Handwriting Identification Using Random Forests and Score-Based Likelihood Ratios

Lead Researchers: Madeline Quinn Johnson and Danica M. Ommen
Journal: Statistical Analysis and Data Mining | **Publication Date:** 03 December 2021
Link: forensicstats.link/HandwritingRecognition-DOI

OVERVIEW

Handwriting analysis has long been a largely subjective field of study, relying on visual inspections from trained examiners to determine if questioned documents come from the same source. In recent years, however, efforts have been made to develop methods and software which quantify the similarity between writing samples more objectively. Researchers funded by CSAFE developed and tested a new statistical method for handwriting recognition, using a score-based likelihood ratio (SLR) system to determine the evidential value.

GOALS

- Apply the SLR system to various handwritten documents.
- Evaluate the system's performance with various approaches to the data.

THE STUDY

CSAFE collected handwriting samples from 90 participants, using prompts of various lengths to get samples of different sizes. These writing samples were broken down into graphs, or writing segments with nodes and connecting edges, then grouped into clusters for comparison.

When comparing the gathered samples, Johnson and Ommen considered two possible scenarios:

- **Common Source Scenario:** two questioned documents with unknown writers are compared to determine whether they come from the same source.
- **Specific Source Scenario:** a questioned document is compared to a prepared sample from a known writer.

They then used **Score-based Likelihood Ratios** (SLRs) to approximate the weight of the evidence in both types of scenarios.

The researchers used three different approaches when generating the known non-matching comparisons for the specific source SLRs:

- **Trace-Anchored Approach:** only uses comparisons between the questioned document (the trace) and a collection of writers different from the specific source (the background population).
- **Source-Anchored Approach:** only uses comparisons between writing from the specific source and the background population.
- **General-Match Approach:** only uses comparisons between samples from different writers in the background population.

Once the SLRs for each scenario were generated, they used random forest algorithms to determine comparison scores, including a pre-trained random forest using all of the gathered data, and one trained according to the relevant SLR.

RESULTS

- In common source scenarios, the trained random forest performed well with longer writing samples, but struggled with shorter ones.
- The specific source SLRs performed better than the common source SLRs because they are tailored to the case at hand.
- In all scenarios, it was more difficult for the SLR system to confirm samples with the same source than with different sources.

FOCUS ON THE FUTURE

The SLRs do not perform well with short documents, possibly due to a mismatch between the number of clusters used and the length of the document. Future work could determine the optimal number of clusters based on the document's length.

Because the SLRs provide data on the strength of forensic handwriting evidence for an open-set of sources, this approach is an improvement on the previous clustering method developed by CSAFE, which used a closed set of known sources.

LEARN MORE

For further details, the full paper can be found here: forensicstats.link/HandwritingRecognition

Additionally, explore relevant publications:

- A clustering method for graphical handwriting components and statistical writership analysis
forensicstats.link/ClusteringMethod
- Bayesian hierarchical modeling for the forensic evaluation of handwritten documents
forensicstats.link/BayesianModeling

FUNDING

csafe
Center for Statistics and
Applications in Forensic Evidence