Science and Justice 57 (2017) 209–220

Contents lists available at ScienceDirect

Science and Justice

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

Technical note

Towards a Bayesian evaluation of features in questioned handwritten signatures

Lorenzo Gaborini*, Alex Biedermann, Franco Taroni

University of Lausanne, School of Criminal Justice, Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 20 June 2016
Received in revised form 10 January 2017
Accepted 22 January 2017

Keywords:
Bayesian networks
Signature evidence
Fourier descriptors
Multivariate likelihood ratio

ABSTRACT

In this work, we propose the construction of a evaluative framework for supporting experts in questioned signature examinations. Through the use of Bayesian networks, we envision to quantify the probative value of well defined measurements performed on questioned signatures, in a way that is both formalised and part of a coherent approach to evaluation.

At the current stage, our project is explorative, focusing on the broad range of aspects that relate to comparative signature examinations. The goal is to identify writing features which are both highly discriminant, and easy for forensic examiners to detect. We also seek for a balance between case-specific features and characteristics which can be measured in the vast majority of signatures. Care is also taken at preserving the interpretability at every step of the reasoning process.

This paves the way for future work, which will aim at merging the different contributions to a single probabilistic measure of strength of evidence using Bayesian networks.

1. Introduction

Handwritten signatures have been employed since centuries as a means of authenticating one's identity on official documents. Their study has been one of the oldest disciplines in forensic science, yet its evaluative part did not achieve the same level of refinement as others.

Several professional groups, such as questioned document examiners, are trained to testify in courts by following established examination protocols for handwritten signatures. However, the usage of handwritten evidence in courts raises a number of issues. The scientific foundations of forensic handwriting comparisons are regularly doubted, in particular the mechanism by which forensic examiners arrive at and state their conclusions is often questioned. Specifically, the evaluation process is highly expert dependent and does not rely on standardized measurements and lines of reasoning, being thus highly dependent on the skill and proficiency of each examiner.

The use of forensic science in legal proceedings is based on the so-called "evaluative" framework: instead of stating a probability for a hypothesis, forensic experts report an expression of strength of support against two competing hypotheses, of forensic and legal interest [1]. To help evaluate the strength of support, a likelihood ratio is used in order to formalise the reasoning of the expert with respect to the relevant scientific findings.

The advantages of this evaluative framework are multiple: while formalised reasoning is much less liable to logical fallacies, experts will not express their beliefs on matters for which a court is responsible, notably on the hypotheses of interest. Further, the approach clarifies that the probability of hypotheses of interest also depends on information other than the scientific findings, allowing thus legal decision makers to incorporate in their reasoning a broad range of collateral case information.

1.1. The defence hypothesis

In forensic science, case-based evidence is collected and assessed under at least two competing hypotheses, those of the prosecution and the defence. In the domain of comparative forensic document examination, evidence takes the form of observed similarities and differences between questioned and reference ("known") items. To assess its value with respect to the competing hypotheses, the forensic scientist needs to evaluate the rarity of such similarities and differences in a given population of potential writers.

The choice and the size of the relevant population is of utmost importance, as it is very easy to overestimate the relevance of a character trait if it is shared by many or all the users of a determinate

* Corresponding author.
E-mail address: lorenzo.gaborini@unil.ch (L. Gaborini).

http://dx.doi.org/10.1016/j.scijus.2017.01.004

writing system [2]. For example, one may compare the writing features of a questioned item against those of two individuals, a number of suspects, or any other set of potential writers. If the relevant population spans a restricted number of individuals, the comparison is said to be a “closed-set”: on the other hand, if the population at large is considered, the situation is labelled “open-set” [3].

As a result, the value of the scientific findings strongly depends on their rarity in the reference population, though some traits might be more discriminating between two individuals rather than among a broader group of writers.

In this article, we mostly focus on closed-set circumstances, leaving the possibility to extend to open-set situations in future works.

1.2. Elements of Bayesian networks

To depict the reasoning using the previously illustrated interpretative framework, consider a single variable H which can assume two mutually exclusive states H_p and H_d, respectively the prosecution and the defence hypothesis. In a questioned signature examination scenario, we may associate e.g., H_p = “Person A has written the questioned signature” and H_d = “An unknown person has written the questioned signature”. Let E be the set of findings, as detected by the expert (e.g., similarities and differences between the questioned signature and the reference specimens). We denote with I the background information on the case, available to the expert.

Relevant to the recipient of expert information are the prior beliefs on H_p and H_d, conditioned by the background information: these are the probabilities $\Pr(H = h_p \mid I)$ and $\Pr(H = h_d \mid I)$, respectively[1]. More precisely, their ratio (called *prior odds*) is the relative strength of belief in H *a priori*.

The role of the expert is to evaluate the probability of having observed E under h_p and h_d: these terms are $\Pr(E|H=h_p, I)$ and $\Pr(E|H=h_p, I)$, respectively.

Bayes' theorem then states that the relative strength of belief in H *a posteriori* is proportional to the prior odds. In formulae:

$$\frac{\Pr(H = h_p \mid E, I)}{\Pr(H = h_p \mid E, I)} = \mathrm{LR}\frac{\Pr(H = h_p \mid I)}{\Pr(H = h_d \mid I)}$$

where

$$\mathrm{LR} = \frac{\Pr(E \mid H = h_p, I)}{\Pr(E \mid H = h_p, I)}$$

is called *likelihood ratio*. We observe that LR provides the expression for the strength of support of E versus the considered H: if $\mathrm{LR} > 1$, E provides more support to h_p rather than h_d, conditioned on the background information, and vice versa. Notice that it does *not* imply that h_p is more probable than h_d. To dissect the definition of the LR, the numerator reads as the probability of having observed E under h_p: referring to the previous example, it amounts to asking “what is the probability of observing the set of concordances and discordances in genuine signatures of Person A?”. The denominator, instead, is the probability of observing the same set of findings in signatures that appear to belong to Person A, but instead have been forged by someone else: this is assessed using the relevant population, defined in Section 1.1. In other terms, the LR is the ratio of two probabilities that account for, respectively, the *intra*- and *inter*-variability of findings.

Note that to apply the evaluative framework, one needs to specify not only the numerical values for the beliefs, but also the dependencies between the variables in terms of conditional probabilities. This can be intuitively represented in a graphical notation, which enables

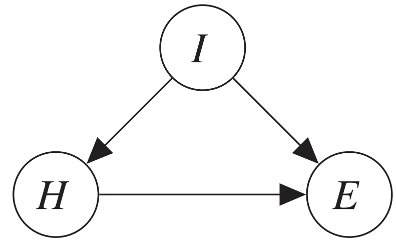

Fig. 1. The example in Section 1.2 as a Bayesian network. Note that the node I is usually omitted.

the forensic scientist to consider cases with multiple variables with differing interdependence. The obtained graphs are named *Bayesian networks* [4]: for instance, the previous example can be represented in Fig. 1. Note that the background information I is used to condition all relevant probabilities associated to H and E. For the sake of simplicity, such dependence is generally just assumed without a clear representation in the network. As a consequence, information I is usually omitted from explicit graphical representation in Bayesian networks.

Bayesian networks are very flexible, and have been used to support evaluative reasoning in very different forensic branches such as firearms [5], printed documents [6], signatures [7], forensic medicine [8] and DNA [9]. A review on the usage of Bayesian networks in forensic science can be found in Ref. [10].

1.3. Hierarchical evidence evaluation

A Bayesian network can be built for a rather generic evaluative procedure (e.g., the two-trace problem in Ref. [11]), but its structure can also be modified in order to accommodate for missing evidence [10]. Specifically, a binary-valued node M can be added to the network, encoding the fact that some evidence E is expected, but has not been retrieved (i.e., is missing). The usual assessments on the probability of E are now conditional on M.

The flexibility of Bayesian networks allows us to assess the probative value in complex cases, using evidence which is more and more case-based. In the works described later on, we envision to build a useful model to capture the essential features of the process of signature comparison, to help evaluating evidence which can always be measured (e.g., physical dimensions of the signatures) up to specific traits of one's signature (e.g., inner angles), which might not always be recognisable.

2. Related work

2.1. Pattern Recognition literature

A large amount of work has been recently done in building automatic classifiers for signatures. Most often, they exploit features which are not easy to describe (either empirically or mathematically), or have limited forensic interest. Such systems can be designed to work in a multi-writer environment, with known

Table 1
Composition and notation of the corpora.

Writer	N°	Description
A	$n_A = 143$	Authentic corpus
F_1	$n_{F_1} = 35$	
F_2	$n_{F_2} = 20$	
F_3	$n_{F_3} = 21$	
F_4	$n_{F_4} = 20$	
$F_{All} = \{F_1, F_2, F_3, F_4\}$	$n_F = 96$	Forged corpora
$W_{All} = \{A, F_1, F_2, F_3, F_4\}$	$n = 239$	Full corpus

[1] The reason for which h_p and h_d are written in lower-case letters is explained in Section 3.