Artificial Intelligence and Law **11:** 251–269, 2003.

Classification system for serial criminal patterns

KAMAL DAHBUR and THOMAS MUSCARELLO
DePaul University, USA
E-mail: dahbur@cs.depaul.edu; muscarello@cs.depaul.edu

Abstract. The data mining field in computer science specializes in extracting implicit information that is distributed across the stored data records and/or exists as associations among groups of records. Criminal databases contain information on the crimes themselves, the offenders, the victims as well as the vehicles that were involved in the crime. Among these records lie groups of crimes that can be attributed to serial criminals who are responsible for multiple criminal offenses and usually exhibit patterns in their operations, by specializing in a particular crime category (i.e., rape, murder, robbery, etc.), and applying a specific method for implementing their crimes. Discovering serial criminal patterns in crime databases is, in general, a clustering activity in the area of data mining that is concerned with detecting trends in the data by classifying and grouping similar records. In this paper, we report on the different statistical and neural network approaches to the clustering problem in data mining in general, and as it applies to our crime domain in particular. We discuss our approach of using a cascaded network of Kohonen neural networks followed by heuristic processing of the networks outputs that best simulated the experts in the field. We address the issues in this project and the reasoning behind this approach, including: the choice of neural networks, in general, over statistical algorithms as the main tool, and the use of Kohonen networks in particular, the choice for the cascaded approach instead of the direct approach, and the choice of a heuristics subsystem as a back-end subsystem to the neural networks. We also report on the advantages of this approach over both the traditional approach of using a single neural network to accommodate all the attributes, and that of applying a single clustering algorithm on all the data attributes.

Key words: data mining, neural networks, pattern classification

1. Introduction

A large police department usually has huge databases comprised of hundreds of thousands of records, each containing data about a criminal offense (crime). The information on each crime may contain patterns in the features related to the general crime's data, the offender(s) data, the victim(s) data and/or the vehicle's data (Dunn 1976; O'Shea et al. 1995). The patterns include and are not limited to: the characteristics of the victims, the type of weapon(s) used in the crimes, the physical characteristics of the offender(s), the geographical area of the crime incident, the particularities of the location of the crime, the characteristics of the get-away vehicle used during the crime, the items taken by the offender(s), any particular clothing items worn by the offender during the crime, etc.

We propose an automated methodology that can systematically identify groups of records as potential patterns for serial criminals, with a good degree of accuracy. Neural networks will be the main tool for the classification of patterns because of their powerful capabilities at such tasks, as will be discussed in detail later. The research will also investigate a methodology to process the data into the form needed for the neural networks to operate on. Heuristics are used to refine the outputs of the neural network. Each one of these issues will be discussed in this paper.

This is an application research project in the area of information systems and artificial intelligence that can be considered to be of empirical value rather than a purely theoretical innovation. We are applying neural networks to an application area that is untried and which has its social and commercial significance. This will be the first time that a hybrid system using neural networks and a rule based heuristics system will be implemented in this application domain. It is also the first time, to our knowledge, that a Kohonen network will be used to build a clustering system for recognizing and grouping potential serial crimes. Our approach to the data analysis and pre-processing of data have proven to be fruitful and will help, from an information systems point of view, to make recommendations to criminal agencies on what information needs to be kept on criminals and what type of format is best for keeping such information. Our categorization and re-categorization will serve as guidelines for criminal experts on what constitutes important information in pattern discovery of serial criminals.

Our approach has also some theoretical contribution, as it divides the complete set of attributes into four groups, which makes the use of the hybrid network of neural networks in this application domain unique in its design. This design enhances the overall classification task, because it simulates, to an extent, the thinking of expert investigators in the field. In addition, this design differentiates between the values of each set of attributes and enforces these values and their corresponding weighted classifications into the overall classifications of the system. The modular approach of this design enables the user of this system to have better control of the system, and allows for multiple checkpoints and validation processes. This helps to overcome the usual problem which neural networks present with respect to explanations-that is, an inability to explain or make clear what was done by the "black box" process. This design caters for individual personal judgment and qualitative analysis of the intermediate and final results, to help achieve the results that are best suited for individual expertise. The design offers a new approach to data mining instead of feeding all the attributes into one single Kohonen network (Kohonen 1988; Xia 1996).

The use of a heuristics system at the final step has its empirical as well as commercial contribution, as we search for a set of heuristics (rules) that will simulate the thinking of an expert in the field and enhance the final classifications. The heuristics system will help overcome any shortcomings in the pre-processing phase as well as the neural network phase. In addition, the heuristics system will allow individual experts in the field to incorporate their own general or special heuristics into the final classifications.

2. Related research

There is a remarkable diffusion of community policing activity in the law enforcement world. There is no consistent definition of the term community policing. The practice has taken on many variations based on the agency, jurisdiction, geographic areas covered, and policing culture. This approach is commonly used, while some of its underlying principles are not so well understood. O'Shea et al. (1995) undertook their original project, conducted under a grant from the National Institute of Justice, seeking to focus on one such principle: the exploration and development of methods to structure innovative technology-based responses to facilitate criminal investigations. That project focused on a deficiency in current police practices, i.e., problem identification, the initial stage of the community policing model. The success of the ability of the police to identify and apprehend criminals is grounded in the capacity of the police agency to accurately analyze data and transform it into useful tactical information.

Mid to large-sized police departments encounter problems in the analysis of case report data, especially when seeking to identify patterns of serial crime. In police parlance, a pattern refers to an individual or group of individuals who are characterized by the fact that they commit a series of criminal offenses, of the same type, using the same method of operation, over an extended period of time. This type of individual is commonly referred to as the serial offender or career criminal.

The accurate and comprehensive identification of problems is fundamental to the problem-oriented policing model. Identification of the career criminal is an important problem area. The Rand study, 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. O'Shea and colleagues posited that 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.

Every department provides officers an opportunity to view data collected in case reports. Mechanisms for reviewing these reports fall into the following categories:

1. Paper files, Clipboards or Notebooks. The officers look for patterns by reading over the reports. This inefficient and ineffective method, given the limits of human information processing, has been in use as long as reports have been filed.
2. PC and mainframe database programs. These automated systems allow investigators to query large data sets to match arrested offenders with incidents possessing similar characteristics. This in some cases will uncover a pattern, albeit reactively.
3. Computer-generated crime map. Relationships can be examined through the mapped display of data. Patterns may not be visible in the map. The pattern may disperse over a large area. The pattern may be masked in an already dense area.

As part of their NIJ project, O'Shea and his colleagues (O'Shea et al. 1995) were interested in formalizing the heuristics used by expert detectives in examining category 1 reports and devising an automated mechanism for applying these heuristics using an automated data system as in category 2. That research team tackled this application area using neural networks for the first time. Instead of the manual system of distributing the records at random to many police experts to visually examine patterns, which has proven to be cumbersome, time consuming and inefficient, they searched for an automated system for distributing cases based on similarities. Their approach included using a back-propagation neural network, and, when that failed, using a k-nearest neighbor algorithm to cluster the data. When using a back propagation network, they trained the network on a data set that included solved criminal cases and tested the network on the crime database. They had no success in this neural network approach because of the use of a supervised network (the back propagation network), which assumes prior knowledge, during training, about all the different possible classes that the crimes will belong to, before the network can generalize about test data. In addition, supervised networks do not work well when the classifications in the data set are continuously changing, which is exactly the case for the crime data. As new crimes are added to the database, the groupings continue to change and consequently the classifications for the records. O'Shea and his colleagues however, had limited success in the second approach. This is mainly due to the way they approached the data. Their approach included limited pre-processing of the data, and application of the k-nearest-neighbor algorithm to all the attributes in a straightforward manner. A major limitation of the k-nearest-neighbor approach was that the researchers still had to choose an optimum maximum size for the output clusters. In addition, the performance complexity of the algorithm created some computational problems. One of the recommendations stemming from the earlier research was that different types of neural networks be tested.

One of the researchers on that project (Muscarello) decided to pursue the use of a different type of neural network, one that was self-organizing, and did not need a predefined training set. The Kohonen network, along with a heuristic based data pre-preparation and a simple expert system were then developed to examine data for potential patterns of serial crime. The results of that research are reported in this document.

3. System design

The type of crimes that the research project will investigate will be limited to armed robberies. However, the methodologies that are discussed in this project for identifying patterns in armed robbery records, are in general applicable to all other types of crimes, with minor (or no) modifications. The choice of armed robbery as a type of crime was made due to the fact that these kinds of crimes usually involve all issues that may be involved in other types of crimes. An armed robbery usually involves an offender, a victim or victims, and a weapon, and sometimes a car, in addition to the general characteristics of a crime. The same issues (objects), however, may not be pertinent to other crimes such as aggravated sexual abuse, as an example, where we may lack the use of a weapon or a get away vehicle.

The research project is divided into three major phases, each of which will be discussed under a separate heading in this paper. Figure 1 illustrates the three different phases in the project design:

(a) Pre-processing phase
(b) Neural network phase
(c) Heuristics system phase

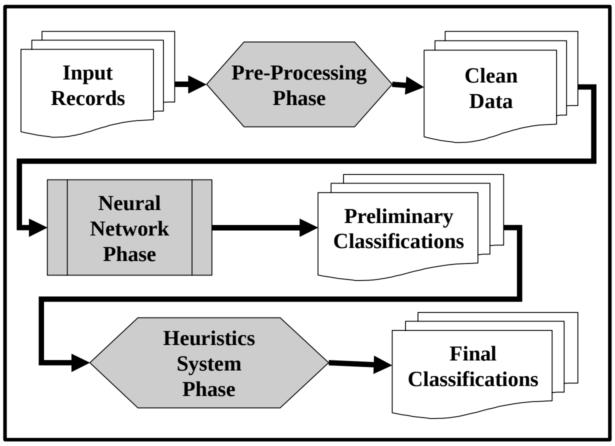


Figure 1. Three phases in system design.

3.1. PRE-PROCESSING PHASE

The concern during the data selection and pre-processing phase of a pattern discovery project is to restructure the data to put it in a format that is useful for the next phase of the project which normally involves the use of a statistical algorithm, a genetic algorithm, a neural network, or any other machine learning methodology that is used for classification tasks (Adiraans and Znatige 1996; Bigus 1996). The issues at hand are the inherent problems in the data, and the format that increases the success of the classification methodology.

The data in the domain can be thought of as four different groups of data: a group that contains the general characteristics of the crime, a group that contains the offender(s) characteristics, a group that contains the victim(s) characteristics, and a group that contains the vehicle (get-away car) characteristics.

3.1.1. Missing values

All of the domain attributes can be categorized as being discrete or continuous numbers. A continuous attribute is one that can take on any real (as opposed to integer) value, such as age and height, for example. Discrete variables can take on any integer value within a specific range, such as race and weapon code, for examples. When a discrete attribute, in a particular record (pattern), is missing (i.e., has no value), the solution in such case is obvious and simple. We add a category value to the attribute (if, it does not exit already) to indicate ‘‘unknown’’. For example, if the offender’s race is missing (or does not belong to the specified range from 1 to 6), we add a new category to the race categories, say category 7, to indicate "unknown". We will assume from here on that the additional "unknown" category is part of every discrete attribute (variable). However, the solution is not so straight forward when dealing with missing values for a continuous attribute. Two alternatives are possible, as we mentioned earlier, either to discard the entire record or substitute an average value for the missing value. We adopt the first approach of discarding the record. The reasoning behind this is two fold. From one angle: substituting the average value for a missing value will falsify the true conception about the fact presented by such attribute. For example, if the weight of an offender is missing then the offender can be thin, medium, or heavy; while substituting the average weight will most likely label him/her as being of medium weight. Secondly, substituting the average weight will ensure that this offender, and similar offenders with missing weights, will be grouped together along with offenders who actually have a numerical value that is equivalent to the average weight.

3.1.2. Standardization

In addition, continuous or discrete values in the fields cannot be used as they are, without modifications. All values, discrete or continuous, should be normalized (standardized) prior to their use. Subtracting the mean from each data value and dividing the result by the standard deviation for the data can accomplish standardization of a continuous variable. Standardizing of a discrete variable can be accomplished in a similar fashion. The standardization of all attributes serves the very important purpose of closing down the gaps among their respective domains. It is also important to note that it is recommended, but not necessary, to use randomized values when using the Kohonen algorithm (Bharath and Drosan 1994), as it is the case in our research.

3.1.3. Categorizations & re-categorization

The standardization of a discrete variable, such as "car color" for example, spreads out the values into equal intervals from the mean, and onto a scale from −4 to 4 (approximately, as the domain for the standard normal curve function, also known as the Gaussian Function). The important issue is that the distance among these categories (normalized values) will be later used to classify their respective records. Therefore, it is safer to put similar categories close to one another to start with, while keeping categories that are completely opposite as far from each other as possible. For our example-attribute: "car color", it would have been more suitable to put the "unknown" category in the middle between the "light" and "dark" categories. This will make the distance between the two opposite categories, "light" and "dark", the farthest, while the "unknown" category is logically in the middle to indicate that the car is as close to being "light" as it is to being "dark". This also will not change the final standardized values, but only changes their respective meanings. The value -1.22 will now correspond to "light", 0 will correspond to "unknown", and 1.22 will correspond to "dark". This technique may not be possible to apply to all discrete variables, as not all of them contain categories that are close to or oppose other categories. Later in this section we will show what re-categorizations will be applied to each of the variables (features/attributes), but first, the re-categorization issue deserves further discussions.

After a careful look at the data attributes, it becomes apparent that some categorizations need to be re-structured. Take for example the offender's eye color. It is virtually impossible for any victim or witness to distinguish a brown from a black (or green from blue) for the eye color of the offender. That holds true for any crime, given that a crime is often committed very quickly, under bad lighting conditions, and/or considering the fear that the victim or witness is subjected to, just to mention a few circumstances that crimes are usually associated with. Therefore, it makes much more sense to re-categorize the attribute "eye color" into three categories: "light", "unknown", and "dark", instead of the many possibilities for values (found in the data) that are used to represent eye color, such as: brown, hazel, green, blue, black, or any combination of light or dark colors.

Other categories, such as the "location" and "taken-codes", are just too many, and it will enhance our classification task tremendously to group these categories into what investigators really look for in such cases. Categorization will be necessary for the year of the vehicle. Ordinary people usually cannot distinguish a 1988 from a 1989 vehicle, and therefore these values become meaningless in this sense. It is more appropriate to categorize such attribute, the vehicle year, into three main categories: "new", "medium", and "old". The "new" category will be used for vehicles that are 1991 or later, the "medium" category will represent vehicles that are between 1985 and 1990, while any vehicle's year below 1985 will fall into the "old" category. A similar categorizations and re-categorizations approach will be applied to other fields in the crimes data, including but not limited to: time and location of crime, type of weapon, items taken, etc.

3.1.4. Grouping

Every offender's information is vital to our classification task. Therefore, each of the offenders involved in a single crime, if the crime is committed by a group of offenders, will be treated as if he/she is totally responsible for the crime, from the point of view of classification of patterns. That is important, as the same offender can be involved in other crimes that he/she committed individually or with different other offenders. Every victim's information, on the other hand, is not that important to our classification task, since the information that is kept on a victim is limited to sex, race, age, and whether the victim was injured or not. This limits the process of discovering patterns to identifying if the offender is targeting a particular group of victims from the sex, age, race point of view, and whether he/she is causing any harm to those victims.

Our approach to this issue will merge multiple records of victims that were involved in a single crime, into a single record. This requires the following minor adjustments in the categorizations of the fields that depict the victim's information:

- Sex – In addition to the "Male" and "Female" categories, we add a new category to represent general sex and that indicates that multiple sexes were victimized by this crime.
- Race – Similar to the sex field, we add another category to indicate multiple races were victimized by this crime.
- Age – The average age of the victims will be used. This will not affect the detection of any patterns that are in this field, as the average age will still reflect if an offender has targeted a particular age group
- Injury – No adjustment in the categorizations of this field is needed. If any of the victims involved in the crime is injured, then the value of this field will reflect a "YES"; otherwise, the value of the field will reflect a "NO". The hypothesis here is that if the offender has injured any victim, then that makes injury a pattern for his/her crimes.

The above adjustments will be able to depict any patterns that are inherent in the victims information, when several victims are subjected to a particular crime; otherwise, the offender is operating at random and reacting to whatever situation he/she is in, which may very well be the case.

We should remind the reader that the pre-processing phase of any data mining system is the most important phase. Clean, consistent and complete data can improve the results tremendously. The Good In Good Out (GIGO) principle applies here, just as it applies in every computer science application. Another issue of importance is the categorization issue. The proper categorizations need to be carefully thought of for each attribute. Categories for each attribute should also be kept at an absolute minimum. The number and type of the categories for each attribute should not be set for all types of crimes. The categories will vary by the nature of the crime, and should reflect that. Much of the preprocessing activity is based on an understanding of the criminal investigative procedure, the knowledge of domain experts, and knowledge of the limitations of the data obtained in victim interviews.

3.2. NEURAL NETWORK PHASE

Many machine learning algorithms, such as genetic algorithms, decision trees, association rules, statistical algorithms and inductive logic programming, have been used in data mining tasks (Adriaans and Znatige 1996), which our task is considered a type of. However, Moustakis et al. (1996) have reported that statistical algorithms and neural networks are rated at the top, and way ahead of, other machine learning algorithms when it comes to clustering and prediction tasks. We investigated, in detail, both of these tools, and decided to use neural networks, and namely Kohonen networks because they offered several advantages over other neural networks and statistical algorithms. The advantages can be summarized as follows:

- The classification results produced by most statistical algorithms rely heavily on the order the data patterns are presented to the algorithm. Such dependency can cause results to vary and the choice of the proper order can be an extremely difficult task, if not impossible, considering the huge volumes of data that are expected in our application domain.
- Most statistical algorithms require the classes that are inherent in the data patterns (target classes), to be linearly separable (i.e., can be separated by a straight line). The complexity of our domain data and the continuously changing quantity and quality of the target classes constitutes a direct violation to the use of such statistical algorithms.
- There are statistical algorithms that do not exhibit the feature of dependency on the patterns' order and do not require linear separability of the classes. These algorithms suffer from the disadvantage of being computationally complex and rely heavily on the choice of multiple parameters that are required by such algorithms.
- Kohonen networks are unsupervised neural networks that use a computationally simple algorithm (Kohonen 1990). The unsupervised environment is a requirement for our domain, as the number and type of classifications are not pre-known. The failure of O'Shea and his colleagues, when they employed a BPNN (a supervised network) is a substantial proof of the necessity to use an unsupervised algorithm. Other unsupervised neural networks suffer, again, from being computationally complex and relying heavily on the proper choice of the network parameters. Kohonen networks were proven to be as efficient, and in some cases more efficient, than other statistical and machine learning algorithms in the experiments published by Kohonen et al. (1998). This in addition to their capabilities of handling vagueness and fuzziness that are inherent in data such as the one we are faced with, and their capabilities to handle large numbers of attributes as a result of their

parallel structures (Kohonen 1988; Adriaans and Znatige 1996). The simplicity of the algorithm for the Kohonen networks has made them very attractive in many application areas especially in data mining (Bigus 1996), classification (Xia 1996) and prediction (Lee et al. 1996).

3.2.1. Hybrid network

The straight forward approach to arriving at the final classifications using neural networks is to feed all the attributes into one huge neural network. This approach assumes that all the attributes are of similar importance in determining the final classifications. O'Shea et al. (1995) report that when criminal investigators are challenged with such classification tasks, they mentally group the crime data into categories they can easily pursue and analyze. While some investigators analyze data of the vehicles' involved in the crime, others may tackle the offenders' characteristics, the particulars of the crime, or the characteristics of the victims. Thus, an investigator will not look at the whole picture (all the data at once) unless he/she finds something interesting in one or more of these groups of characteristics. Our methodology simulates, to some extent, the actual approach that criminal investigators take by grouping similar attributes and classifying them accordingly, instead of feeding all the attributes into one huge neural network. We group the features into different groups of similar features (similar importance), and then present each group of features to an independent Kohonen network. These Kohonen network modules will work independently and in parallel, each of them producing its own output. The outputs will then be combined to produce the required pattern classifications. Figure 2 shows the project network design that pictorially illustrates our approach.

To simulate the methodology used by criminal investigators when faced with such classification tasks, we employ a network of cascading Kohonen neural networks. Each intermediate Kohonen network will be responsible for classifying a group of attributes such as the characteristics of the offenders for example. The resulting classifications will be joined again and fed into a final Kohonen network to arrive at the final classifications for the project. The number of input neurons for the Kohonen network is dictated by the number of attributes, while the output neurons will be set constant to the maximum number of input patterns in the data files.

Upon the introduction of each input pattern, the network will apply the basic Kohonen algorithm for updating the network weights. When the last input pattern gets introduced, we may assume that the network has learned the density function that is inherent in the input patterns. Each input pattern will then be compared to the final weights of the network, and the set of

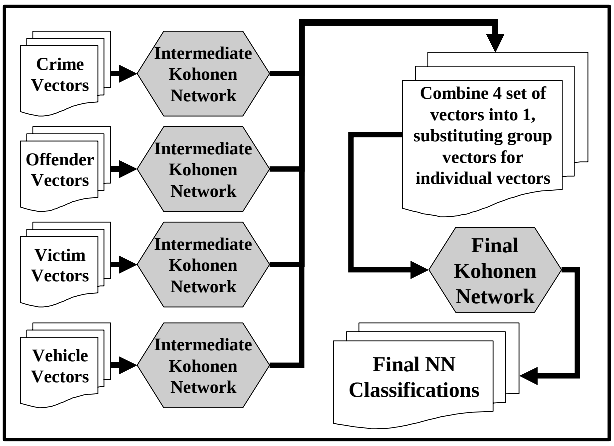


Figure 2. Design of the neutral network phase.

weights that has a closer distance with the input pattern will be declared the proposed class for that input pattern (Kohonen 1990). Upon completion of training, each intermediate Kohonen network has learned the classifications inherent into the input patterns, and the network's final weights represent those classifications. Each input pattern will then be compared to all the final weights of the network, and the weight vector that is closest to the input pattern will be declared to be the proposed class for that input pattern. That same class (final weight vector) will be, most likely, associated with multiple input patterns, and hence, represents the center for its members.

The four groups of characteristics are then joined into a single table consisting of records, each of which represents an offender. In other words, each record will convey the information that an offender O committed crime C using car R on victims V, where O, C, R, and V are the offender's characteristics, crime's characteristics, car's characteristics, and victims' characteristics respectively. But, instead of using the original data fields for O, C, R and V, we use the corresponding vectors from their respective class' centers, achieved from the intermediate Kohonen networks. The records in this cumulative table will then be fed into the final Kohonen network. The results of the final Kohonen network will constitute the preliminary classifications suggested by the system for the original input patterns.

3.3. HEURISTICS SYSTEM PHASE

Our concern during this phase is directed at enhancing the classifications that were produced by the last phase. We employ heuristics that depict how an actual expert investigator in the field would look at the classifications produced by the neural networks. Different investigators normally employ different heuristics based on their experience and success rate with their approach. For example, while one investigator might be concerned with any criminal attacking a particular group of victims (older ladies for instance), another investigator might not be concerned with such facts operating under the assumption that a criminal will attack randomly if he/she finds the crime circumstances to be convenient.

We focus on general heuristics that we find necessary to employ for the purpose of enhancing the resulting preliminary classifications, instead of focusing on what each investigator might consider important, as this will always be left to his/her personal judgment:

The first of such heuristics is directed at rectifying a shortcoming in the neural network classifications. While neural networks will group crimes based on their time of occurrence, they will not group crimes that occurred before midnight with those that occurred after midnight, even if those crimes are similar when it comes to all the other attributes. That is due to the big gap among such times (i.e., time $=$ 23:40 and time $=$ 00:30 are too far apart numerically, even when standardized). While this heuristic enhances the classifications by adding records to each class, the next heuristic extracts records from each class in an effort to also enhance the classifications produced by the neural networks.

The next heuristic starts by identifying those records, within each class, that represent crimes that were committed on the same day. Such a group of records suggest that the crimes were committed by the same serial criminal and on the same day. If so, then the serial criminal must have had enough time to move from one location (address) to another. The heuristic aims at verifying such assumptions or rejecting them. We employ a specialized software system, such as MapInfo, to change the address field in such records from an actual numerical address to geographical coordinates and distances. Using such a system, the distance between any such two locations is determined, then divided by the average speed of forty miles per hour, to determine the time needed to move from one of the two locations to the other. If the difference between the two actual times for the crimes is not greater than the time needed to move between the two crimes' locations, then the two crimes do not belong to the same class. Under such circumstances, it is hard to determine which one of those records does not belong to the class, when such two (or more) records exist in the same class. Therefore, both of these sets of records will be marked for special identification purposes, pending further field investigation by actual experts in the field.

We also chose to employ another heuristic that is aimed at identifying records, within each class, that correspond to the same crime being committed by more than one offender. Such records normally have the same case number, and thus are easily identifiable through a simple query.

What such situation suggests is that the criminals have similar characteristics, which is totally a different issue of investigation than the one we are interested in. We are interested in identifying patterns that will lead to further investigation of the cases that we identified, no matter how many offenders are involved in such cases. Thus we eliminate such records from the class.

4. Results and conclusion

The results that were obtained from the research were very encouraging. The system provides criminal investigators with a tool for identifying patterns in criminal databases. The results are, by no means, meant to be conclusive in terms of indicting such serial criminals, but will serve as a very fine starting point for investigators, and the classifications produced by the system will command the worthiness of re-opening the relevant actual crime files for further field investigations. Our excitement with the results was substantiated with the reaction that we received from a few experts in the field. They valued the results and had very optimistic feelings as to how such a system can prove very beneficial in practical use.

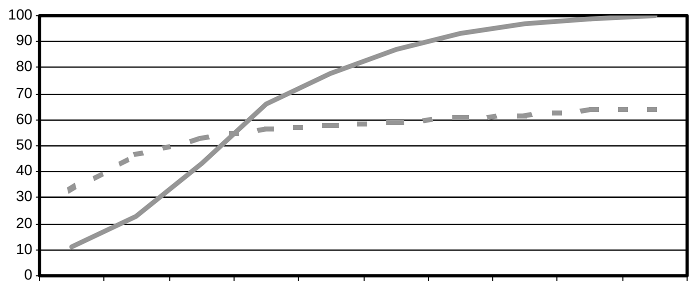
The system offers a modular design of neural networks that allows each expert investigator to customize the system to his/her style of investigations. For example, if the victims' characteristics are not very important to a particular investigator, he/she can select a threshold for the victims' group of characteristics that reduces the number of victims' classes to one if he/she chooses to do so. The fine-tuning of each neural network allows investigators (users) to incorporate their own style of thinking and mental approach, as well as their systematic practices into the system.

Although our results looked encouraging from a qualitative point of view, we needed to quantify these results. Therefore, we computed some statistics about these results as compared to the test data. The test data consisted of crime records that were classified by police officers for serial criminal patterns. A unique pattern number was used to identify crime records that belong to the same class. To eliminate any word confusion, we will label patterns (classes) identified by police as "actual-patterns", while the classifications (patterns) identified by our system will be labeled as "predicted-patterns". We computed two main statistics: the percentage of actual-patterns that are spread over ten predicted-patterns or less and found that value to be 100% In fact, it was found that 84% of actual patterns are spread over five predicted patterns or less; as well as the percentage of predicted-patterns that contained ten actual-patterns or less and found that value to be 64%.

The qualitative results can be interpreted in our domain to mean that instead of distributing crime files at random to many different field investigators; we need only ten investigators to distribute all predicted-patterns among them, and each of these investigators has a 64% chance of being allocated a set of predicted-patterns that contained at least one actual-pattern and a maximum of ten actual patterns. Figure 3 summarizes these quantitative results graphically.

Although better values for these statistics would have been more desirable (i.e., decreasing the number of needed investigators as well as the number of actual-patterns contained within each predicted-pattern), we remind our reader that these results constitute a significant improvement over the manual current system for case assignments. An actual-pattern under random assignment may end up being distributed to several officers, and each officer may have at his/her desk cases that have nothing in common. We should also point out that the actual-patterns in the test data were identified, by police investigators, on a crime file basis rather than being identified on an offender basis (as it is the case in our methodology). For example, if an offender O1, who committed a crime C1, was identified as the same serial offender O2, who committed a crime C2, then the police records will identify every offender associated with crimes C1 and C2 using the same actual-pattern number. In our methodology, only those two offenders' records could be classified in the same predicted-pattern, while other offenders associated with those two crimes could belong to other predicted-patterns (as their attributes may suggest). This causes the values for our two statistics to decline, as a single case would most likely be dispersed across different predicted-patterns.

It is also important to point out that the manual system results in only 1% success rate in closing cases as serial patterns, while it is estimated that some 80% of the cases should fit these patterns (O'Shea et al. 1995). This point and our results using only four attributes confirm what was suggested by O'Shea et al. (1995) that this combination of attributes is considered by a panel of experts to be the core data used in the investigations, no matter what special techniques are employed by the detectives. They also suggest that a system



	1	2	3	4	5	6	7	8	9	10
Cumulative percentage of actual-patterns	11	23	43	66	78	87	93	97	99	100
Cumulative percentage of predicted-patterns	33	47	53	57	58	59	61	62	64	64

Figure 3. Summary for quantitative results.

such as ours is a valuable tool for investigators, simply because the number of attributes and the sizes of the data files in this domain do not allow investigators to look at the whole picture, but only a very narrow view of it.

5. Research contribution

Our approach to the data analysis and pre-processing of data have proven to be fruitful and will help, from an information systems point of view, to make recommendations to criminal agencies on what information need to be kept on criminals and what type of format is best for keeping such information. Our categorization and re-categorization will serve as guidelines for criminal experts on what constitutes important information to pattern discovery of serial criminals.

The research represents the first time that a hybrid system using neural networks and a rule based heuristics system has been implemented in this application domain. It is also the first time, to our knowledge, that a Kohonen network has been used to build a clustering system for recognizing and grouping potential serial crimes. The approach enhances the overall classification task, because it simulates, to an extent, the thinking of expert investigators in the field, the modular approach of this design enables the user of this system to have better control of the system, and allows for multiple check-points and validation processes. This helps to overcome the usual problem that neural networks present with respect to explanations – that is, an inability to explain or make clear what was done by the "black box" process. The design offers a new approach to data mining instead of feeding all the attributes into one single Kohonen network (Kohonen 1990; Bigus 1996). The approach of hybrid design of Kohonen networks can be adopted in several other application areas including, and not limited to: marketing and consumer behavior studies, medical/insurance field and risk analysis, and last but not least fraud detection for credit cards, insurance and/or medical billing. These are explained as follows:

- *Marketing*: A similar approach can be used to detect patterns in customers purchasing behaviors and interests. The grouping of similar attributes and feeding them into a hybrid network of Kohonen neural networks will allow better control of intermediate and, consequently, the final classifications produced by the system. The manufacturer or retailer can then incorporate general marketing heuristics as well as their own special set of heuristics to enhance the final classifications. We are confident that such an approach will prove to be more successful than feeding the huge number of attributes into a single neural network in the business areas of product pricing, expanding the customer base, as well as market and sales forecasting which all are similar in nature to marketing.

- *The Medical/Insurance Field*: Patterns in people/patients records can help identify and/or predict certain future risks with respect to diseases that people may be vulnerable to. Similar to the above analysis with respect to marketing, our approach can help enhance the classifications that are needed by medical researchers to treat or prevent the occurrence of certain diseases such as heart attacks or strokes, as well as to help insurance companies run a more cost effective business by identifying and reducing the number of high-risk situations.
- *The Field of Fraud Detection*: This is a very similar application area, in nature, to the area of the reported research, as both areas identify patterns of criminal behavior. The area of fraud includes, and is not limited to: credit card fraud, insurance fraud, and medical billing fraud.

6. Further Research

The research that was presented here is continuation of a complete project [1] that is aimed at providing criminal investigators with computerized tools to identify serial criminals through patterns in criminal behavior. The tools will prove to be of great practical and social value. As part of this valuable research, our future research will include:

- Refining the use of attributes to accommodate weather, sunlight, sunrise and sunset, and exact geographical coordinates of crimes.
- Identifying and employing other heuristics that are either general or specialized in nature. Examples of such heuristics include taking into consideration the type of day that the crimes occurred on, and the geographical districts that are targeted by offenders. This will be beneficial in identifying patterns of offenders that strike on weekends or holidays or operate within a particular geographical region.
- Incorporating into the system the capability of accommodating investigators who want to employ their own heuristics that are not pre-known to the system, if they wish to do so.
- Extending the system to classifications of other categories of crime.
- Automating the steps involved in the system, such that they can be operated via a graphical user interface (GUI) that is very friendly and easy to use by non-computer specialized users, which most criminal investigators are assumed to be. This includes providing help facilities that suggest particular hints on decisions that the user might have to make, such as in the case of the categorization issue.

- The data files that such a system normally has to deal with are very huge, and consequently will impact the performance of the system in the traditional sequential environment. Further research may examine porting the system, and all its detailed steps, to a parallel environment.
- A graphical methodology is needed to present the user with the huge amount of classification results, produced by the system. Such a methodology should allow the user to move the system into the next phase (step) when he/she sees the proper classifications have been achieved for a particular group of categories.
- Incorporating an explanation subsystem in the overall system to give both the implementers and the users the flexibility of knowing, at any given time, the answers to his/her "how"(s) and "why"(s) on the operation of the system.

References

Adriaans, P. and Znatige, D. (1996). Data Mining. Addison-Wesley.

Ball, G. H. and Hall, D. J. (1967). A Clustering Technique for Summarizing Multivariate Data, Behavioral Science 153–155.

Batchelor, B. G. and Wilkins, B. R. (1969). Method for Location of Clusters of Patterns to Initialize a Learning Machine, Electronic Letters 481–483.

Bharath, R. and Drosen, J. (1994). Neural Network Computing, Windcrest/McGraw-Hill.

Bigus, J. P. (1996). Data Mining with Neural Networks: Solving Business Problems- from Application Development to Decision Support, McGraw-Hill.

Carpenter, G. A. and Grossberg, S. (1987). ART2: Self-Organizing of Stable Category Recognition Codes for Analog Input Patterns, Applied Optics.

Carpenter, G. A. and Grossberg, S. (1987). A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine, Computer Vision, Graphics and Image Processing 54–115.

Carpenter, G. A. and Grossberg, S. (1998). The ART of Adaptive Pattern Recognition, Computer 77–88.

Carpenter, G. A. and Grossberg, S. (1990). ART3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architecture, Neural Networks 129–152.

Chung, F. L. and Lee, T. (1994). Fuzzy Competitive Learning, Neural Networks 539–551.

Cover, T. M. and Hart, P. E. (1967). Nearest Neighbor Pattern Classification, IEEE Transaction on Information Theory 21–27.

Dahbur, K. and Muscarello T.(2001). An artificial Intelligence Approach to Pattern Discovery in Criminal Databases, Proceedings of the IASTED International Symposia on Applied Informatics, 11–19.

Diday, E. (1973). The Dynamic Clusters Method in Non-Hierarchical Clustering, International Journal on Computer Information Science 61–68.

Dunn, C. (1976). Patterns of Robbery Characteristics and Their Occurrence Among Social Areas, Analytic Report 15, Utilization of Criminal Justice Statistics Project, Criminal Justice Research Center.

Kohonen, T. (1988). An Introduction to Neural Computing, Neural Networks, 1: 3–16.
Kohonen, T. (1990). The Self-Organizing Map, Proceedings of IEEE, 78 (9).
Kohonen, T., Barna G., and Chrisley, R. (1998). Statistical Pattern Recognition with NN: Benchmarking Studies, Proceedings of IEEE International Conference on Neural Networks, 1–61 – 1–68.
Lee, D. H., Payne, J. S., Byun, H. G., and Persand, K. C. (1996). Application of Radial Basis Function Neural Networks to Odour Sensing Using a Broad Specificity Array of Conducting Polymers, Proceedings of International Conference on ANN 299–304.
MacQueen, J. Some Methods for Classification and Analysis of Multivariate Data, Proceedings of Fifth Berkeley Symposium on Probability and Statistics, University of California Press, Berkeley.
McKinzie, P. and Alder, M. (1994). Unsupervised Learning: The Dog Rabbit Strategy, IEEE International Conference on Neural Networks 616–621.
Moustakis, V. S., Lehto, M., and Salvendy G. (8-1996). Survey on Expert Opinion: Which Machine Learning Method may be used for Which Task? Journal of HCI, 221—236.
Nayler, J., Higgins, A., Li, K. P., and Schmoldt, D. (1988). Speaker Recognition Using Kohonen's Self-Organizing Feature Map Algorithm, Abstracts of the First Annual INNS Meeting.
O'Shea, T., Gardiner, J., Illingworth, W., Maltza, M., and Muscarello, T. (1995). Application of Artificial Intelligence Methodology to the Problem of Identifying Career Criminals in Large Crime Databases, Project Report.
Pao, Y. H., Park, G.H., and Sobajic, D.J. (1994). Learning and Generalization Characteristics of Random Vector Functional Link Net, Neurocomputing 163–180.
Park, D. C. and Dagher, I. (1994). Gradient Based Fuzzy C-Means (GBFCM) Algorithm, Proceedings of IEEE Conference on Neural Networks 1626–1631.
Xia, X. (1996). Neural Network Models in Predicting Insurance Insolvency and Detecting Insurance Claim Fraud, Ph.D. Dissertation, University of Texas at Austin.