Forensic Science International 216 (2012) 146–157

Contents lists available at SciVerse ScienceDirect

Forensic Science International

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

Using subsampling to estimate the strength of handwriting evidence via score-based likelihood ratios

Linda J. Davis [a,*], Christopher P. Saunders [b], Amanda Hepler [b,1], JoAnn Buscaglia [c]

[a] Department of Statistics (MS 4A7), George Mason University, Fairfax, VA 22030, United States
[b] Document Forensics Laboratory (MS 1G8), George Mason University, Fairfax, VA 22030, United States
[c] FBI Laboratory, Counterterrorism and Forensic Science Research Unit, Quantico, VA 22135, United States

ARTICLE INFO

Article history:
Received 15 February 2011
Received in revised form 14 September 2011
Accepted 21 September 2011
Available online 22 October 2011

Keywords:
Forensic science
Likelihood ratio
Subsampling
Handwriting evidence
Statistical evidence evaluation
Forensic document comparison

ABSTRACT

The likelihood ratio paradigm has been studied as a means for quantifying the strength of evidence for a variety of forensic evidence types. Although the concept of a likelihood ratio as a comparison of the plausibility of evidence under two propositions (or hypotheses) is straightforward, a number of issues arise when one considers how to go about estimating a likelihood ratio. In this paper, we illustrate one possible approach to estimating a likelihood ratio in comparative handwriting analysis. The novelty of our proposed approach relies on generating simulated writing samples from a collection of writing samples from a known source to form a database for estimating the distribution associated with the numerator of a likelihood ratio. We illustrate this approach using documents collected from 432 writers under controlled conditions.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

In the National Research Council's report to Congress in February 2009 [1], the current practice in the evaluation of nuclear DNA evidence is presented as a model for other branches of forensic science to follow in the presentation of the strength of evidence. One framework used in DNA analysis since the 1990s is based on quantifying the strength of evidence via a likelihood ratio [2,3].

A brief history of the use of a likelihood ratio to quantify the strength of evidence is given by Good [4]. Also, several texts (e.g., [5,6]) reveal the trend toward the interpretation of evidence using a likelihood ratio and provide a thorough discussion of its applicability to evidence interpretation.

Recently, the likelihood ratio paradigm has been studied as a means for quantifying the strength of evidence for a variety of types of forensic evidence, including speech [7], earmarks [8], latent fingerprints [9,10], footwear marks [11], glass fragments [12,13], and handwriting [14,15]. These references represent a small fraction of the research related to likelihood ratios in forensic sciences other than DNA. However, they serve to illustrate that the use of a likelihood ratio approach may provide a viable estimate of the strength of evidence.

The likelihood ratio paradigm is based on comparing the plausibility of evidence under two propositions (or hypotheses), usually denoted as the prosecution proposition (H_p) and the defense proposition (H_d). A number of recent papers (e.g., [16,17]) have discussed the proper choice of these two propositions in a variety of scenarios. A likelihood ratio is not testing a pair of competing propositions but rather seeking a measure of the relative support of a particular piece of evidence for the validity of one proposition H_p vs. another proposition H_d. The interpretation of a likelihood ratio equal to say V is that the evidence is V times more likely to have been observed if H_p is true than if H_d is true.

The form of a likelihood ratio makes explicit one of the key principles in the interpretation of forensic science information — that interpretation is only meaningful when two or more competing propositions are addressed [18]. Whenever observations are to be interpreted as evidence, they must be viewed from at least two perspectives; these may reflect, for instance, the forensic scientist's understanding of the respective positions of prosecution and defense in an adversarial system.

Although the concept of a likelihood ratio as a comparison of the plausibility of evidence under two or more propositions is straightforward, a number of issues arise when one considers

* Corresponding author. Tel.: +1 703 993 4835; fax: +1 703 993 1700.
E-mail addresses: ldavisc@gmu.edu (L.J. Davis), csaunde6@gmu.edu (C.P. Saunders), abhepler@innovativedecisions.com (A. Hepler), joann.buscaglia@ic.fbi.gov (J. Buscaglia).
1 Present address: Innovative Decisions, Inc., 1945 Old Gallows Road Suite 207, Vienna, VA 22182, United States.

0379-0738/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2011.09.013

how to go about computing a likelihood ratio. In forensic DNA analysis, such computations involve estimating the relative frequency of different DNA profiles in relevant populations. In areas outside of forensic DNA analysis (such as glass fragments, latent fingerprints, and handwriting), however, additional issues arise. Some of these issues are related to the numerator not necessarily being one, as is often the case with (single source) DNA evidence. Other issues are related to evidence being in the form of continuous measurements. This involves taking ratios of probability density functions instead of probabilities to form a likelihood ratio, and estimating probability density functions instead of probabilities. (Forming likelihood ratios based on continuous measurements is also currently being studied in the context of evidence-based medical diagnostic testing [19–21].)

One promising alternative when a likelihood ratio cannot be estimated without making perhaps unfounded assumptions about the underlying processes that generate the evidence is a score-based approach, which has recently been summarized in [22]. In this approach, a likelihood ratio is calculated as a measure of the strength of evidence when the evidence is the result of a comparative analysis between two samples. This approach has been applied to various types of evidence: signatures [23], voice recordings [7], and fingerprint evidence [10].

The purpose of this paper is to illustrate one possible application of a score-based approach to comparative handwriting analysis. The novelty of our proposed implementation of a score-based approach relies on generating simulated writing samples from a collection of writing samples from a known source to form a database for estimating the distribution associated with the numerator of a likelihood ratio. Estimation of this distribution is central to the calculation of a likelihood ratio and rarely there are sufficient writing samples available to estimate it accurately.

This paper is organized as follows. We begin with an overview of the likelihood ratio paradigm, including the rationale for using a score-based likelihood ratio. Then, we state our proposed approach to evaluating evidence in comparative handwriting analysis. In particular, we describe in detail the components required to implement our proposed approach, including three necessary databases of samples. Then, we turn to the main focus of this paper: how to generate via subsampling a database of (simulated) writing samples that can be used to estimate the distribution associated with the numerator of a likelihood ratio. We illustrate the proposed approach using a data set of handwriting samples collected under controlled conditions and processed using a proprietary feature extraction method. Finally, we illustrate how this subsampling technique can also be used to study the effect of the size of the questioned document (as measured by number of characters in it) on the strength of evidence.

2. Overview of the likelihood ratio paradigm

A likelihood ratio is an empirical tool to evaluate and compare propositions, often concerning the likely source of some evidence. It is one approach to answering the question: What does this evidence tell me about proposition *A* vs. proposition *B*? It does so by assigning plausibility weights to propositions that aid in the selection of the proposition that provides the best explanation for the evidence.

In the following discussion, we treat the evidence as the realization of some random variables that have probability distributions. A likelihood ratio is then calculated as the ratio of the distributions of these random variables either in terms of probabilities (for discrete evidence) or probability density functions (for continuous evidence) under two propositions, evaluated at the realized value of the evidence. In this scenario, a likelihood ratio (*LR*) can be defined as:

$$LR \equiv \frac{g[E|H_p, I]}{g[E|H_d, I]}$$

for given prosecution proposition (H_p) and the defense proposition (H_d), where:

I is the background information,
E is the evidence (new information) assumed to be a realization of some random variables, and
$g[\cdot|H, I]$ is the probability distribution associated with the evidence given background information I and assuming proposition H is true. When the evidence involves discrete measurements, $g[\cdot|H, I]$ is a probability; when the evidence involves continuous measurements, $g[\cdot|H, I]$ is a (continuous) probability density function.

The background information I may be items of information relating to time and location of the discovered evidence, and/or possibly eyewitness statements related to the evidence. This body of information forms a framework of circumstances within which the scientific examination of the new information, i.e., the evidence, is carried out [24]. The inclusion of I in the above formula for a likelihood ratio makes explicit that interpretation of scientific evidence is carried out within a framework of circumstances and that its interpretation is affected by the structure and content of that framework.

2.1. Interpretation

One interpretation of a likelihood ratio (based on Bayes' theorem [5]) is as the multiplicative factor by which the ratio of the probabilities of two propositions changes with observation of the evidence. For example, suppose the trier of fact can quantify beliefs about guilt prior to consideration of the new information, i.e., the evidence. A likelihood ratio for each piece of evidence can be used to update prior (before viewing evidence) belief to posterior (after viewing evidence) belief given the new evidence according to the following formula:

$$\underbrace{\frac{Pr(H_p|E, I)}{Pr(H_d|E, I)}}_{\text{posterior probability ratio}} = \underbrace{\frac{g[E|H_p, I]}{g[E|H_d, I]}}_{\text{likelihood ratio}} \times \underbrace{\frac{Pr(H_p|I)}{Pr(H_d|I)}}_{\text{prior probability ratio}}$$

Formally, the trier of fact could then reach a final decision using the posterior odds in favor of H_p (e.g., the suspect's guilt) given the evidence. In this paradigm, the trier of fact is responsible for quantifying prior beliefs about H_p and H_d, while the forensic scientist is responsible for providing a summary of the evidence needed to update these beliefs given the evidence. Attempts to date to utilize such an interpretation in court have met with some opposition [25].

This relationship between prior and posterior odds via multiplying by a likelihood ratio follows directly from the laws of probability and therefore is reasonable as a framework for interpreting evidence. The prior odds are the "odds" in favor of H_p before observing the evidence; the prior odds are based in part on an evaluation of any available background information. The posterior odds are the "odds" in favor of H_p after observing the evidence. A likelihood ratio then details how the two are related. (Note: Unless H_p and H_d are exhaustive, the prior and posterior odds are not "odds" in the usual sense of being the ratio of a probability to one minus that probability; in general, the prior and posterior odds are relative risks, i.e., the ratio of two probabilities.)

Download English Version:

https://daneshyari.com/en/article/96108

Download Persian Version:

https://daneshyari.com/article/96108

Daneshyari.com