Knowledge-Based Systems 21 (2008) 563–572

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Constructing Bayesian networks for criminal profiling from limited data

K. Baumgartner [a,1], S. Ferrari [a,*,2], G. Palermo [b,3]

[a] Pratt School of Engineering, Duke University, P.O. Box 90300, 176, Hudson Hall, Research Drive, Durham, NC 27708-0005, USA
[b] Psychiatry and Neurology Department, Medical College of Wisconsin, Milwaukee, WI 53226, USA

ARTICLE INFO

Article history:
Received 21 March 2007
Accepted 21 March 2008
Available online 30 March 2008

Keywords:
Criminal profiling
Crime analysis
Automation
Bayesian networks
Performance metrics

ABSTRACT

The increased availability of information technologies has enabled law enforcement agencies to compile databases with detailed information about major felonies. Machine learning techniques can utilize these databases to produce decision-aid tools to support police investigations. This paper presents a methodology for obtaining a Bayesian network (BN) model of offender behavior from a database of cleared homicides. The BN can infer the characteristics of an unknown offender from the crime scene evidence, and help narrow the list of suspects in an unsolved homicide. Our research shows that 80% of offender characteristics are predicted correctly on average in new single-victim homicides, and when confidence levels are taken into account this accuracy increases to 95.6%.

1. Introduction

The study of criminal behavior for the purpose of identifying the characteristics of an unknown offender and the motivation for the crime is commonly known as *criminal profiling*. In current practice, criminal profiling relies primarily on the personal experience of criminal investigators and forensic psychologists, rather than on empirical scientific methods [31]. As such, it may be subject to errors caused by cultural biases and misinterpretation [24,31,32,43]. After clearing a criminal case, investigators file the background characteristics and psychological diagnosis of the convicted offender together with the forensic evidence obtained from the crime scene. With the increased availability of computer and information technologies, law enforcement agencies have been able to compile databases with detailed offender and crime scene information from major felonies, such as murder, rape, and arson. Consequently, important authors have advocated that machine learning techniques will play a significant role in developing decision-aid tools for police investigations [4,17,27,32,42]. The most significant contributions to date have been recently reviewed in [17]. Rule-based systems have been proposed in [4] for knowledge acquisition from a database with modus operandi information. Research on inductive profiling has employed statistical analysis to classify offender behavior into categories or *dichotomies*, based on the crime scene evidence [12,25,34,35,37,39,41]. While this research has been successfully implemented to predict the approximate residence location of serial homicide offenders [35], it has been unable to identify psycho-behavioral offender profiles in single-victim (non-serial) homicides. This shortcoming has been attributed to the complexity of human behavior and to the large number of relevant variables, both of which limit the applicability of behavior classification techniques [2,31,32].

In this paper, a novel Bayesian network (BN) approach to criminal profiling is presented. The approach consists of learning a BN model of offender behavior from data and, subsequently, implementing the model for profiling by means of an inference engine. The database used in this paper is similar to the modus operandi database described in [4]. However, the BN approach is not limited by decisive "if-then" relationships, because it views the relationships among all variables as probabilistic. Unlike inductive profiling, the BN approach does not require to postulate behavior categories *a priori* and, consequently, it is capable of identifying psycho-behavioral profiles in single-victim single-offender homicides (Section 6). Also, the inferred offender characteristics include confidence levels that represent their expected accuracy. Thus, when provided with a BN profile, the police can easily establish what are the reliable predictions in the investigated case.

Implementing BN models for inference has proven valuable in many applications, including medical diagnosis, economic forecasting, biological networks, and football predictions [1,19,20,23,30]. This literature shows that the effectiveness of BN inference and prediction is highly dependent on the sufficiency of the training database. While various approaches have been proposed for dealing with insufficient databases [11,14–16,21,26,30,40], there are no general guidelines for establishing whether a given database

* Corresponding author. Tel.: +1 919 660 5484; fax: +1 919 660 8963.
E-mail address: sferrari@duke.edu (S. Ferrari).

[1] K. Baumgartner is a graduate student in the Pratt School of Engineering.
[2] S. Ferrari is with Faculty of Mechanical Engineering and Material Science and the Faculty of Electrical and Computer Engineering.
[3] G. Palermo is with Faculty of Psychiatry and Neurology.

0950-7051/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2008.03.019

is insufficient. In [45], it was shown that the size of a sufficient database depends on the number of variables, their domain, and the underlying probability distributions. But, while the variables and the domain definitions are known from the problem formulation, the underlying probability distribution is often unknown *a priori*. This paper presents a set of performance metrics that can be used to determine the sufficiency of an available database without knowledge of the underlying joint probability distributions (Section 4). Although a police database may include hundreds of cleared cases, they may still be insufficient to train a BN model due to the large number of relevant variables, and to the complexity of their relationships [3]. Therefore, in Section 5 these performance metrics are implemented to determine the size of a sufficient database with single-victim single-offender homicides. Subsequently, a BN model is trained using a newly modified K2′ algorithm that improves performance once the database size is fixed (Section 5). In Section 6, the trained BN model is applied to infer the characteristics of unknown offenders from the crime scene evidence. The results show that when the confidence level is taken into account, the average accuracy of the BN predictions is 95.6%. For comparison, the evidence from two homicide cases has been presented to a team of expert criminologists. Based on the evidence alone, the experts predict 53% of all offender variables correctly. Whereas, in the same two cases the BN predicts 86% of all offender variables correctly, and displays 80% average accuracy in 1000 other homicide cases. Also, offender characteristics that cause disagreement among the experts are predicted correctly and with a high confidence level by the BN. Finally, the structure of the BN model indicates what are the most significant relationships among the variables and, thus, it could be used for the scientific development of hypothesis on criminal psychology.

2. Background on Bayesian network inference and training

A Bayesian network (BN) approximates the joint probability distribution for a multivariate system based on expert knowledge and sampled observations that are assimilated through training [18,22]. A BN consists of a *directed acyclic graph* (DAG) and an attached parameter structure comprised of *conditional probability tables* (CPTs) that together specify a joint probability distribution [22]. The DAG $\Lambda = \{\mathscr{X}, \mathscr{S}\}$ is composed of a set of directed arcs $\mathscr{S}$ that represent the dependencies among a set of variables or nodes $\mathscr{X} = \{X_1, \ldots, X_n\}$ known as *universe*, such that $\mathscr{S} = \{(X_i, X_j) | X_i, X_j \in \mathscr{X}, X_i \neq X_j, j > i\}$. A node X_i represents an event, proposition, or mathematical quantity that has a finite number of mutually exclusive instantiations (denoted by lower case letters), and is said to be in its j^{th} instantiation when $X_i = x_{i,j}$. $\Theta = \{\theta_1, \ldots, \theta_n\}$ is the parameter structure that is attached to Λ, where θ_i is the conditional probability $p(X_i|\pi_i)$ attached to node X_i, and the set π_i represents the immediate parents of X_i.

In this research, the nodes $\mathscr{X}$ and their instantiations are defined by criminologists and psychologists. The BN arcs $\mathscr{S}$ and parameters Θ are learned from the database $\mathscr{T}$ in this sequence. Structural training determines the set of arcs that "best" describes the database by considering all possible arcs between the nodes. The compatibility of each hypothesized structure with the training data is assessed by a scoring metric that approximates the conditional probability of $\mathscr{S}$ given $\mathscr{T}$, $p(\mathscr{S}|\mathscr{T})$ [18]. Since $p(\mathscr{T})$ is independent of $\mathscr{S}$, the joint probability $p(\mathscr{S}, \mathscr{T})$ can be maximized in place of $p(\mathscr{S}|\mathscr{T})$. A tractable scoring metric, known as *K*2, is obtained from $p(\mathscr{S}, \mathscr{T})$ using the assumptions in [6], which include fixed ordering of variables in $\mathscr{X}$:

$$\mathscr{G} = \log\left(\prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(\overline{N}_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!\right), \tag{1}$$

where r_i is the number of possible instantiations of X_i, and q_i is the number of unique instantiations of π_i. N_{ijk} is the number of cases in $\mathscr{T}$ in which $X_i = x_{i,k}$, and $\overline{N}_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. Then, the BN structure that displays the highest compatibility with the data is sought by maximizing (1). Subsequently, the structure is held fixed, and the CPTs are computed by the Maximum Likelihood Estimation algorithm (MLE) (reviewed in [8,33]).

The BN (Λ, Θ) represents a factorization of the joint probability over a discrete sample space,

$$p(\mathscr{X}) = p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i|\pi_i), \tag{2}$$

for which all probabilities on the right-hand side are given by the CPTs. Therefore, when a variable X_i is unknown or *hidden*, Bayes' rule of inference can be used to calculate the posterior probability distribution of X_i given evidence of the set of l variables, $\mu_i \subset \mathscr{X}$, that are conditionally dependent on X_i,

$$p(X_i|\bar{\mu}_i) = \frac{p(\bar{\mu}_i|X_i)p(X_i)}{p(\bar{\mu}_i)}, \tag{3}$$

where $p(X_i)$ is the prior probability of X_i. The likelihood function is factored as $p(\bar{\mu}_i|X_i) = \prod_j p(\bar{\mu}_{i(j)}|X_i)$, where $\bar{\mu}_{i(j)}$ is the evidence of the j^{th} variable in μ_i. The marginalization required to obtain $p(\bar{\mu}_i)$ is simplified using (2):

$$p(\bar{\mu}_i) = \sum_{i=1}^{n} p(\bar{\mu}_i, X_i) = \sum_{k=1}^{r_i} p(X_i = x_{i,k}) \prod_{j=1}^{l} p(\bar{\mu}_{i(j)}|X_i), \tag{4}$$

The posterior probability in (3) is used to obtain the prediction $\overline{X_i} = \arg\max_k p(X_i = x_{i,k}|\bar{\mu}_i)$, and its posterior probability is the *confidence level* of the prediction. Furthermore, by identifying conditional independencies among nodes from the so-called *Markov separation properties*, inference of hidden variables can be completed efficiently even in large networks [9].

Bayesian networks are particularly well suited to criminal profiling because they learn from data, and utilize the experience of criminologists in selecting the nodes and node ordering. The confidence levels provided for the offender profile inform detectives of the likely accuracy of each prediction. In addition, the graphical structure of the BN represents the most significant relationships between offender behavior and crime scene actions, which may be useful in developing new scientific hypothesis on criminal behavior.

3. Bayesian network approach to criminal profiling

This research develops an approach for obtaining a BN model of criminal behavior that (1) captures the most significant relationships among the relevant criminal profiling variables, and (2) is used to predict the profile of an unknown offender given evidence from the crime scene. The methodology consists of using expert knowledge to define the BN universe, and the fixed node ordering for structural training (as shown in Section 2). The universe $\mathscr{X}$ consists of 57 binary variables that have been identified as relevant to the criminal process by criminal investigators and forensic psychologists. A sample of these variables is illustrated in Table 1, and the complete list is shown in [38]. Each variable $X_i \in \mathscr{X}$ is binary, and represents a characteristic or event that is either present or absent at the crime. $\mathscr{X}$ is partitioned into set $\mathscr{E} = \{E_1, \ldots, E_k\}$ containing $k = 36$ *evidence variables* that are observable from the crime scene, and set $\mathscr{Y} = \{Y_1, \ldots, Y_m\}$ containing $m = 21$ *offender variables* that characterize the offender and, thus, are unknown or *hidden* at the crime scene.

The BN model structure, $\mathscr{S}$, and parameters, Θ, are learned using a police database of cleared single-victim single-offender homicides, $\mathscr{D} = \{C_1, \ldots, C_d\}$. Each case C_i is a complete observation