Forensic Chemistry 3 (2017) 41–51

Contents lists available at ScienceDirect

Forensic Chemistry

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

Full length articles

Investigative probabilistic inferences of smokeless powder manufacturers utilizing a Bayesian network

Dana-Marie K. Dennis [b], Mary R. Williams [b], Michael E. Sigman [a,b,*]

[a] Department of Chemistry, University of Central Florida, Orlando, FL 32816-2366, United States
[b] National Center for Forensic Science at the University of Central Florida, Orlando, FL 32826, United States

ARTICLE INFO

Article history:
Received 30 August 2016
Received in revised form 2 December 2016
Accepted 5 December 2016
Available online 26 January 2017

Keywords:
Smokeless powders
Bayesian networks
Prior probabilities
Posterior probabilities
Conditional independence

ABSTRACT

Chemical and physical characteristics for 169 smokeless reloading powders were utilized in the development of a Bayesian network for inference of the powder manufacturer. The chemical characteristics of the smokeless powders were encoded using the most intense ions in their total ion spectra from gas chromatography-electron ionization-mass spectrometry (GC-EI-MS), which were previously determined from agglomerative hierarchical cluster analysis. Physical characteristics included as network nodes were the average kernel diameter and length, shape, color, luster, absence/presence of a bias cut and absence/presence of a perforation, which are commonly considered in casework. A Bayesian network was compiled using R code, written in-house. Performance of the network was validated by 100 iterations of stratified cross validation, withholding 10% of the data for testing and using the remaining 90% of the data to develop probability tables for the network. Posterior probabilities of the powder manufacturers were calculated for each test sample, and manufacturer inferences were made based on the highest posterior probability. The sensitivity and specificity of the fully instantiated network was examined for each manufacturer. Other performance metrics, including the positive and negative predictive values (PPV and NPV), which take into account the prevalence of each manufacturer, were also examined. The PPV ranged from 0.59 to 0.81 for individual manufacturers when all nodes of the network were instantiated. The NPV for fully instantiated networks ranged from 0.82 to 0.99 for individual manufacturers.

1. Introduction

Smokeless powders are low explosive propellants which undergo decomposition by deflagration [1,2], and are categorized as single base, double base or triple base, as defined by their primary energetic materials [3]. Smokeless powders also contain additional organic compounds that function as plasticizers, deterrents, stabilizers, flash suppressants, opacifiers, and dyes [3–5]. The powders have a range of physical properties, including kernel shape, presence or absence of a perforation, color, kernel size, and in some cases colored markers are added to the powder. The manufacturer's aim is to optimize the ballistic performance of the product [5–9].

Single and double base powders are available for civilian purchase in the United States from sporting goods and internet retailers, and are typically used to manually reload ammunition. Commercial availability of smokeless powders contributes to their use in improvised explosive devices (IEDs) such as pipe bombs [10]. In the forensic sphere, the goal of smokeless powder analysis is to identify particles as smokeless powders, and, if possible, to determine the product identity and/or manufacturer of the smokeless powder [2]. Identification of some brands (products) has been reported if the colored markers are present as a physical attribute in the smokeless powder [2,8]; however, identification of brands using markers could be misleading since the marker color may vary between lots [8]. Furthermore, brand identification becomes increasingly difficult in the absence of colored markers, if the physical features and chemical formulation of a product has been altered between production lots, or by a combination of these factors.

Laboratory protocols for the analysis of intact (pre-blast) smokeless powders commonly include stereomicroscopy [2], Fourier transform infrared spectroscopy (FT-IR) [11,12], and gas chromatography – electron ionization – mass spectrometry (GC-EI-MS) [13]. The results obtained from these analyses combined with notation of the powder's physical attributes (shape, dimensions, color, etc.) are often used in a database search with the goal of identifying the powder's brand, or to generate a short list of poten-

* Corresponding author at: National Center for Forensic Science at the University of Central Florida, P.O. Box 162367, Orlando, FL 32826, United States.
E-mail address: michael.sigman@ucf.edu (M.E. Sigman).

http://dx.doi.org/10.1016/j.forc.2016.12.001
2468-1709/

tial products [2,8]. Product identification from a database search is limited if the database content is neither exhaustive nor dynamic. Smokeless powders industrial practices include altering product formulation and morphology to improve ballistic performance, as well as purchasing domestic and imported products or reworking surplus military powders for repackaging and retail [8,14]. These practices reinforce the need for continual purchase and analysis of smokeless powder products to ensure that database records accurately reflect product availability in the market. Perhaps the most pertinent forensic question arises when a database search has been conducted, and assertions are to be made regarding the product's identity. The question addresses whether a recovered sample and a control sample are the same or different products. Concluding that the recovered and control samples are different brands would be reached when either, or both, the chemical formulations or the physical attributes of the powders are different; however, when the two samples share the same physical and chemical characteristics, the problem is more challenging. Categorical rather than probabilistic statements are typically made as to the product's identity, although the general need for statistically based assertions in forensic science has been addressed [15]. It is generally held that categorical statements should be avoided in favor of statements of evidentiary value in all areas of forensic science; however, at the time of this writing, categorical statements remain accepted in many areas of forensic science in the United States. Our ability to replace categorical statements is dependent on continued research and collaboration between the academic and practicing forensic communities. Specifically in the area of smokeless powder analysis, a recent study which was conducted using data for a set of intact smokeless powder database samples demonstrated the limited evidentiary value that could be expressed in probabilistic assertions of same product or different product, based on chemical and physical characteristics [10]. In this work we examine the use of a Bayesian network to make a probabilistic inference regarding the manufacturer of a smokeless powder. Inference of the manufacturer, rather than product/brand, was chosen to limit the size of the probability tables associated with each node in the network, and to allow for more reliable estimates of the probability tables from a relatively small number of recently purchased powders where the manufacturer is definitely known. The use of Bayesian networks also allows for calculation of the posterior probability of each manufacturer, even when only part of the chemical and physical parameters for a powder are available. The potential absence of some chemical or physical characteristics for smokeless powders in casework samples was an important factor in the choice of a Bayesian network model for this work. In the models presented here, as with other possible models, the quality of the calculated posterior probabilities is dependent upon having a good estimate of the relevant population when creating the model. The Bayesian network models presented here are tested by cross validation with a random stratified draw from the database samples. This ensures that the test set is representative of the population used to make the models and consequently, the results can be considered as representing a best case scenario. Estimation of the distributions of relevant populations for forensic applications continues to be a challenge to advances in many areas of forensic science. This work does not address the problem of estimating manufacturer prevalence in a casework relevant population.

Bayesian networks are probabilistic graphical structures comprised of nodes and edges. Within the network structure, the nodes, which are depicted by circles, represent random variables or events. The edges are shown as arrows to denote conditional relationships between the nodes, resulting in a directed acyclic graph (DAG). A node located at the arrow's tail is referred to as a parent node, a node at the arrow's head is referred to as a child node, and a node without an arrow pointing into it is known as a root node. Fig. 1 illustrates a simple Bayesian network structure comprising four nodes: **A**, **B**, **C**, and **D**. Nodes **A** and **C** are parent nodes; **A** is the parent of **B** and **C**, and **C** is the parent of **D**. Node **A** is also a root node, since there is no arrow directed into the node.

Within the network, each node represents either a discrete or continuous random variable and contains an exhaustive list of mutually exclusive states [16]. States represent possible outcomes for the random variable, and each state has a probability value ascribed to it. A table associated with each node encodes the probability distribution across all states, or combination of states, within the node. A root node's table encodes unconditional probabilities for the states within that node, and the table associated with a child node encodes conditional probabilities for all states within the child node, where each state is conditioned on the states of the parent node(s) [17]. Referring to Fig. 1, the table associated with node **A** encodes unconditional probabilities, while the tables associated with nodes **B**, **C**, and **D** encode conditional probabilities. The probabilities for each state in nodes **B** and **C** are conditioned on each state in node **A**, and the probabilities for each state in node **D** are conditioned on each state in node **C**. An example of a simple Bayesian network and the associated probability tables is shown in Appendix A.

The probabilities across all states of a child node, conditioned on each state in its parent node, sum to one. A requirement which must be satisfied within the structure of a Bayesian network is that of conditional independence, where two child nodes must be conditionally independent given their parent. In Fig. 1, nodes **B** and **C** must be conditionally independent given node **A**. Consequently, a Bayesian network describes the probabilistic and independence relationships between a set of random variables by conditioning child nodes upon their parent nodes. Additionally, the joint probability distribution for a set of random variables, $X_1, \ldots, X_n$, is decomposed to the product of their probabilities conditioned on their parents [16]. The expression describing this relationship, known as the *chain rule or Markov property*, is given in Eq. (1).

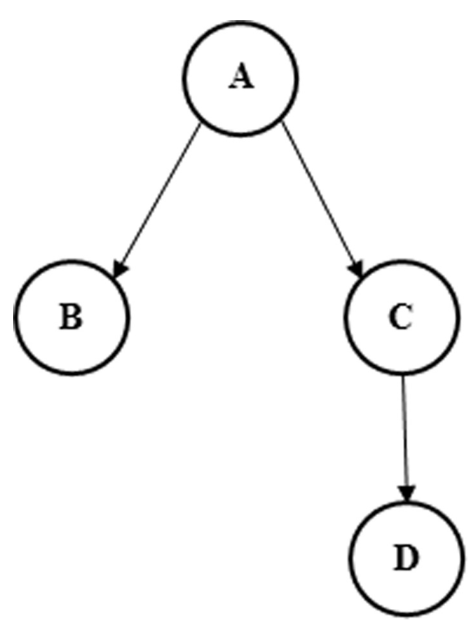

Fig. 1. A simple Bayesian network structure is illustrated. Node A is the parent of nodes B and C; node D is the child node of C. Tables associated with nodes B, C, and D encode conditional probabilities, while the table associated with node A encodes unconditional probabilities. Nodes B and C must be conditionally independent given node A.