines for choosing the technical parameters used to control the inner workings ~~of these systems.~~ They are programmed by trial and error in many cases. They do not incorporate rules or procedures and, so, cannot explain their reasoning. Because of this it is frequently impossible to tell if a Neural Network, even one that scores 100% on training tests, is actually identifying those features which its developers believe it to be identifying.

Knowledge Representation: The Knowledge Base of a neural net consists of the connection weights generated automatically in the training of the net. There is no means by which we can examine a net and determine the purpose or meaning of a node or group of nodes in the net. In addition, a Neural Network has no built in capability to explain exactly why a choice or decision is made, nor to trace a logical reasoning path. Such systems are most successful at learning and recognizing relatively stable patterns of data. The problem with the Police Department crime data is that it is ***fluid*** and ***dynamic***.

The knowledge representation of a neural network is implicit. For example, the net may give correct answers, but it is difficult to ~~there is no way of~~ identify~~ing~~ the purpose of a neuron in the net. The implicit knowledge is hidden from the developers due to its algorithmic nature. It is also nearly impossible to match the implicit knowledge to the explicit knowledge used by human experts. This explicit knowledge of problem solving in the domain can easily be represented as heuristic rules.

DISCUSSION OF FINDINGS RELATED TO NEURAL NET

Learning Modes

FINDING: Employing supervised learning with undifferentiated

crime patterns resulted in suboptimal performance of the neural net. There were no visible patterns to be used as training patterns. Unless the data base is changed to include a more comprehensive collection of variables, there is little chance that patterns can be abstracted for use in training a net.

One of the major findings of the research into the use of neural nets for categorizing crime patterns relates to the learning mode used by the net. There are two basic modes of learning: 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.

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.

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.

Choice of Neural Net Architecture

FINDING: The Back Propagation Network architecture is insufficient to the task of differentiating patterns of activity given the data available.

The neural net architecture used in this research was the Back Propagation net model. This is the most commonly used type of net. However, it has shortcomings. This type of net is best used when a system must *generalize* input training patterns into categories. The greatest use of such networks is in the area of signal processing. Sejnowski(1987) used such nets to convert written text into phonemes. This type of categorization succeeds because a net can identify input patterns and categorize based on the matching of input patterns to those on which it has trained. In the example cited, phoneme input patterns are known and may easily be presented for training.

The data used in this research yields no such representation of patterns. Our discussion of the data has shown that the CPD data base does not contain data at a sufficiently fine level of detail to allow us to categorize patterns of crimes which could be attributed to individual criminals. We are however able to identify gross levels of criminal activity fitting certain higher level patterns or categories.

FINDING: The Back Propagation Network architecture will not be suitable to the task over a lengthy period of time.

A single input pattern (data from one case report) might possibly fit many existing patterns. For example, there will be many known patterns of strong arm street robberies committed by young black males. These will involve elderly ladies as victims, with the proceeds being the purses of the victims. These patterns will occur in many of the city's police districts, but are not being committed by the same criminal. Each time that a new case pattern of this type is entered into the net, it could fit numerous categories or patterns to greater or lesser extent. This is a one to many mapping.

This problem of conflict is not handled well by Back Propagation networks. One possible neural net solution to this problem may be the use of the Elman network (Freeman, 1994; Skapura, 1995). This network adds a number of "context nodes" to the input pattern layer. These are used to establish contexts for inputs. They may allow differentiation of patterns occurring at different times.

In effect our system added context by preprocessing the input data using some simple heuristics and then feeding the data through a nearest neighbors pattern recognition component. The output from this component was far more successful at grouping potentially related patterns of criminal activity than was the neural net.

FINDING: Neural nets will require frequent retraining when new patterns are to be examined. There is no science for determining how often and to what extent this should be done.

When a Back Propagation network is trained it is presented with input patterns until the weight adjustment converges and no further adjustments are made. Once convergence occurs, the net moves from training mode into production. The presentation of new data patterns, different from those seen on training, will require retraining of the network. Retraining must be done using old as well as new patterns so that the net will learn the new patterns without forgetting the old.

The problem in this situation is that we know that although most criminals are creatures of habit, new patterns occur in the data. New criminals will be out on the street. Some criminals may move or alter their patterns in certain ways. Some criminals may be removed from the population, thus causing cessation of certain patterns. We do not know how long patterns stay in effect. Thus we do not know how often to attempt to retrain the system. There is also no scheme for which patterns might be useful to keep in the training set.

Data Processing

FINDING: We will have to conduct a more extensive examination of the data before proceeding with development of neural nets using alternative architectures.

We must ensure that data contains no inconsistencies. That is, there should be no errors in the training sets. A consistency can occur when a number of similar input patterns are associated with different output patterns. This problem was mentioned above. It is not wise to eliminate any inconsistent patterns. We believe that there is sufficient data to differentiate by location and time so that we can handle this problem.

ALTERNATIVE NEURAL NET ARCHITECTURES FOR FUTURE RESEARCH

Adaptive Resonance Theory

One Neural Net based solution to this problem would be the use of Adaptive Resonance Theory or ART networks. ART type networks are more complicated than simple Back Propagation networks. They are dynamic and do not separate learning from production modes. These nets do not require knowledge of a precise number of classes in training data. The ART net will examine an input pattern and attempt to match it to known patterns. If the net has seen the input before, it will strengthen that pattern on recall. If the pattern is new, then the ART net adds a new layer to encode new information about this pattern and link the new layer to existing layers. Essentially the ART net uses both feed forward and feed backward to adjust its weights as data are analyzed. This establishes an arbitrary number of categories. Networks of this type are used in vision applications.

This type of network may be the subject of future research in this area. Although it would be a possible solution to the problem of recognition of new patterns, the problem of conflicts (as discussed above) would still remain.

Probabalistic Neural Nets

These nets are a special case of feed forward neural networks. The node architecture is multilayer and fully connected. There is one hidden layer. Neuron activation uses a Gaussian function. Supervised training may be done, in some cases, in a single pass using input sets as exemplars. Advantages to this type of net are the functions used, which are universal approximators utilizing local maxima. The nets also train quickly because back propagation, which is time consuming, is not used. Disadvantages revolve around poor performance with sparse or noisy data. The type of data present in the existing Police Department data base would probably lead to errors in classification.

IV. SYSTEM DEVELOPMENT CHAPTER

INTRODUCTION

This chapter explains the method used to implement the system. We begin by explaining the initial pattern recognition approach using the neural network. This initial approach did not adequately categorize case data. A modified nearest neighbor approach with decision trees was then used. We describe that method. Finally we describe the operation of the system interface.

INITIAL APPROACH USING NEURAL NETS

The initial methodology used in this project involved the building of a system which used the most commonly applied artificial intelligence programming technologies to complement each other. Rule based expert systems are the most commonly fielded type of artificial intelligence systems, closely followed by neural networks. Each has strengths and weaknesses. When used together the systems complement each other well.

Rule based expert systems are good at implementing expert reasoning ability by encoding heuristics or rules for problem solving. This procedural knowledge is available for use by the system and is available to the system user in the form of explanations. A neural net excels at pattern recognition tasks.

The trend in AI systems development methodology is to use a number of tools in combination (Newquist, 1990). This allows the synthesized system to combine the best aspects and outcomes of logic and reasoning (expert systems) with pattern matching, recognition, and pattern analysis (neural nets).

The project described herein was initially designed to use a neural net as a pattern recognizer. The net was to be designed and trained to examine department wide data, initially for robbery reports, and to categorize the individual cases reported into appropriate groups or patterns. It was hoped that the net would be able to learn how to recognize patterns in the data that would indicate certain groups of case reports were related. The neural net, as the critical component of the system developed, was to be used by the crime analyst in examining monthly departmental crime data. The system within which this net was embedded would examine monthly data and prepare lists of grouped cases. The analyst would then be able to use a rule based component to assist in preparation, routing, coordination of cases across areas or districts, or to assist in case assignment. Heuristic rules could also be used prior to neural net categorization to prepare raw data.

A number of factors led to the decision to focus on the pattern recognition component of the system and to "hide" the heuristic knowledge or rules in the system code.

1) The Information Technology infrastructure of the

Department does not currently support the use of the system by individual detectives. Equipment is needed. The Department data base would need to be reorganized and a network capability designed to allow crime unit analysts to access data, case reports, crime analysis applications, and areas, districts and units.

2) Data entry must be reorganized. The case report forms do not capture all the necessary data. Patrol officers sometimes do not complete forms properly. Data is frequently entered incorrectly. In addition, narrative data is not captured in a way in which it can be easily coded automatically.

3) The entire data entry, case assignment procedure would need to be reengineered to adequately address items 1) and 2) above.

4) As has been stated, the detectives often use the MO file, in conjunction with the case report, in a reactive way. This method of decision making has an effect on the placement of the technology components in the overall system, and the way in which the system will be used. Crime analysts will examine the data in ways that are somewhat different from those of detectives. The crime analysts will operate at a more strategic or "big picture" level than will the detectives.

NEURAL NET DEVELOPMENT METHODOLOGY

The development of the backpropagation neural net for this project followed the general principles described in the above sections. The neural net developer performed the following tasks:

1)Meta-Analysis

The neural net developer attended one of the focus group meetings with the project team. He participated in post group discussions with project team members. This session and notes and transcripts of the other focus group sessions gave the neural net developer an overview of the types of data elements captured in investigative reports, their use and usefulness in the development of cases.

2)Data Analysis

The neural net developer obtained complete case report data on robbery cases for a one year period. This consisted of a download of the pertinent data elements. Data was obtained from the CPD RAMIS system. Case report data elements are contained in Appendix A. Narrative sections of the case reports were not used because this data is not coded by CPD in its data base.

The data was subjected to an initial statistical analysis and pareto analysis. A number of data elements were frequently left blank or contained data values indicating that the field value was unknown. These issues are addressed in greater detail in the section on Operational Findings and Recommendations.

The results of focus group observations and the data analysis indicated that 10 of the data elements could possibly be used as input variables to sufficiently differentiate crime patterns.

3)Neural Net Configuration

A Back Propagation neural network was constructed for use in the pattern categorization. This net used 3 layers configured as follows:Input layer containing 10 input variables

- Time of occurrence
- Beat
- Location Type code
- Offender Race
- Offender Age
- Offender Height
- Offender Weight
- Offender Eye Color
- Offender Hair Color
- Weapon Used code

The variables in the input layer are of two kinds:

a) Scalar variables (scaled, continuously variable values in an arbitrary range.) These include time, age, height, weight. The range of values represented in the data must be compressed (usually to the range 0 through 1.) Data in the entire training set must be used to set these normalized values.

b) Categorical variables (representing data having one of n possible arbitrary categorical values.) These include the rest of the above referenced input variables. Frequently, multiple binary inputs may be used. Each may contain as many binary positions in the pattern vector as there are feature values.

One Internal layer containing 5 nodes

The number of hidden layers used in such a net is normally one. The number of nodes implemented in that layer may generally be from 20% to 50% of the number of input nodes to that layer.

Output layer containing one node

Pattern Classification results in the categorization of the test cases, or new cases being examined, into one of a number of possible crime pattern categories. Unfortunately, the categories ultimately used were only those eight learned by the net during training on sets of known pattern data. The net was not able to recognize and categorize cases into new patterns. This problem will be described in detail in the discussion following. It is the main reason why the system was implemented using the nearest neighbors pattern classification method adopted.

4) Training

A selected set of training exemplars was obtained from the Chicago Police Department. These consisted of closed cases which were known to fit certain patterns of crimes committed by certain serial offenders. The training proceeded until the net converged and weights were frozen.

Hebb's Learning Rule, one of the most common, was used by the net. Simply put, the rule holds that when any 2 nodes are simultaneously excited, the connection strength between them is increased. The converse also applies. A connection's strength can be adjusted downward by inhibitory activity.

5) Testing

The neural net developer tested the net against unknown cases for categorization.

6) Proof of Concept Net

This initial net was refined to create a proof of concept net. Refinement proceeded through two iterations of development including training with additional exemplars.

PATTERN RECOGNITION USING MODIFIED NEAREST NEIGHBOR METHODS

Problems in the training and performance of the Neural Network led to a search for another method of classifying patterns of criminal activity. The work done by Pattern

Associates, Inc. satisfied this need. In comparisons of the classification capabilities of Neural Nets versus nearest neighbors methods and k-clustering methods, the nearest neighbors methods generally proved more accurate, especially when the cluster size increased. (Balakrishnan et al, 1994) The method used involved the following phases (Frey, 1996).

1) Feature Fusion

Raw attribute vectors were reduced to a set of dimensions which comprised the decision space. This transformation will reduce a large input vector (30 or more data elements) to a small decision space of 2 to 5 dimensions. Input values from a group of features can be used to create one output which represents a single dimension of the decision space. In this project only simple feature fusion was done basically as data preprocessing. Future research should attempt the complete representation of an appropriate decision space for this domain.

A standard approach to achieve this mapping uses regression equations built for each dimension. The equation variables are sets of features. This approach is unacceptable in our domain for a number of reasons. First, much of the data in the data sets used is categorical. Many numerical values have uneven distributions of values and much noisy or missing data. These problems make regression an unlikely tool as regression equations usually "like" well-behaved numerical data.

2) Neighbor Identification

Nearest neighbor analysis is done (Stanfill and Waltz, 1986.) Any case is compared to prior cases stored in memory. The Euclidean distance is calculated between the position in k-space of the target case and stored prior cases. Nearest neighbors are grouped by choosing a given number of them that are closest.

3) Examine Neighbors

Outcomes for each neighbor are examined. For numerical values the values of neighbors are averaged to determine the outcome for the test case. For binary valued neighbors, proportion is used to determine the outcome. For categorical outcomes, that with the highest proportion or frequency among the neighbors is chosen. Weighted voting may be used if weights are assigned to each member of the subset and factored into the equation.

BINARY CLASSIFICATION TREES

Our method of segmenting or categorizing data involves the creating of binary splits in the cases, thus identifying category patterns (Frey, 1996). The database is arranged as a flat file. Predictor variables are selected and preprocessed if necessary.

All variables and values are assessed for their strength in defining the points at which categorization splits are applied. Variables are ~~looked at~~ examined in order of importance. Importance is indicated by the assignment of priorities. The splitting creates an inverted decision tree. Successive splits become less predictive. The process continues until divisions are not meaningful or there are two few remaining cases to split out. This process can use all types of data. In addition, missing data is not a problem.

A Binary Classification Tree is generated for the data set. The advantages of the approach taken actually increase with the size of the data set.

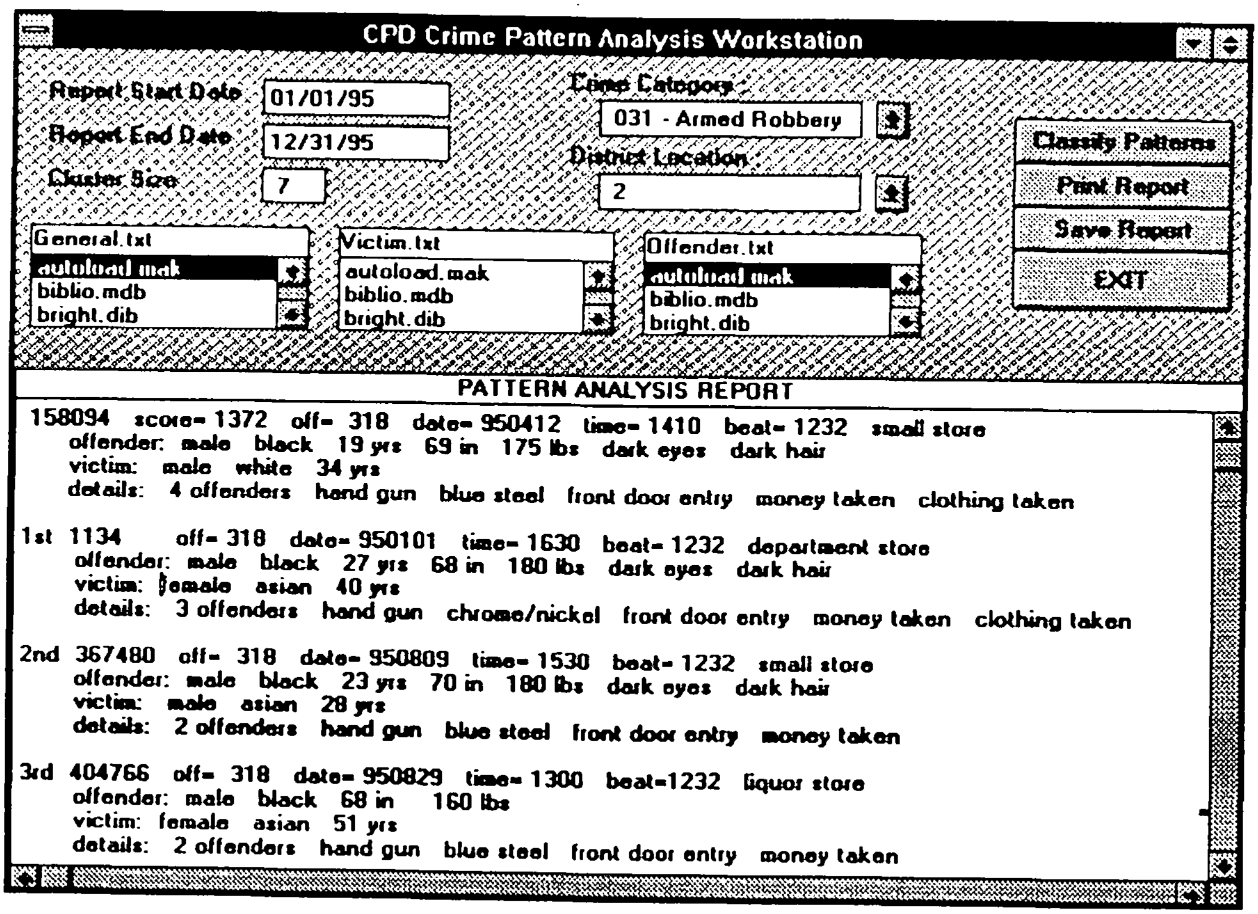
This method has decided advantages. One among them is that because the decisions rely on localized information (with regard to decision space), the use of a small database results in random guesses as output. This matches what is known of human novice behavior with sparse data. As the number of known cases grows, the forecasts should gain in accuracy. Regression and other global optimization methods improve rapidly at first and then become asymptotic. Humans and neighborhood voting schemes appear to improve continually with experience. In essence, standard regression treats data relationships as additive, creating a weighted linear equation. Regression models do not handle higher order data interactions, non-linear relationships, or non-gaussian distributions that well. Regression will also throw out

STRONG STATEMENTS!

outliers in the data set.

SYSTEM INTERFACE

The interface represented in Figure 2 allows the user to supply certain information that can be used to limit the scope of case reports being analyzed.



Items include:

- Starting date for the report period analysis (mm/dd/yy format)

- End date for the report period analysis (mm/dd/yy format)
- Size of the cluster (number of nearest neighbors to find for each seed case)
- Criminal category to examine from the following list of Departmental category codes:
    - 01 Murder
    - 02 Criminal Sexual Assault
    - 03 Robbery
    - 04 Battery
    - 05 Assault
    - 06 Burglary
    - 08 Theft
    - 10 Arson/Explosives
    - 11 Fraud/Embezzlement
    - 12 Stolen Property Crimes
    - 13 Property Damage
    - 14 Weapons Violations
- Location (limited now to departmental location/jurisdiction)
- Data files to be used in the analysis (contained in a fixed working directory):
    - Name of file containing general case report information
    - Name of file containing victim information
    - Name of file containing offender information

The user may then use command buttons to perform the following actions:

- Extract and clean up data for processing (call to data preparation program)
- Analyze the cleaned data (call to nearest neighbors heuristic program) This action also results in a display of the analysis report in the Report Display frame on the bottom half of the screen.
- Print the displayed report
- Save the displayed report to a file
- Exit the program

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 files used (some in excess of 70 megabytes.)

These include technical and operational needs/issues:

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

The operating system and basic architecture of personal computers is still 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 the 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 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 to Information systems personnel. But no matter how the IS Department would view the full implementation of analysis systems (which should theoretically be available for use by any investigative personnel), budgetary limitations would prohibit hiring of the substantial numbers of personnel needed.

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 the 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 are moving the delivery platform of the system to an environment which will support the needs enumerated above. This involves the use of the WorldWideWeb(WWW or Web) as the network or 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 layer will be graphic, and will be built with tools that will allow great flexibility in choice of hardware platforms. The most innovative aspect of the user interface is that it will run in a state-maintaining Web browser. In addition, the interface will run 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.)

Connectivity between the layers, and with the rest of the world, will be built on the Internet and the WorldWideWeb (WWW or Web). Data bases, applications, and decision support tools can be located on net nodes accessible via Internet connection.

Front ends to the system resources will be built using scripted windows which will be provided by the server as needed by clients. This method allows the transmission of small script files which can be run interactively by the client. These scripted interfaces can be run by a simple public domain interpreter. By using this method of input combined with a web browser which allows for interactive real-time use of applications over the web, end users will have full use of all remote systems. Regardless of the physical method of connecting users to remote machines, the protocols for such connections have generally been difficult to initiate and support. The recent development of browsers for the Web has dramatically decreased the effort, and increased the robustness, of client-server connections.

The initial methods of providing client-server connections over the Web have used a stateless mode; forms, consisting of Hypertext Markup Language (HTML) documents are sent to the desired server using the Universal Resource Locator (URL) addressing method. The forms are received by a Common Gateway Interface (CGI) program on the server, and are parsed and passed to a database. Any action by the server must generate a new HTML document, which is passed back to the client. The system is stateless since the connection between the client and the server is not maintained. While useful, the CGI approach is not optimal. Database interactions are much more useful if connectivity (state) is maintained. This allows the use of interactive objects such as buttons, pick lists, etc., and allows testing for individual responses

The popularity of the Web has been sparked by technological advances to basic Web technology. These advancements have transformed the Web into an environment for robust interactive applications. Such interactive environments are called "document-centric" since the documents themselves become the focus for doing work. The Web document becomes the environment for doing work, through embedded applets which can run interactively and pull data from any necessary Web location which they are authorized to visit.

The advantage to using the newly developing methods and environments is that they allow for use of a single integrated environment for rapid Web application development. The application development tools themselves are actually downloaded Weblets, which means that any machine capable of accessing the Web can be used as a development workstation. The environment is designed specifically for component based application development, allowing reusable modules to be developed, used, and stored from within a server-side "component database." By using a scripting language such as Tcl (developed and supported by University of California, Berkeley and now used by SUN) Weblet applications can be generated on the fly, combining only the application components needed by the user into the downloaded weblet program.

Security for both the browser user and the server can be handled through Public Key Encryption technology. This allows authentication, access control, validation of code, and data encryption to be controlled through an Industry standard and universally entrusted method. Weblets can communicate through messaging technology with other Internet applications using the industry standard socket protocols. Distributed processing is supported through remote procedure call (RPC) technology. This allows flexible partitioning of program execution by, for example, allowing user interface, numerical computation engines, and data and data access components of an application to reside and run on optimal machines on the network.

Dynamic program capabilities combined with the distributive processing feature allow for the development of "performance adaptive" applications which can decide at time of use where various program parts should run. Multiple versions of an application don't have to be created. This also allows for system recovery and reconfiguration in the event that parts of the net are disabled.

The environment extends from the Weblet running in the Web page, back through levels of server applications, to legacy data. One set of tools and one language can be used to create all components of the enterprise system. Pieces of the system which have been developed in other languages or with other tools can be easily integrated within the environment. This stands in contrast to other Internet development tools, which require re-engineering of existing applications for integration. The use of this technology allows the leveraging of the Web as a low cost distribution channel, as well as a cooperative interactive working environment.

V. RECOMMENDATIONS

One of the most important findings in this study revolves around the fuzziness and uncertainty of the data available. Because much of the data is missing, inaccurately described (by victims or witnesses), inaccurately transcribed onto reports by patrol officers, inaccurately entered into the data base system, or fuzzy (height of 72 -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. In addition, weather information could be used.)

2) Fuzziness in data must be addressed.

a) 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." "He was not very old; maybe 20 to 25."

b) In many cases the data is described in an unequivocal crisp manner (yes or no answers; or a selection from an enumerated list of possible answers.) It is often the case that we have reason to doubt the correctness of these answers. (e.g. How accurate is the description of offenders who are seated in a vehicle at midnight?)

3) Preprocessing of data must be expanded. The data should be examined and manipulated to clean up errors and noise. The feasance should also be mapped using fuzzy set theoretic functions. Data elements should also 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. Here, 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.

4) The system must be fully implemented and tested. The system must be implemented and fully tested in a "production" environment. The system is being installed for use by a case officer in an Area detective office. It will be used in examining data for case assignments. It's use can then be monitored and examined more fully as time permits.

5) Other Neural Net learning algorithms should be tested. As has been explained in previous sections of this report, the particular Neural Net algorithms tested in this project were not amenable to the type of data used. A further test of Neural Nets should include the testing of unsupervised learning algorithms. Of potential use would be the ART or Kohonen type nets.

Appendix A
Files, Codes and Categories

Tables (AA), (AB) and (AC) show the fields included in the victim table, the offender table and the crime table respectively. Table (AD) shows some of the codes that are used in these tables. Table (AE), on the other hand, shows fields and/or categorizations that are part of the crime report and may be part of crimes tables.

Field	Starting Position	Field Length	Field Type	Remarks
RDNum	1	7	Alpha	RDNum from police case report
Sex	8	1	Alpha	Sex of victim (Male/Female)
Race	9	1	Numeric	Race of victim
Age	10	3	Numeric	Age of victim
Injury	13	1	Alpha	Injured (Yes/No)

Table (AA) Fields in the Victim File

Field	Starting Position	Field Length	Field Type	Remarks
RDNum	1	7	Alpha	RDNum from police case report
Sex	8	1	Alpha	Sex of offender (Male/Female)
Race	9	1	Numeric	Race of offender
Age	10	3	Numeric	Age of offender
Height	13	3	Numeric	Height of offender
Weight	16	3	Numeric	Weight of offender
Eyes	19	3	Alpha	Color of eyes
Hair	22	3	Alpha	Color of hair
Complexion	25	8	Alpha	Complexion (Med., Light ... etc.)
VehYr	33	2	Numeric	Make year of vehicle
VehMake	35	4	Alpha	Brand name of vehicle
Color	39	6	Alpha	Color of vehicle
Body	45	2	Alpha	Body of vehicle (2 DR ... etc.)
License	47	8	Alpha	License plate number
State	55	2	Alpha	State of license plate

Table (AB) Fields in the Offender File

Field	Starting Position	Field Length	Field Type	Remarks
RDNum	1	7	Alpha	RDNum from police case report
OffenseCode	8	4	Alpha	Type of crime (code)
Date	12	6	Numeric	Crime date
Time	18	4	Numeric	Crime time (24 hr. clock)
Location	22	3	Alpha	Type of location (alley, park ... etc.)
Beat	25	4	Alpha	Beat number from case report
StNum	29	5	Alpha	Street numbers
StDirection	34	1	Alpha	Street direction (N, S, E, W)
StName	35	24	Alpha	Name of street
NumVictims	59	3	Numeric	Number of victims
NumOffenders	62	3	Numeric	Number of offenders
WeapCode	65	1	Numeric	Weapon code
WeapTpe	66	2	Numeric	Weapon type
Feature	68	2	Numeric	Features in the weapon
EntryPoint	70	2	Numeric	Entry point code
ExitPoint	72	2	Numeric	Exit point code
Alarm	74	2	Numeric	Alarm on premise? Circumvented?
SafeMethod	76	2	Numeric	Safe burglary method
Residence	78	2	Numeric	If residence, where were occupants?
TakenCode1	80	1	Numeric	The code of items .....
TakenCode2	81	1	Numeric	that were taken .....
TakenCode3	82	1	Numeric	by the offender .....
TakenCode4	83	1	Numeric	in the crime incident.
TakenCode5	84	1	Numeric	"same"
TakenCode6	85	1	Numeric	"same"

Table (AC) Fields in the General Crime File

Field	Code	Description
Race	1	Black
	2	White
	3	Black-Hispanic
	4	White-Hispanic
	5	American Indian / Alaskan Native
	6	Asian / Pacific Islander
Eyes	1	Brown
	2	Blue
	3	Green
	4	Unknown

Table (AD) Codes used in the Tables

Field	Code	Description
Hair	1	Black
	2	Light Brown
	3	Dark Brown
	4	Blonde
	5	Red
	6	Gray
	7	Bold
Complexion	1	Light
	2	Medium
	3	Dark
	4	Olive
	5	Unknown
TakenCode	1	Money
	2	Jewelry
	3	Furs
	4	Clothing
	5	Office Equipment
	6	TV, Radio, Stereo
	7	Household Goods
	8	Consumable Goods
	9	Firearms
	10	Narcotics / Dangerous Drugs
	11	Vehicles
	12	Others
	13	None

WeapCode	1	Used
	2	Displayed
	3	Unknown
Feature	1	Chrome Nickel
	2	Blue Steel
	3	Short Barrel
	4	Long Barrel
	5	Sawed Off
	6	Other
	7	Unknown
	8	DNA (Does Not Apply)

Table (AD) Codes ... Continued

Field	Code	Description
WeapType	1	Handgun
	2	Shotgun
	3	Rifle
	4	Knife
	5	Vehicle
	6	Blunt Instrument
	7	Thrown Object
	8	Explosive
	9	Liquid / Gas
	10	Bottle / Glass
	11	Razor
	12	Pry Tool
	13	Hands, Feet
	14	Other
	15	DNA (Does Not Apply)
EntryPoint	1	Front Door
ExitPoint	2	Rear Door
	3	Window
	4	Roof
	5	Floor
	6	Side Door
	7	Other
	8	Unknown
	9	DNA (Does Not Apply)
Alarm	1	DNA (Does Not Apply)

	2	On Premise (Y/N)
	3	Alarm Circumvented (Y/N)
SafeMethod	1	Punch
	2	Force
	3	Explosive
	4	Drill
	5	Removed
	6	Peel
	7	Open
	8	Unknown
	9	DNA (Does Not Apply)

Table (AD) Codes ... Continued

Field	**Code**	**Description**
Residence	1	Work
	2	Visiting
	3	Vacation
	4	Wedding
	5	Funeral / Wake
	6	Other
	7	Unknown
	8	DNA (Does Not Apply)

Table (AD) Codes ... Continued

Field	Code	Description
UCRCode UCR code used to represent armed or strong-armed robberies	031A	Armed robbery - handgun
	031B	Armed robbery - other firearms
	0312	Armed robbery - knife
	0313	Armed robbery - other dangerous weapon
	0320	Strong-Arm
	033A	Attempted armed robbery - handgun
	033B	Attempted armed robbery - other firearm
	0334	Attempted armed robbery - knife
	0337	Attempted armed robbery - other dangerous weapon
	0340	Attempted Strong-Arm
Facial-Hair	1	Clean Shaven
	2	Mustache
	3	Beard
	4	Unknown
Marks/Scars	1	Tattoo
	2	Scar
	3	Birth Mark
	4	Unknown
Color Color of vehicle	1	Black
	2	White
	3	Gray / Silver
	4	Blue
	5	Red
	6	Green
	7	Brown
	8	Yellow
	9	Unknown / DNA (Does Not Apply)

Table (AE) Other Fields and/or Categorizations

Field	Code	Description
VehMake	1	Chevy
	2	Buick
	3	Nissan
	4	Ford
	5	Cadillac
	6	Mazda
	7	Audi
	8	Oldsmobile
	9	Mercury
	10	Mercedes
	11	BMW
	12	Pontiac
	13	Jeep
	14	Unknown / DNA (Does Not Apply)
	15	Hyundai
	16	Accura
	17	Plymouth
	18	Suzuki
	19	Honda
	20	Lincoln
	21	Toyota
	22	GMC
	23	Dodge
	24	Infiniti
	25	GEO

Table (AE) Other Fields/Categorizations (Continued)

LITERATURE CITED

Aldenderfer, M. and R. Blashfield (1984). **Cluster Analysis.** Newbury Park, CA.: Sage Publications.

Automation Review in *Best's Review: Life/Health Insurance,* **89**(5):136-137 (1988).

Balakrishnan, P.V., Cooper, M.C., Jacob, V.S. and Lewis, P. (1994). *A study of the classification capabilities of Neural Networks using unsupervised learning : A comparison with K-means clustering.* **Psychometrika,** 59(4):509-525.

Block, R. and Dabdoub, M. (1993). **Workshop on Crime Analysis Through Computer Mapping Proceedings.** Chicago, Illinois: Illinois Criminal Justice Information Authority.

Blumstein, A. et al. (1986). **Criminal Careers and Career Criminals. Vol. I and II.** Washington, D.C.: National Academy Press.

Buchanan, B.G. and Shortliffe, E.H. (1984). **Rule-Based Expert Programs.** Reading, MA: Addison-Wesley.

Chaiken, J. and Chaiken, M. (1982). **Varieties of Criminal Behavior.** Prepared for the National Institute of Justice, U.S. Department of Justice, Report R-2814/1-1015. Santa Monica, CA: Rand Corp.

Chester, M. (1993). **Neural Networks: A Tutorial.** Englewood Cliffs, N.J.: Prentice-Hall.

Clifford, H. and W. Stephenson (1975). **An Introduction to Numerical Taxonomy.** New York: Academic Press.

Denning, P.J. (1986). *Towards a science of expert systems,* **IEEE Expert,** 1, 2: 80-83.

Eck, J.E. (1979). **Managing Case Assignments: The Burglary Investigation Decision Model Replication.** Washington DC: Police Executive Research Forum.

Eck, J.E. (1983). **Solving Crimes: The Investigation of Burglary and Robbery.** Washington DC: Police Executive Research Forum.

Eck, J.E. (1992). *Criminal Investigation.* **In What Works in Policing?** Ed. by G. Cordner & D. Hale. Cincinnati, OH: Anderson.

Everitt, B. (1980). **Cluster Analysis.** New York: Halsted.

Freeman, J.A. (1994). **Simulating Neural Networks with Mathematica.** Reading, MA: Addison-Wesley.

Frenzel, C.W. (1992). **Managing Information Technology.** Boston, MA: Boyd & Fraser Publishing Co.

Frey, Peter (1996). **Conceptual Overview of Neighborhood Voting System.** Tech. Report. Pattern Associates, Inc.

Fryer, B. (1993). **Chemical Bank's Weapon for Risk Management, PC AI,** Jan.-Feb.: 26-29.

Gaines, L.K., B. Lewis & R. Swanagin. (1983). *Case Screening in Criminal Investigations.* **Police Studies,** 6: 22-29.

Greenwood, P. (1970). **An Analysis of the Apprehension Activities of the New York City Police Department.** N.Y.: Rand Institute.

Greenwood, P. (1989). **Selective Incapacitation.** Report prepared for the National Institute of Justice, U.S. Department of Justice. Report R-2637-DOJ. Santa Monica, CA: Rand Corp.

Greenwood, P., J. Chaiken & J. Petersilia. (1977). **The Investigative Process.** Lexington, MA: Lexington Books.

Hink, R.F. and Woods, D.L. (1987). *How humans process uncertain knowledge: An introduction for knowledge engineers,* **AI Magazine, 8,** 3: 41-53.

Isaacs, H. (1967). *A Study of Communications, Crimes, and Arrests in a Metropolitan Police Department,* in **Institute for Defense Analysis: Task Force Report: Science and Technology.** Report to the President's Commission on Law Enforcement and Administration of Justice: 88-106.

Kahneman, D. and Tversky, A. (1972). *Subject probability: A judgment of representativeness,* **Cognitive Psychology, 3**:430-454.

Kahneman, D. and Tversky, A. (1982). *On the study of statistical intuitions*, **Cognition, 11**, 2: 123-141.

Kahneman, D. Slovic, P. and Tversky, A. (1982). **Judgment Under Uncertainty: Heuristics And Biases.** New York: Cambridge Univ. Press.

Kahneman, D. and Tversky, A. (1973). *On the psychology of prediction*, **Biological Review, 80**: 237-251.

Keen, P.W. and Scott-Morton, M.S. (1987). **Decision Support Systems: An Organizational Perspective.** Reading, MA: Addison-Wesley.

Leinweber, D. (1988). *Knowledge-based systems for financial application*, **IEEE Expert, 3**(3):18-31.

Liebowitz, J. and DeSalvo, D.A. (eds.) (1989). **Structuring Expert Systems: Domain, Design, and Development.** Englewood Cliffs, N.J.: Yourdon Press, Prentice-Hall.

Maltz, M. (1991). **Mapping Crime in Its Community Setting: Event Geography Analysis.** N.Y, N.Y: Springer-Verlag.

McGowan, D. (1991). *High Tech Sleuthing*, **Computers in Healthcare**, Dec.: 33-34,39.

Mena, J. (1994). **The Adaptive Tax Collector**, *AI Expert*: 39-41.

Mettrey, W. (1987). *An assessment of tools for building large knowledge-based systems*, **AI Magazine, 8**, 4: 81-92.

Newell, A. and Simon, H.A. (1972). **Human Problem Solving.** Englewood Cliffs, N.J.: Prentice-Hall.

Newquist, H.P. (1990). *Bloodhounds and Expert Systems.* **AI Expert** Mar.: 66-69.

Newquist, H.P. (1990). *AI Trends*, **PC AI**, March-April: 24-30.

Nilsson, N.J. (1980). **Principles Of Artificial Intelligence.** Palo Alto, CA: Tioga Publishing Co.

Pao, Yoh-Han (1989). **Adaptive Pattern Recognition and Neural Networks.** Reading, MA.: Addison-Wesley.

Petersilia, J. and Honig, P. (1980). **The Prison Experience of Career Criminals.** Prepared for the National Institute of Justice, U.S. Department of Justice, Report R-2511-DOJ. Santa Monica, CA: Rand Corp.

Rolph, J., Chaiken, J. and Houchens, R. (1981). **Methods for Estimating Crime Rates of Individuals.** Prepared for the National Institute of Justice, U.S. Department of Justice, Report R-2730-NIJ. Santa Monica, CA: Rand Corp.

Romesburg, H.C. (1984). **Cluster Analysis For Researchers.** Belmont, Ca.: Lifetime Learning Publications.

Rosenbaum, D. (1994). **The Challenge of Community Policing: Testing the Promises.** Thousand Oaks, CA.: Sage Pub.

Rothi, J.A. and Yen, D.C. (1990). **Why American Express Gambled on an Expert Data Base,** *Information Strategy,* Spring: 16-22.

Rumelhart, D.E. and McClelland, J.L. (1986). **Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Volumes 1 and 2.** Cambridge, MA: Bradford Books, MIT Press.

Rummelhart, D.E. and McClelland, J.L. (1986). **Parallel Distributed Processing: Explorations in Microstructures of Cognition, Volumes 1 and 2.** Cambridge, MA: MIT Press.

Salzberg, S. (1988). **Machine Learning Moves Out of the Lab,** in *AI Expert,* Feb.: 44-52.

Schacter, R.D. and Heckerman, D.E. (1987). *Thinking Backward for Knowledge Acquisition,* **AI Magazine** 3: 55-61.

Schacter, R.D. and Heckerman,D.E. (1987). *Thinking backward for knowledge acquisition,* **AI Magazine,** 8, 3: 55-61.

Sejnowski, T.J. and Rosenberg, C.R. (1987). *Parallel networks that learn to pronounce English text.* **Complex Systems, (1)**: 145-168.

Shpilberg, D. (1987). *One giant step for insurers, Best's Review: Property/Casualty Insurance,* **87**(1): 54-60.

Skapura, D. M. (1995). **Building Neural Networks.** Reading, MA: Addison Wesley.

Sneath, P. and R. Sokal (1973). **Numerical Taxonomy.** San Francisco, CA: W.H.Freeman.

Snyder, C. (1987). *From research to reality: The leading edge of expert Systems,* **Insurance Software Review, 12**(3): 22-24,26-27,30.

Stanfill, G., and Waltz, D. (1986). *Toward memory-based reasoning.* **Comm. of the ACM, 29**: 1213-1228.

Tracy, P., Wolfgang, M. and Figlio, R. (1990). **Delinquency Careers in Two Birth Cohorts.** N.Y.: Plenum Press.

Tschudi, F. (1988). *Matrix Representation of Expert Systems* in **AI Expert,** Oct.: 44-53.

Waldo, G. (1989). **Career Criminals.** Beverly Hills, Ca.: Sage Pub.

Waterman, D.A. (1986). **A Guide to Expert Systems.** Reading, MA: Addison-Wesley Publishing Co.

Weiner, N. and Wolfgang, M. (1989). **Violent Crimes, Violent Criminals.** Newbury Pk., CA.: Sage Pub.

Williams, W. (1971). *Principles of Clustering.* **Annual Review of Ecology and Systematics** 2: 303-326.

Zurada, J. (1992). **Artificial Neural Systems.** St. Paul, MN: West Publishing Co.