SILVIA FERRARI,
KELLI C. BAUMGARTNER,
GEORGE B. PALERMO,
ROBERTA BRUZZONE,
and MARCO STRANO

© EYEWIRE

COMPARING BAYESIAN AND NEURAL NETWORKS FOR DECISION SUPPORT IN CRIMINAL INVESTIGATIONS

Network Models of Criminal Behavior

Criminal psychology is beginning to focus on modeling the effect of mental processes and environmental stimuli on offender behavior [1]. This article presents a knowledge-based-system approach to deriving empirical models of this behavior. The inputs to the system are the psychological characteristics and environmental variables that reflect the mental state of the offender and the circumstances of the crime, while criminal actions are viewed as the outputs. Although criminologists and psychologists can identify relevant input and output variables, the processes underlying criminal behavior are only partly understood and thus cannot be modeled from first principles.

Once a criminal case is cleared, investigators file a record that includes background characteristics and psychological diagnoses of the convicted offender as well as forensic evidence obtained from the

Digital Object Identifier 10.1109/MCS.2008.924037

crime scene. This practice creates large databases of crime-scene and offender information for major felonies, such as murder, rape, and arson. Consequently, knowledge-based systems trained from data can potentially assist in the development of tools to support criminal investigations.

In this article, we construct neural network (NN) and Bayesian network (BN) models of criminal behavior from databases of single-victim homicides. We illustrate how these models can be used to estimate the profile of the offender, including a psychobehavioral portrait that helps detectives select and interrogate suspects. A broader implication to the field of psychology is the development of scientific hypotheses from relationships and patterns that emerge through the analysis of network models trained on data.

CLASSICAL CRIMINAL PROFILING

Criminal profiling is the study of behavior that identifies the characteristics of the offender, his or her modus operandi, and the motivation for the crime. Criminal profiling techniques are intended to assist investigators by narrowing the scope of the investigation, predicting the social and psychological characteristics of the offender, and suggesting strategies for apprehending and interviewing suspects [3]. In current practice, a team of medical examiners, detectives, and psychologists attempts to reconstruct the motives for the crime as well as the events that took place at the crime scene. The investigative team then tries to deduce the psychobehavioral portrait of the offender based on professional training and previous investigations. This approach, known as *behavioral evidence analysis*, is limited by the team's ability to analyze patterns from data and is subject to individual prejudices and biases [4]–[5].

An alternative approach, known as *inductive profiling*, attempts to generalize behavioral patterns obtained from crimes of convicted offenders to crimes of unknown offenders [6]–[8]. Inductive profiling techniques rely on dichotomic psychology methods, which classify the offender's behavior and profile into one of two categories. One example of inductive profiling is the FBI model, which classifies offenders as either organized or disorganized [3]. The *organized* offender, who displays characteristics of maturity and resourcefulness, carries out a crime that is methodical and premeditated. In contrast, the *disorganized* offender does not plan the crime, and thus the crime scene shows evidence of haphazard behavior [9], [10]. Another example is circle theory, which classifies offenders as either *commuters* or *marauders*, depending on the distance traveled from their home base when committing a crime [6]. By defining these behavioral categories, multidimensional scaling and clustering techniques are used to classify offenders based on the forensic evidence [11], [12]. Circle theory has been used successfully in predicting the approximate residence location of serial homicide offenders [13].

The literature on inductive profiling suggests that different styles of homicide reflect differences in the personality and background of the offender [14]. However, inductive profiling techniques have not been able to produce accurate offender profiles in single-victim homicides [11]–[15]. This shortcoming is typically attributed to both the complexity of human behavior and the large number of situational variables, such as the place where the aggression began, the disorder ensued by a fight, and the interaction between the victim and offender. Consequently, as advocated in [14] and [15], a more realistic and utilitarian understanding of mental processes leading to criminal actions requires development beyond behavior classification.

By using NN or BN models of offender behavior in single-victim homicides, relationships between the behavioral and situational variables can be considered simultaneously. In this article these relationships are learned and quantified by means of training algorithms that utilize observations of the criminal process without postulating offender categories a priori. A trained network model has two main applications. First, the arcs in the network represent correlations between the variables and thus can be used to form new hypotheses and theories of criminal psychology. Second, the network model can be used in new criminal investigations to infer offender characteristics from the forensic evidence obtained from the crime scene.

BEHAVIORAL MODEL

The development of a behavioral model relies on two basic premises, namely, behavioral consistency across offenses and causal relationships among offender characteristics, such as criminal record, psychiatric disorders, and the behavior exhibited during a crime. Experts in psychology, sociology, forensic medicine, and psychiatry [25] have identified variables that are relevant to single-victim homicides and have organized these variables according to the criminal profiling (CP) taxonomy in Table 1. Examples of these variables and their possible values are provided in Table 2. As illustrated in Figure 1(a), external and internal stimuli, such as environmental and psychological attributes, are viewed as input variables driving the criminal behavioral process. Crime-scene circumstances that are beyond the offender's control are considered to be uncertain disturbances. Once a behavioral model linking the inputs to the outputs is obtained from a database of cleared criminal cases, the model is inverted to obtain a profile of the offender in an unsolved case, as shown in Figure 1(b).

Knowledge-based systems, such as BNs and NNs, can be used to obtain a behavioral model and help solve a new criminal case. BNs describe relationships among variables by means of a joint probability mass function, whereas NNs use a multivariate nonlinear function obtained from the superposition of basis functions, such as sigmoids. The network parameters that best explain the data are determined through training, which specifies a model of the

TABLE 1 Taxonomy of criminal profiling variables. These variables model offender behavior in single-victim homicides. This taxonomy, as well as each variable's definition and range, are determined by a multidisciplinary team of experts through panels organized by the International Crime Analysis Association. This study identifies up to 105 candidate variables that represent the offender's actions and decisions during the crime as well as the stimuli that drive them. Also, for each cleared offense the value of each variable can be determined from investigations and interviews with the offender after he or she is convicted and incarcerated.

Crime scene analysis (CSA)	Physical elements and characteristics of the crime scene that represent consequences of offender behavior, such as time and place where the victim is found, neighborhood's ethnic and social characteristics, correspondence between where the victim is found and where the murder took place, and crime typology (premeditated, rape, or arson).
Victimology assessment (VA)	Victim characteristics, such as background information, age, sex, education, and occupation.
Medical examiner report (MER)	Medical characteristics of the victim obtained by an examiner at the scene of the crime and during the autopsy, such as the cause and time of death, type of lesions, and signs of self-defense.
Offender assessment (OA)	Background and demographic information about the offender, such as age, sex, and family status, as well as his or her criminal characteristics, such as criminal career, forensic awareness, and geographical residence with respect to the victim.
Transactions with the victim (TV)	Attributes of the relationship and interactions between the offender and the victim before and during the crime, such as the social and cultural differences between the offender and the victim.
Psychological and psychopathological profile (PP)	Characteristics of the mental state of the offender before, during, and after the crime; attributes of the offender psychology, personality, intelligence, and attitude toward the crime, such as the motive; and the actions of the offender while the investigation was taking place.
Psychiatric diagnosis (PD)	Psychiatric disorders observed in the offender, such as anxiety, depression, or forms of paraphilia.

system. The database of cleared criminal cases is divided into a training set and a validation set. The training set is used to estimate the network parameters, while the validation set is used to evaluate model predictions. When the model predictions are not satisfactory, an alternative set of models is considered by changing the variables and architecture of the network.

BN MODELS

BNs are probabilistic models that combine Bayes' rule and graph theory to provide a mathematical description of an uncertain situation or random process [19], [20], such as a crime, as shown in Figure 2. A random process produces precisely one outcome from the sample space S, which is the set of all of its possible outcomes. If a random-process outcome is not numerical, it is associated with a number, called the *value*, by a real-valued function on S, known as the *random variable*. For example, the gender of the offender has two possible outcomes, which may be associated with the numerical values $\{1, 2\}$ by a discrete random variable called *gender*. A BN models a multivariable random process by a pair $B = (G, \Omega)$ comprised of a directed graph

TABLE 2 Example of criminal profiling (CP) variables used by a team of experts for modeling the behavior of offenders in single-victim homicides. Each variable is assigned to a node in the network model that takes on a finite set of values, referred to as the variable range, which are determined by expert criminologists and psychologists. The cardinality of the range of each variable, shown in the third column, illustrates the dimensionality of the problem.

Random Variable	Definition (CP Group)	Range Cardinality
Z_9	Typology of offender's judicial record (OA)	4
Z_{13}	Immigration history (OA)	2
Z_{14}	Typology of offender's relationship with the victim (TV)	8
Z_{18}	Social-cultural differences between offender and victim (TV)	6
E_6	Place where the aggression began (CSA)	14
E_{23}	Origin or source of the crime weapon (CSA)	2
E_{14}	Typology of housebreaking signs (CSA)	5
E_{29}	Objects or accessories left on the victim (VA)	12

$G = (U, A)$ and a parameter structure Ω. The set $U = \{X_1, \ldots, X_n\}$ of random variables associated with the random process is the *universe*, and the union of their domains is the sample space S. Although continuous random variables can be included [21, pp. 521–540], we assume that each random variable $X_i \in U$ is discrete and, thus, has finite range $\{x_{i,1}, \ldots, x_{i,m_i}\}$, where $x_{i,j}$ denotes the jth value of X_i.

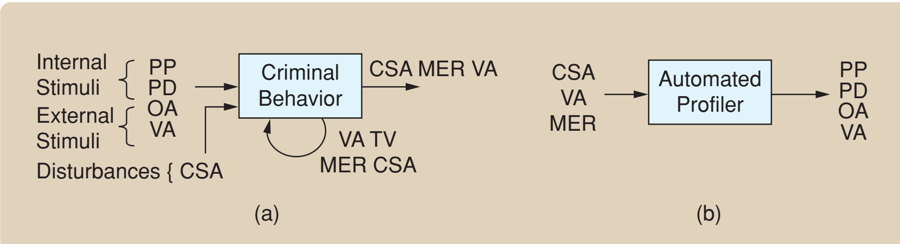


FIGURE 1 Example of (a) criminal behavioral model and (b) automated profiler, with inputs and outputs defined as in Table 1. The inputs driving criminal behavior are comprised of internal stimuli, which are determined by the offender's psychological characteristics and diagnosis, as well as external stimuli that may be reflected in the offender's background and transactions with the victim. Some of the crime scene variables can be viewed as disturbances, while others can be used to describe how the crime took place and, thus, comprise the state. The crime scene analysis, including the medical examiner report and victimology assessment, are the output variables observable from the investigations. These variables can be used as inputs to the (b) automated profiler, derived from the (a) behavioral model, to predict all of the offender's variables.

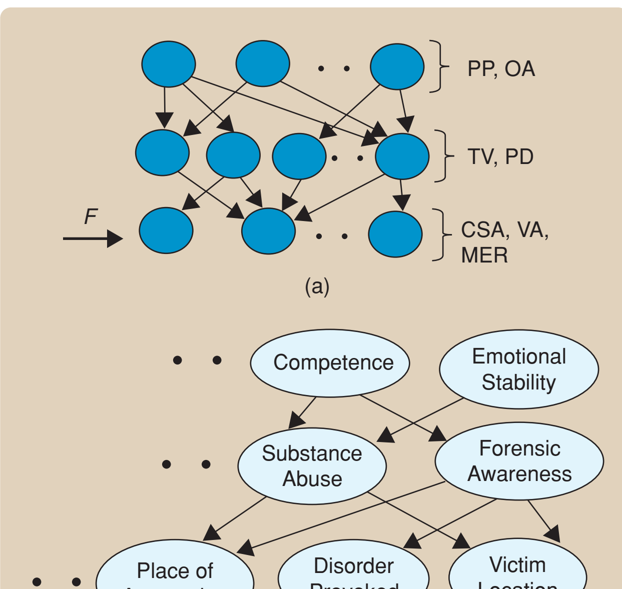


FIGURE 2 General Bayesian network (BN) model of (a) criminal behavior and (b) selected nodes and connections within the BN structure. The offender's psychological characteristics and background may determine his or her psychiatric disorders and transactions with the victim. These variables, in turn, are believed to determine the offender's behavior at the crime scene as well as the victim's attributes, which are observable from the investigation. When this structure is learned from real cases, it can be used to reveal interdependencies among the variables, such as those highlighted in (b).

The random variables in a BN are represented by nodes, which are connected by directed links or *arcs* in the set A. An arc from a random variable X_i to a random variable X_j represents conditioning of X_j on X_i, indicating that knowing the value of X_i provides partial knowledge about the value of X_j. This knowledge is captured by the conditional probability mass function $p(X_j|X_i)$, which is *attached to* X_j. X_j is a *child* of X_i, and X_i is a *parent* of X_j. When a random variable X_j is conditioned on more than one random variable in U, X_j has attached a multivariate conditional probability mass function $p(X_j \mid pa(X_j))$ conditioned on all of its parent nodes in U, which are denoted by $pa(X_j)$. When a random variable X_i has no parents, it has attached a prior probability mass function $p(X_i)$. In a BN, every prior and conditional probability mass function is expressed as a conditional probability table (CPT). A CPT expressing a multivariate probability mass function for n random variables $X_1, \ldots, X_n$ is an n-dimensional array of size $m_1 \times m_2 \times \cdots \times m_n$, where m_i denotes the number of values in the range of X_i. For example, a two-dimensional CPT expressing the conditional probability mass function of the random variable E_{23} conditioned on Z_{12} is shown in Figure 3, where the rows correspond to values in the range of E_{23}, and the columns correspond to values in the range of Z_{12}. The BN parameter structure Ω is the set of all CPTs associated with the universe U.

Training a BN is the process by which the set A of arcs and the parameters Ω are estimated from data. Given Ω and A, the BN B defines a joint probability mass function that is specified in terms of the factorization

$$p(U) = p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid pa(X_i)), \tag{1}$$

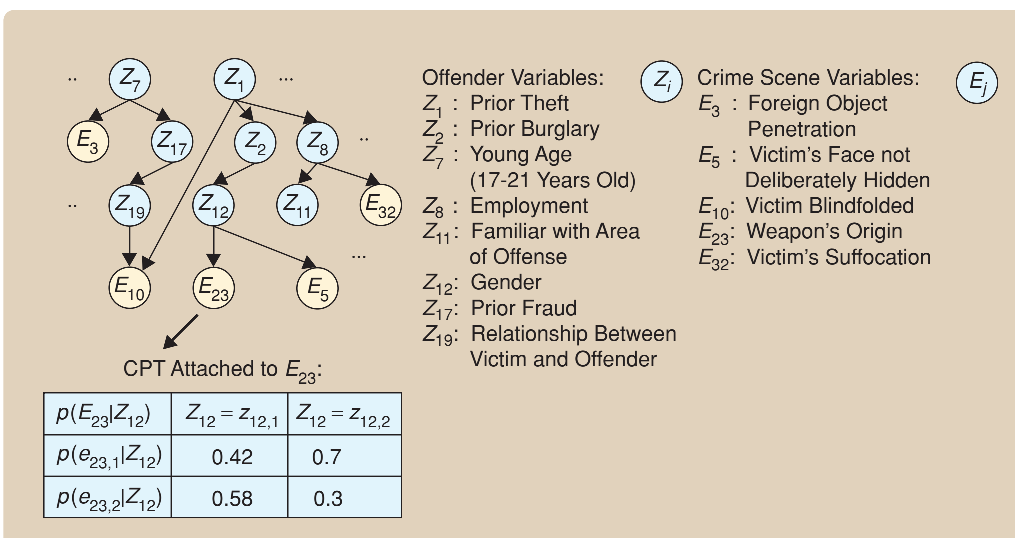
obtained from the multiplication rule of probability calculus, that is, $p(X_1, \ldots, X_n) = p(X_1) \cdot p(X_2|X_1) \cdot p(X_3|X_1, X_2) \cdots p(X_n|X_1, \ldots, X_{n-1})$ [22, p. 24]. Here, $p(X_i, X_j)$ denotes the joint probability mass function of X_i and X_j, and $(\cdot)$ denotes the multiplication of probabilities [22, p. 24]. Thus, from (1) the joint probability mass function $p(U)$ over the universe, which is expressed as an n-dimensional array, is obtained by multiplying the n BN CPTs $p(X_i \mid pa(X_i))$, $i = 1, \ldots, n$, where, each CPT $p(X_i \mid pa(X_i))$ with v parent nodes in the set $pa(X_i)$ is a $(v+1)$-dimensional array in Ω. The set A of arcs, which constitutes a *BN structure*, can be obtained from expert knowledge or from a training database D. A training database $D \equiv \{C_1, \ldots, C_d\}$ is a set of samples or *cases*, where each case C_i is obtained from the random process and contains the measured values of all random variables in U. When expert knowledge is unavailable, A can be obtained from D by means of a structural training algorithm that utilizes the frequency of events observed from d cases. Batch training algorithms for this purpose are reviewed in [21, Part IV]. Given the structure A, the parameters Ω that best explain the database D are determined by parameter-training algorithms based on maximum likelihood estimation, as explained in the following sections.

Once the BN model of a random process is established, measured values of a subset of the random variables in U can be used to estimate the remaining ones. A random variable that is not measured, due to a partial observation process, is a *hidden node*. For example, in an unsolved crime, most of the nodes representing offender characteristics are hidden, while most of the nodes representing the evidence can be measured from the crime scene. Bayesian inference consists of computing the posterior probability mass function of hidden nodes based on the joint probability mass function $p(U)$ as well as sample information.

Sample information is *hard evidence* when it consists of samples of random variables in U and is *soft evidence* when it consists of partial information about a random variable in U. Partial information obtained from the measurements may not reveal the value of a discrete random variable but may narrow its range. In this case, the soft evidence consists of a *finding* e_i for a random variable X_i, where e_i is an array of size $1 \times m_i$, that is, a row vector, containing ones and zeros that indicate which values in the range of X_i are possible and which are ruled out with certainty by the measurements, respectively. For example, consider the random variable E_6 in Table 2 representing the place where the aggression began, with $m_6 = 14$ possible values, such as 1) the victim's home, 2) someone else's home, 3) a public place, and 4) the victim's professional studio or office. Even after the crime has taken place, the investigators may not be able to determine where the aggression began, but may be able to rule out locations 1) and 4) based on the evidence. Then, the finding on E_6 is a 1×14 array containing zeros for values 1 and 4, as well as ones for values 2, 3, and 5–14. If no evidence is available about X_i, then e_i is set equal to the $1 \times m_i$ ones vector, all of whose entries are 1, that is, $e_i = 1_{1\times m_i}$.

By expressing probability mass functions and findings as tables and arrays, the operations of multiplication $(\cdot)$ and marginalization (Σ) of probabilities can be carried out using the algebra of potentials [19, pp. 12–17], also known as the tabular method [22, pp. 93–94]. According to this method, given a two-dimensional joint probability table, that is, a matrix, $p(X_j, X_i)$, and a finding e_i on X_i, the two-dimensional joint probability table $p(X_j, X_i, e_i)$ is obtained by multiplying each row of $p(X_j, X_i)$ entrywise with e_i, that is, $p(X_j, X_i, e_i) = p(X_j, X_i) \cdot e_i$. When the joint probability mass function is factorized using the multiplication rule, $p(X_j, X_i) = p(X_j) \cdot p(X_i \mid X_j)$, the same result can be obtained by multiplying the CPT attached to X_i by e_i, that is, $p(X_j, X_i, e_i) = p(X_j) \cdot p(X_i \mid X_j) \cdot e_i$. To calculate the marginal probability mass function $p(X_i)$, the tabular method sums all of the entries corresponding to X_i in the joint probability table $p(X_j, X_i)$ according to the marginalization of probabilities [22, p. 93]. Other types of sample information, such as likelihood evidence, can be incorporated using Jeffrey's rule [23].

The posterior probability mass function of the hidden nodes in a BN can be obtained from the joint probability mass function (1) using Bayes' rule of inference as explained by the following theorem.



$p(E_{23} \vert Z_{12})$	$Z_{12} = z_{12,1}$	$Z_{12} = z_{12,2}$
$p(e_{23,1} \vert Z_{12})$	0.42	0.7
$p(e_{23,2} \vert Z_{12})$	0.58	0.3

FIGURE 3 Slice of the Bayesian network model of criminal behavior learned from data. The arcs, which are learned from the training set using the K2′ algorithm, represent directional conditioning among the criminal profiling variables. The conditional probability tables (CPTs) attached to every node are learned from the training set after the structure is fixed. As an example, a CPT is shown for the evidence node E_{23}, representing the strength of the relationships between this node and its parents (adapted from [29]).

Theorem

Let $B = (G, \Omega)$ be a BN over the universe $U = \{X_1, \ldots, X_n\}$, and let $F = \{e_1, \ldots, e_n\}$ be a set of findings on the variables in U. Then

$$p(U, F) = p(X_1, \ldots, X_n, e_1, \ldots, e_n) = \prod_{i=1}^{n} p(X_i \mid pa(X_i)) \cdot e_i, \tag{2}$$

where Π denotes repeated multiplication of probabilities, and the posterior probability mass function for $X_i \in U$ is

$$p(X_i \mid F) = \frac{1}{p(F)} \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{j_k=1}^{m_k} \cdot p\left(x_{1,j_1}, \ldots, x_{i-1,j_{i-1}}, X_i, x_{i+1,j_{i+1}}, \ldots, x_{n,j_n}, F\right). \tag{3}$$

For large BNs, efficient inference algorithms exploit Markov separation properties, which consist of conditional independence statements derived from the graph structure and the available evidence [21, Part I]. The random variables X_i and X_j are *conditionally independent* given the value of the random variable X_l if $p(X_i \mid X_l) = p(X_i \mid X_j, X_l)$. Using the factorization in (1) and the rules of probability calculus, it can be shown that a random variable X_i is conditionally independent of its nondescendants given its parents $pa(X_i)$. A node X_j is a descendant of X_i if there exists a forward path connecting X_i to X_j in the directed graph structure. Markov separation properties are used for graphical manipulations that simplify the factorization (1) represented by the original BN structure. The result is an approximate but efficient computation of the joint probability $p(U)$, which constitutes the main difficulty in obtaining the posterior probability mass functions required to infer the hidden nodes.

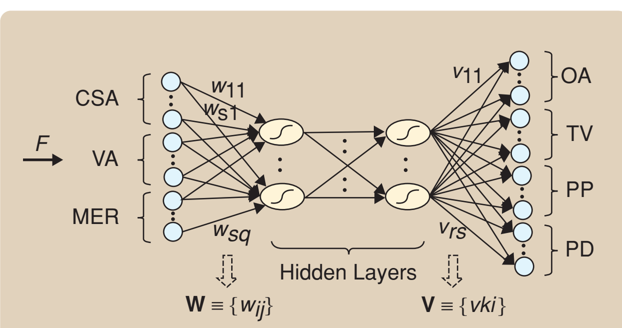


FIGURE 4 Structure of a neural network model of criminal behavior with multiple hidden layers containing s basis-function nodes, with q crime scene variables and r offender variables. The neural network input nodes correspond to the observable variables, comprising the crime scene analysis, the victimology, and the medical examiner report. The output nodes correspond to the offender nodes. Scalar adjustable parameters, referred to as weights, are attached to each arc, and are grouped into the matrices **W** and **V**. Once these parameters are computed through training, the NN model can be used to predict the offender's assessment, transactions with the victim, psychological profile, and psychiatric diagnosis, given the findings F obtained from the crime scene.

FEEDFORWARD NN MODELS

Unlike BNs, NN architectures are based on a division of the random variables in U into input and output variables. Like BNs, feedforward NNs provide a mathematical system representation based on observed data. A feedforward architecture consists of a layer of input nodes that represent input variables, one or more layers of nodes that represent basis functions, and a layer of output nodes that represent output variables. Starting with the input layer, all of the nodes in each layer are connected to all of the nodes in the next layer unidirectionally, forming a feedforward structure in which information about the inputs is transmitted forward to compute the output variables. The input variables in U comprise a vector $\mathbf{p}$, while the outputs in U comprise a vector $\mathbf{z}$. A feedforward NN with one layer of s basis functions represents the nonlinear transformation

$$\mathbf{z} = \mathbf{V}\sigma[\mathbf{W}\mathbf{p}], \tag{4}$$

where the matrices **W** and **V** contain the adjustable parameters of the NN model. The basis-function layer consists of an operator σ that takes the $s \times 1$ input vector $\mathbf{n} = \mathbf{W}\mathbf{p}$ and returns the $s \times 1$ vector output

$$\sigma[\mathbf{n}] \equiv [\sigma(n_1) \cdots \sigma(n_s)]^T.$$

$\sigma(\bullet)$ is a nonlinear function. A commonly used function is the exponential sigmoid $\sigma(n) \equiv (e^n - 1)/(e^n + 1)$.

A database can be used to train both BNs and NNs over the same universe U. However, once an NN model is obtained from data, its implementation consists of estimating the output variables from the values of the input variables. When only some of the input variables are observed, or when only the outputs are the observable variables, estimating the unknowns requires solving nonlinear systems of equations obtained by inverting (4). Once a BN model is obtained, evidence about any subset of variables in U can be used in (2)–(3) to infer all of the unknowns.

For criminal profiling, an NN model of criminal behavior is determined by grouping the set of observable variables $E = \{\text{CSA, VA, MER}\}$ into the input vector $\mathbf{p}$, and the set of offender variables $Z = \{\text{OA, TV, PP, PD}\}$ into the output vector $\mathbf{z}$. The number of basis-function layers and nodes is determined through testing. The weights that best match the database D are computed by a backpropagation training algorithm [24]. An illustrative neural model of criminal behavior is shown in Figure 4. When a new crime is investigated, the neural model is used to compute the offender characteristics Z from the values of the crime scene variables E, through the output equation (4). Evidence variables that are unavailable from the crime scene can be computed by inverting (4) numerically.

Comparison of BN and NN Models

Network models are a convenient tool for representing systems with many variables because they are inherently distributed (Figure 5) and are characterized by scalable algorithms, such as junction tree and backpropagation, that exploit the network architecture to simplify computation. The network architecture is a directed graph comprised of a set of nodes and a set of arcs that represent relationships between the nodes. In NNs, nodes that represent scalar basis functions are hidden. In the BN literature, the term hidden nodes refers to random variables that are not observable. In BNs arcs represent statistical correlations, whereas in NNs they represent functional relationships. While NN training algorithms usually produce architectures that are fully connected, the arcs in BNs are placed between only some of the nodes to capture conditioning. As a result, BN training produces hierarchical architectures that are sparsely connected but have many more layers than NNs.

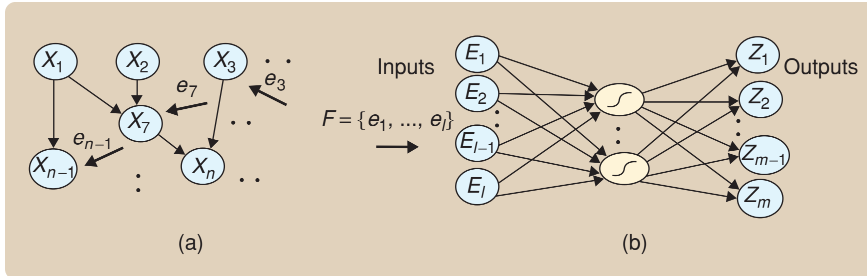


FIGURE 5 Comparison of (a) Bayesian and (b) neural network architectures. The nodes represent variables in the system, while arcs represent (a) directional conditioning and (b) functional relationships. In the case of neural networks, the hidden nodes contain nonlinear scalar functions of their inputs. The two architectures also differ as to where the evidence F or observations are injected into the network. In Bayesian networks evidence can be injected into any node, whereas in neural networks evidence is injected into the input variable nodes.

In network models, adjustable parameters are associated with arcs and nodes to quantify the functional or statistical relationships they represent. In BNs the parameters are organized in probability tables attached to the nodes. NN parameters consist of scalar numbers that are attached to arcs to represent scalar multiplication or to hidden nodes to represent an adjustable parameter of the basis function. The implementation of these network models for criminal profiling is discussed and compared in the following sections.

BEHAVIORAL NETWORK MODELS FOR AUTOMATED DECISION-SUPPORT SYSTEMS

The criminal profiling universe U_{cp} contains 57 discrete random variables that are partitioned into a set E containing 36 crime-scene variables and a set Z containing 21 offender variables. A sample of these variables and the cardinalities of their ranges are shown in Table 2, while the complete list is provided in [12] and [29]. For example, the range of a random variable Z_{14} representing the relationship between the offender and the victim includes values such as psychological subjection, psychological dominance, passionate love and jealousy, hate and envy, dependence, and ambivalent hate/love. The database D, which contains cleared homicide cases committed by various offenders, consists of complete observations of all of the random variables in U_{cp}.

BN Training with Cleared Homicide Cases

Feasible network models are obtained by designing BN and NN architectures that are consistent with the criminal profiling universe U_{cp} and its partition $\{E, Z\}$. The database D of cleared cases is partitioned into a training set T and a validation set V. BN models are trained in two steps, by first determining the arcs and then the parameters. The arcs are determined by maximizing the probability that a hypothesized structure $\hat{A}$ is compatible with the given database, that is, $p(\hat{A} \mid T)$. Since $p(T)$ is independent of $\hat{A}$, the joint probability $p(\hat{A}, T) = p(\hat{A} \mid T)p(T)$ can be maximized in place of $p(\hat{A} \mid T)$.

A tractable approximation of the joint probability $p(\hat{A}, T)$ can be obtained by assuming as in [26] that prior to training all models are equally likely; all cases in T occur independently, for example, there are no serial murderers; and all variables in U_{cp} can be ordered based on expert knowledge to allow arcs only in the forward path of the BN. With these assumptions, the joint probability can be factored into

$$p(\hat{A}, T) = p(\hat{A}) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(\bar{N}_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!\,, \tag{5}$$

where n is the number of variables in U_{cp}, r_i is the cardinality of the range of X_i, q_i is the number of unique instantiations of $pa(X_i)$, N_{ijk} is the number of cases in T with $X_i = X_{i,k}$, and $\bar{N}_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

An objective function or scoring metric for structural training is defined as a monotonically increasing function of the joint probability $p(\hat{A}, T)$, such as

$$J = \log \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(\bar{N}_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!\,. \tag{6}$$

Since prior to training all models are considered to be equally likely, the probability $p(\hat{A})$ is a known constant for any structure and can be removed from the scoring metric. Also, the node index i can be removed from (5) to consider only single-attribute scores based on the BN factorization property (1). The optimal BN structure A^*, which is the hypothesized arc structure that displays the highest

We construct neural network and Bayesian network models of criminal behavior from databases of single-victim homicides.

compatibility with the data T, is obtained by maximizing J with respect to $\hat{A}$ by means of the K2 greedy search algorithm described in [26].

The number of hypothetical structures increases exponentially with n [27], which affects the amount of data required to determine a reliable structure [28]. The modified K2′ algorithm given in [28] and [29] eliminates arcs between the evidence nodes by assuming that the evidence nodes are independent given the evidence, thereby reducing the search space and requiring less training data than the K2 algorithm. Once the BN structure is determined, the parameters Ω are obtained by means of a maximum likelihood estimation algorithm that maximizes their probability given the hypothesized architecture $\hat{A}$ and the training data T, that is,

$$\Omega^* = \arg\max_{\Omega}\{\log(p(\Omega \mid T, \hat{A}))\}\,.$$

In the maximum likelihood estimation algorithm, each entry of the CPT attached to node X_i is computed by normalizing the frequency of its values' observations in the training set T, for every observed combination of the parents' possible values [30]–[31].

Behavioral Network Analysis

A key application of behavioral network models is the development of scientific hypotheses based on relationships and patterns that emerge through the analysis of these models. A slice of a BN model obtained from a database T containing 200 homicide cases cleared by the U.K. police between the 1970s and the early 1990s is shown in Figure 3, taken from [29]. The arcs represent the most significant relationships among the nodes in U_{cp} that are learned from data. An illustrative CPT is provided in Figure 3 for node E_{23}, which represents a weapon obtained from the crime scene, and is strongly influenced by node Z_{12}, representing the gender of the offender. From this CPT it can be seen that female offenders ($Z_{12} = z_{12,2}$) are more likely to use a weapon obtained from the crime scene ($E_{23} = e_{23,1}$). Also, from the arcs connecting nodes Z_7, E_3, and Z_{17} (Figure 3), as well as from the corresponding CPTs (not shown, for brevity), it can be seen that young offenders ($Z_7 = z_{7,1}$) tend to have no prior record of fraud ($Z_{17} = z_{17,2}$), and to have a prior record of perpetrating foreign-object penetration ($E_3 = e_{3,1}$). These relationships are consistent with psychiatrists' understanding of young offenders' behavior. According to psychiatrists, offenders under the age of 21 are more likely to display psychopathic behavior and to be moved by extreme anger that may result in extremely violent acts, such as foreign-object penetration. As expected by criminologists, the BN model indicates that offenders with a prior record of theft (Z_1) are more likely to have a prior record of burglary (Z_2) and to be unemployed (Z_8), and that a strong relationship exists between the state of these priors and the gender of the offender (Z_{12}).

Although the interdependencies described above are known to psychiatrists and criminologists, several connections revealed by the network structure are considered new and interesting. For instance, the arc between the nodes representing a prior relationship between the victim and the offender (Z_{19}) and the victim being blindfolded (Z_{10}) can be given the following interpretation. Presumably, an offender blindfolds the victim to avoid eye contact and feelings of shame that may be brought about by the familiarity between them. Likewise, the relationship between the prior record of fraud (Z_{17}) and Z_{19} is new as is the connection between the offender employment and the victim suffocation (Z_{32}). A possible explanation is that an unemployed offender who is exasperated by his or her status and is reproached by the victim may employ suffocation in a reactive fury, especially if the victim attempts self-defense. Suffocation is believed to be a sign of a close relationship between the offender and the victim. But it may also occur when an intruder, such as a burglar, unexpectedly finds and kills a person who reacts to the intrusion. However, these last two relationships are not represented in the graph structure (Figure 3). The arcs between offender gender and nodes E_{23} and E_5, representing a weapon obtained from the crime scene and the victim's face being deliberately hidden, are consistent with the investigators' understanding of female offenders. Female offenders typically employ weapons that already are present in the home, such as kitchen knives, whereas male offenders, especially intruders, tend to bring their own weapon. Also, female offenders tend to cover the victim's face during the crime when the victim is a relative who spawns ambivalent feelings or after the crime to avoid guilt and remorse.

The graph structure of the BN behavioral model provides a map of correlations that are easily conveyed to researchers in the fields of criminal psychology and social sciences. Conversely, relationships hypothesized by psychologists and criminologists could be used to initialize the BN structure. Initialization consists of impeding or enforcing arcs between certain nodes based on expert knowledge and, by reducing the search space, can improve the effectiveness of the structural training algorithms.

TABLE 3 Example of offender profile variables inferred by the behavioral Bayesian network (BN) model, given the evidence from a homicide scene. The true value of each offender's characteristic is known from the validation set. The BN predictions are accompanied by a confidence level (CL), which represents the likelihood of the value predicted for each variable. Also, the evidence from the crime scene is shown to three teams of experts, labeled A, B, and C. Working independently, each of the three expert teams produces an offender profile that includes the example variables in this table, with an average accuracy ranging from 53% to 66%.

Offender Profile	First Homicide Case				
Variable	**True Value**	**BN (CL)**	**Expert Team A**	**Expert Team B**	**Expert Team C**
Z_1 = Prior theft	None	None (66%)	None	None	None
Z_2 = Prior burglary	No	No (80%)	None	Yes	Yes
Z_7 = Young age	No	No (83%)	No	Yes	Yes
Z_8 = Employment	Unemployed	Employed (50%)	Unemployed	Unemployed	Unemployed
Z_{11} = Familiarity CS	Yes	Yes (85%)	Yes	No	Yes
Z_{12} = Gender	Male	Male (74%)	Male	Male	Male
Z_{17} = Priors of fraud	None	None (67%)	None	None	None
Z_{10} = Prior victim/ offender relationship	None	Sexual (56%)	Nonsexual	Yes	Yes

Investigative Decision-Support Systems

In this section we show that a BN behavioral model can constitute a valuable decision-support tool for investigators. Using an approximate inference algorithm, the offender variables are inferred from the evidence obtained from the crime scene of an unsolved homicide. As a result, an approximate posterior probability mass function is computed for all hidden variables, that is, $p(Z_i \mid F)$ is obtained for all $Z_i \in U_{cp}$. The same algorithms can also be used to infer offender profiles from partial observations of the crime scene and forensic variables E. In a partial observation of E some variables may be hidden due to missing evidence or human error.

Two distinct homicide cases drawn from the validation set V are used to illustrate BN inference of offender profiles. Tables 3 and 4 show the true values of the offender characteristics along with the BN predictions. After $p(Z_i \mid F)$ is obtained from the inference algorithm, the offender variable Z_i is predicted by selecting the value with the highest posterior probability, that is, $\hat{Z}_i = z_i^* \equiv \text{argmax}_j \{p(z_{i,j}|F)\}$, where F is the set of findings obtained from the crime scene. The maximum posterior probability $p(z_i^* \mid F)$, known as the confidence level (CL), is shown in tables 3 and 4 for each offender variable. In the first homicide, six of the eight offender variables included in Table 3 are predicted correctly by the BN, while the two incorrectly predicted variables have a low confidence level (approximately 50%). In the second homicide (Table 4), seven of the eight variables are predicted correctly, but the incorrect prediction carries a fairly high confidence level of 81%. When we consider the complete offender profile comprised of 21 variables, approximately 80% of the variables are predicted correctly by the BN in both homicides.

For comparison, the evidence from the two homicides is also provided to three teams of experts. Team A has its main expertise in forensic psychiatry, while teams B and C are composed of criminologists and police investigators. Based on the evidence provided, Team A interprets the first homicide to be nonsexual and perpetrated by an offender related to the victim, such as a relative or

Table 4 Example of offender profile variables inferred by the behavioral Bayesian network (BN) model, given evidence from a second homicide scene. The true value of each offender's characteristic is known from the validation set. The BN predictions and confidence levels (CLs) are compared to the predictions of three independent teams of experts, labeled A, B, and C, who were given the same crime-scene evidence as the BN.

Offender Profile	Second Homicide Case				
Variable	**True Value**	**BN (CL)**	**Expert Team A**	**Expert Team B**	**Expert Team C**
Z_1 = Prior theft	None	None (75%)	Yes	Yes	Yes
Z_2 = Prior burglary	None	None (99%)	None	None	None
Z_7 = Young age	Yes	No (81%)	Yes	No	No
Z_8 = Employment	Unemployed	Unemployed (76%)	Unemployed	Unemployed	Unemployed
Z_{11} = Familiarity CS	Yes	Yes (90%)	Yes	No	No
Z_{12} = Gender	Female	Female (99%)	Female	Male	Female
Z_{17} = Priors of fraud	None	None (78%)	Yes	None	None
Z_{10} = Prior victim/ offender relationship	None	None (99%)	Nonsexual	None	None

We present an approach for deriving network models of criminal behavior that draws on knowledge-based systems as well as the fields of criminology and offender profiling.

personal acquaintance. The close relationship between the offender and the victim is evidenced by the type of wounds described, found mostly on the face, neck, and upper part of the body. The impulsive nature of the crime is demonstrated by the fact that no property is taken and the victim is left at the crime scene. The age of the offender is unclear, but he or she may have a history of violence or disorderly conduct and may be suffering from intermittent explosive behavior. The second homicide (Table 4) is believed to be nonsexual and premeditated, because of an attempt to poison the victim. Furthermore, familiarity between the offender and the victim is supported by the fact that the victim is suffocated. The subsequent stab wounds may reveal the fury of an offender who suffers from psychiatric problems, is easily angered, and has a past history of disorderly conduct and theft. The crime may have followed an argument over a valuable object. Although stabbing is generally perpetrated by male offenders, the use of poison leads the team to conclude that the offender may be female.

The offender profile presented by each team, which includes the variables illustrated in tables 3 and 4, achieves an average accuracy ranging from 53% to 66% depending on the team. Therefore, the BN predictions for these two homicides are on average more accurate than those presented by the experts. The BN model can provide accurate predictions and high confidence levels ($\geq$ 80%) even for variables that cause disagreement among the experts, such as the offender criminal record in Table 3 as well as the gender and crime scene familiarity in Table 4. Another interesting finding is that the performance of expert teams varies according to the group of variables and to the team's expertise. For example, the team specializing in forensic psychiatry makes better predictions regarding offender assessment (OA) and psychological profile (PP and PD) variables in both homicides. In contrast, teams B and C make better predictions regarding the transactions with the victim (VT) and the offender criminal record (OA). Since the BN model performance is independent of the variable taxonomy groups described in Table 1, the BN model can be used to complement the expertise of a team investigating a crime.

When the BN model is used to infer the offender profiles from 47 validation cases, it is found to predict 78.8% of all offender variables correctly, as shown in Table 5. When only variables with high confidence levels are taken into account, the average accuracy increases to 95.6%. Numerical simulations show that the percent confidence level of each hidden node Z_i is representative of

TABLE 5 Average percent accuracy of the offender variables predicted by the BN model, given the crime-scene evidence, organized by confidence level (CL). The average accuracy is obtained by dividing the number of correct predictions by the total number of hidden nodes over all validation cases. The results in each row illustrate how the ratio of the number of nodes predicted correctly over the total number of predictions increases with increasing CL. Investigators can utilize this decision support system by relying on those variables that are inferred with high confidence, while continuing to investigate the remaining ones.

Confidence Level (CL)	Correct Predictions/ Total Predictions (Number of Nodes)	Average Accuracy
$\geq$50%	780/987	79.03%
$\geq$60%	713/866	82.33%
$\geq$70%	618/725	85.24%
$\geq$80%	501/573	87.43%
$\geq$90%	244/255	95.6%

TABLE 6 Accuracy of a selected group of offender variables, averaged across the validation set. The accuracy of these predictions depends on the underlying statistical processes, on the quality of the network models, and on the evidence used for inference. These results show that, when everything else is equal, the Bayesian network model outperforms the neural network model for all but one of the variables' predictions.

	Average Accuracy (%)	
Inferred Offender Variable	Bayesian Network Model	Neural Network Model
Z_1 = Prior theft	57.4%	51.1%
Z_2 = Prior burglary	72.3%	65.9%
Z_7 = Young age	87.2%	76.6%
Z_8 = Employment	46.8%	53.2%
Z_{11} = Familiarity with crime scene	93.6%	85.1%
Z_{16} = Psychiatric condition	68.1%	53.2%
Z_{17} = Priors of fraud	72.3%	70.2%
Z_{10} = Prior victim/ offender relationship	98.9%	72.2%

its average percent accuracy, as illustrated by Table 5. For example, the average accuracy of offender variables with $CL \geq 0.9$ is 95.6%, whereas the average accuracy of variables with $CL \geq 0.7$ is 85.2%. Thus, in an unsolved homicide the confidence level indicates which predictions are reliable and can be used to narrow the list of suspects as well as which predictions remain uncertain and require further investigation. The percent accuracy and confidence level of each variable averaged across the validation set are illustrated in tables 6 and 7, respectively. These tables identify variables that display low confidence on average and, therefore, may be misleading during most investigations. For instance, low accuracy may be due to evidence that carries high errors or biases, or to a variable that is weakly related to the rest of the criminal profiling universe U_{cp}.

Figure 6 shows that NN models obtained from the same training set display a lower accuracy (72%, on average) than BNs. Architectures with larger hidden layers or with different definitions of the input and output vectors are found to display an even lower accuracy (as low as 52%). To understand the impact of the training set size on the performance of these network models, forward-sampling techniques are utilized to produce increasingly large databases. In Figure 6, the average percent accuracy of an NN model with three hidden layers and 321 hidden nodes is compared to the accuracy of a BN model trained with the same variable-size training set. For a fixed underlying joint probability mass function, the NN model displays worse performance than a BN model trained with the same training set. Furthermore, NNs do not provide a measure of confidence for their outputs and thus cannot inform investigators of the reliability of each prediction in an unsolved case. Unlike BNs, NNs require hard evidence for their inputs and complete cases for training. Their implementation may thus not be practical given the uncertainty surrounding criminal investigations. Although missing input data can be handled by solving nonlinear systems of equations, we found the inverse solutions to be unreliable and to deteriorate the performance by approximately 30% on average.

CHALLENGES AND OPPORTUNITIES

The results in tables 3–7 show that a useful BN model of criminal behavior can be obtained from a database of cleared single-victim homicide cases. Through approximate inference, this model can be used to predict the offender's background, psychological profile, and transactions with the victim, given evidence from the crime scene. For each variable inferred by the BN, investigators obtain a confidence level that represents the probability of each prediction and that is indicative of its accuracy. The inference of CP variables that display low predictive accuracy on average can be improved by identifying a better CP universe, as well as by collecting observations void of significant errors or biases. Police databases often contain incomplete cases in which some of the observations are missing due to difficulties encountered during the investigations and interrogations. The expectation-maximization (EM) parameter-training algorithm [32],

TABLE 7 Offender nodes grouped by average confidence level, with a few examples provided in the third column. The employment state of the offender tends to be inferred with low confidence, and is accompanied by low accuracy (Table 6), whereas the offender's familiarity with the crime scene and his or her prior relationships with the victim are typically predicted with fairly high confidence and accuracy. If expert detectives and psychologists believe that variables in the lower ranges are crucial to the investigative process, the Bayesian network model can be modified by identifying a better set of criminal profiling variables, or by improving the training database.

Confidence Level Range	Number of Nodes	Offender Variables (Examples)
$CL < 50\%$	1	Z_8 = Employment
$50\% \leq CL < 60\%$	1	Z_1 = Prior theft
$60\% \leq CL < 70\%$	3	Z_{16} = Psychiatric condition
$70\% \leq CL < 80\%$	5	Z_2 = Prior burglary Z_{17} = Priors of fraud
$80\% \leq CL < 90\%$	6	Z_7 = Young age Z_{12} = Gender
$90\% \leq CL < 100\%$	5	Z_{10} = Prior victim/offender relationship Z_{11} = Familiarity with crime scene

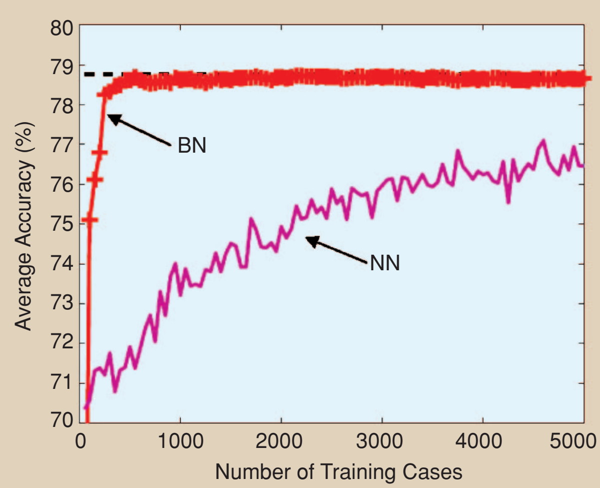


FIGURE 6 Comparison of Bayesian network and neural network predictive performance as a function of training set size. These network models are trained and validated with synthetic data generated by using forward sampling of a joint probability mass function obtained from real data. This numerical study shows that, in addition to providing desirable features such as confidence levels, Bayesian network models require less data than neural networks to meet their maximum predictive performance (shown by the dashed line).

[33] can be used to include all available cleared cases in the training set, despite the missing data. Typically, BN algorithms are based on the assumption that the data are identically sampled, but this assumption is often violated during criminal investigations. Hence, more systematic data-collection procedures are required to minimize errors due to prejudices and biases, to improve the completeness of training databases.

The use of automated criminal profiling software can potentially reduce the number of suspects in a given case and shorten the investigation times. Ideally, the model is continually updated by insertions from newly cleared crimes using incremental training algorithms. A comprehensive BN model of criminal behavior can also contribute to our understanding of crime and offender typologies. By inspecting the BN structure learned from data, it is possible to discover previously unknown links between the variables. Sensitivity analysis can be valuable for testing psychological hypotheses as well as for assessing offender profiles obtained from evidence compromised by investigative errors or biases, as caused by human emotions and prejudices.

The validity of present-day profiling may be questionable, and has raised numerous critiques due to the blurred borderlines between instinct and intuition [34]. Also, the procedures adopted by human profilers are known to possibly compromise the entire process, from the collection of evidence to the offender assessment [34]. The presence of cultural baggage passed on from one profiler to another with little exposure to actual crime scenes, the stress on psychodynamics, and the poor application of in-depth psychology offer little information about real patterns of criminal behavior. National data for 2002 report 23 million cases of victimization in the United States, 5.3 million of which were personal victimizations. The national clearance rate for violent crimes was as follows: murder, 64% (16,204 cases); aggravated assault, 56.5% (894,348 cases); and rape, 44.5% (95,136 cases). Including indexed property crimes in this data, which often degenerate into rape or murder, the clearance rate for burglary was 13.0% (2.2 million reported cases) and for arson was 16.5% (74,921 cases) [2]. These low clearance rates indicate that in spite of attentive police investigation, a large number of offenders are not apprehended and continue their criminal activities. Because of the high number of offenses it may prove difficult for the police to carry out a thorough case analysis. In consideration of the limitations of existing profiling techniques and clearance rates, BNs can contribute greatly to the field of criminology and law enforcement.

CONCLUSIONS

We present an approach for deriving network models of criminal behavior that draws on knowledge-based systems as well as the fields of criminology and offender profiling. A behavioral model provides a mathematical representation of the multidimensional interdependencies between variables that determine or reflect offender behavior at the crime scene. Once a valid network model is obtained from a database of cleared cases, it can be used for decision support in an unsolved case by inferring the offender profile from the forensic evidence obtained from the crime scene. More research is needed to determine which computational attributes should be provided to criminologists and psychologists interested in developing novel profiling theories and techniques. For example, providing confidence levels that denote the probability that the predicted variables are correct can be valuable in narrowing the list of suspects, because the variables with the highest confidence can be given priority over the remaining variables. Also, sensitivity analysis and structural training algorithms can be used to identify the most significant relationships among the variables and to determine how sensitive the linked variables are to each other. By presenting this collaborative research and discussing the challenges and opportunities of the profiling application, we hope to motivate further interdisciplinary research on behavioral models and dynamics.

ACKNOWLEDGMENTS

The authors wish to thank Antony Pinizzotto and Gabrielle Salfati for their suggestions and continuing contributions to this study.

REFERENCES

[1] M. Strano, "A neural network applied to criminal psychological profiling: An Italian initiative," *Int. J. Offender Therapy Comparative Criminology*, vol. 48, no. 4, pp. 495–503, 2004.

[2] Bureau of Justice Statistics, *Crime Data 2003* [Online]. Available: http://www.fbi.gov/ucr/ucr.htm

[3] S.A. Egger, "Psychological profiling: Past, present, and future," *J. Contemporary Criminal Justice*, vol. 15, no. 3, pp. 242–261, 1999.

[4] M. Rogers, "The role of criminal profiling in the computer forensic process," *Computers Securities*, vol. 22, no. 4, pp. 292–298, 2003.

[5] B. Turvey, *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*. New York: Academic, 1999.

[6] D. Canter, *The Psychology of Place*. London: Architectural Press, 1977.

[7] D. Canter and R. Heritage, "A multivariate model of sexual offense behavior: Developments in offender profiling," *J. Forensic Psychiatry*, vol. 1, no. 2, pp. 185–212, 1989.

[8] D. Canter, *Criminal Shadows*. New York: HarperCollins, 1994.

[9] R.K. Ressler, A. Burgess, and J.E. Douglas, *Sexual Homicide: Patterns and Motives*. New York: Lexington Books, 1988.

[10] R.N. Kocsis, R.W. Cooksey, and J.I. Harvey "Psychological profiling of sexual murders: An empirical model," *Int. J. Offender Therapy and Comparative Criminology*, vol. 46, no. 5, pp. 532–554, 2002.

[11] P. Santtila, H. Häkkänen, D. Canter, and T. Elfgren, "Classifying homicide offenders and predicting their characteristics from crime scene behavior," *Scandinavian J. Psychol.*, vol. 44, pp. 107–118, 2003.

[12] C.G. Salfati, "Offender interaction with victims in homicide: A multidimensional analysis of frequencies in crime scene behavior," *J. Interpersonal Violence*, vol. 18, no. 5, pp. 490–512, 2003.

[13] D.K. Rossmo, *Geographic Profiling*. Boca Raton, FL: CRC, 2000.

[14] G.B. Palermo and R.N. Kocsis, *Offender Profiling: An Introduction to the Sociopsychological Analysis of Violent Crime*. Springfield, IL: Charles C. Thomas, 2004.

[15] L. Alison, C. Bennell, and D. Ormerod, "The personality paradox in offender profiling: A theoretical review of the process involved in deriving background characteristics from crime scene actions," *Psychol., Public Policy, and Law*. vol. 8, no. 1, pp. 115–135, 2002.

[16] B.E. Turvey, "The impressions of a man: An objective forensic guideline to profiling violent serial sex offenders," 1997 [Online]. Available: http://www.corpus-delicti.com
[17] B.E. Turvey, *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*. San Diego, CA: Academic, 1999.
[18] B. Snook, D. Canter, and C. Bennell, "Predicting the home location of serial offenders: A preliminary comparison of the accuracy of human judges with a geographic profiling system," *Behav. Sci. Law*, vol. 20, pp. 109–118, 2002.
[19] F.V. Jensen, *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag, 2001.
[20] M.I. Jordan and T.J. Sejnowski, Eds., *Graphical Models: Foundations of Neural Computation*. Cambridge, MA: MIT Press, 2001.
[21] M.I. Jordan, Ed., *Learning in Graphical Models*. Cambridge, MA: MIT Press, 1998.
[22] D.P. Bertsekas and J.N. Tsitsiklis, *Introduction to Probability*. Boston, MA: Athena Scientific, 2002.
[23] J. Pearl, "Jeffrey's rule, passage of experience, and neo-Bayesianism," *Knowledge Representation and Defeasible Reasoning*, H.E. Kyburg, R.P. Loui, and G.N. Carlson, Eds. Norwell, MA: Kluwer, 1990, pp. 245–265.
[24] M. Reidmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in *Proc. IEEE Int. Conf. Neural Networks (ICNN)*, San Francisco, CA, 1993, pp. 586–591.
[25] M. Strano and R. Bruzzone, *Telematic Journal of Clinical Criminology* [Online]. Available: http://www.criminologia.org/index.html
[26] G. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Mach. Learn.*, vol. 9, pp. 309–347, 1992.
[27] R.W. Robinson, "Counting unlabeled acyclic digraphs," *Combinatorial Mathematics*, C.H.C. Little, Ed. Berlin, Germany: Springer-Verlag, pp. 28–43, 1977.
[28] K.A.C. Baumgartner, S. Ferrari, and G.B. Palermo, "Constructing Bayesian networks for criminal profiling from limited data," *Knowledge-Based Systems* [Online]. Available: http://dx.doi.org/10.1016/j.knosys.2008.03.019
[29] K.A.C. Baumgartner, S. Ferrari, and C.G. Salfati, "Bayesian network modeling of offender behavior for criminal profiling," in *Proc. Conf. Decision and Control*, Seville, Spain, 2005, pp. 2702–2709.
[30] D. Heckerman, D. Geiger, and D. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Mach. Learn.*, vol. 20, pp. 197–243, 1995.
[31] S.-Z. Zhang, H. Yu, H. Ding, N.-H. Yang, and X.-K. Wang, "An application of online learning algorithm for bayesian network parameter," in *Proc. Int. Conf. Machine Learning and Cybernetics*. vol. 1, 2003, pp. 153–156.
[32] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, ser. B, vol. 39, pp. 1–22, 1977.
[33] R.J.A. Little and D.B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. New York: Wiley, 2002.
[34] R.N. Kocsis and G.B. Palermo, "Ten major problems with criminal profiling," *Amer. J. Forensic Psychol.*, vol. 26, no. 2, pp. 45–67, 2005.

AUTHOR INFORMATION

Silvia Ferrari (sferrari@duke.edu) is an assistant professor of mechanical engineering and materials science at Duke University, where she directs the Laboratory for Intelligent Systems and Controls. Her research interests include robust adaptive control of aircraft, learning and approximate dynamic programming, and distributed sensor management. She received the B.S. from Embry-Riddle Aeronautical University and the M.A. and Ph.D. from Princeton University. She is a member of IEEE, ASME, SPIE, and AIAA. She is the recipient of the ONR young investigator award (2004), the NSF CAREER award (2005), and the PECASE award (2006). She can be contacted at Duke University, Department of Mechanical Engineering and Materials Science, Box 90300, Durham, NC 27708-0005 USA.

Kelli C. Baumgartner received the B.S. in aerospace engineering from Embry-Riddle Aeronautical University. She received the M.S. and Ph.D. from Duke University. Her research interests include Bayesian networks, distributed systems, and launch vehicle flight design and control. She is a member of IEEE, ASME, and AIAA. She is currently a researcher in the mission analysis branch at Analex Corporation, a subsidiary of QinetiQ North America, Brook Park, Ohio.

George Palermo is a clinical professor of psychiatry at the Medical College of Wisconsin and has published numerous books and articles on violence, forensic psychiatry, criminology, and criminal profiling. He is an adjunct professor of criminology and law studies at Marquette University in Milwaukee and the director of the Center for Forensic Psychiatry and Risk Assessment in Milwaukee. He is the editor-in-chief of the *International Journal of Offender Therapy and Comparative Criminology*, a member of the Board of Directors of the International Academy of Law and Mental Health, and a member of various national and international psychiatric and criminological organizations. In 2002, he received the Lex et Justitia award from the Department of Criminology and Law Studies, Marquette University. He has also received several awards in Italy and is Commendatore of the Italian Republic.

Marco Strano is a psychologist and criminologist for the Italian State Police on active service by the Center of Neurology and Medical Psychology, where he carries out research activities focused on psychological support investigation tools. He is the president of the International Crime Analysis Association, and develops research and academic education activities for the Psychiatry and Psychology Institute of the Catholic University of Rome. He is a member of the American Society of Criminology, the American Psychological Association, the New York Academy of Sciences, the International Association of Crime Analysts, and the International Academy of Law and Mental Health. He is the author of numerous scientific publications on forensic psychology and criminology. His research interests are criminal profiling, crime prevention, intuitive policing, and investigative psychology.

Roberta Bruzzone is a psychologist and criminologist, specializing in psychodiagnostics and criminal and forensic psychology. She is the vice president of the International Crime Analysis Association (ICAA) and technical editor of the *Telematic Journal of Clinical Criminology*. She has participated in training exercises at Italian universities and the University of California and teaches various training courses on criminological and information technology at the Italian State Police Academies. She is a member of the Scientific Board of the Italian State Communication Police, the International Association of Crime Analysts, and the International Academy of Law and Mental Health. She is the coordinator of the ICAA Criminal Profiling Lab, whose principal research interests are criminal profiling, crime prevention, psychological aspects of criminal investigation, intuitive policing, and investigative psychology.