Two-Stage Approach for Forensic Handwriting Analysis

Ashlan Simpson and Danica Ommen

Introduction

Prior Research: The feature-based and score-based likelihood ratios assess evidence within an evaluation framework. The evaluation framework leads to loosely defined error rates causing difficulties with method assessment.

Motivation: The 2016 PCAST report for strengthening the foundations of forensic science recommends that all methods have clearly defined error rates for use in justice proceedings.

Current Research: The Two-Stage framework meets these recommendations by allowing for the quantification of error rates.

Error Rates: The rate of false associations is called the Random Match Probability (RMP), and the rate of false exclusions is called the Random Non-Match Probability (RNMP).

Materials

Writing Exemplars: CSAFE collected handwritten samples from 241 writers aged 18-60 years old. Participants provided three writing samples for each of the three prompts in three sessions leading to 27 samples per participant.

Processing: Each handwriting sample was broken down into smaller segments and then clustered according to 40 common shapes using the 'Handwriter' Package. We use the proportion of segments in each cluster as data to represent each handwritten sample.

Methods

Two-Stage Approach:

Stage One: If the comparison score is low ($< \tau$), declare an association.
If the score is large ($> \tau$), declare an exclusion.

Stage Two: If you associated, find the significance of the match by computing the probability that two documents from different writers would also match (RMP). If you excluded, find the significance of the non-match by computing the probability that two documents from the same writer would also non-match (RNMP).

Choosing τ: If τ is too large, there will be many false associations, but if τ is too small there will be many false exclusions. We produced two values for τ using the roc function from the pROC package, one that produces optimal error rates and one that produces equal error rates.

Distance metrics: We use three distance metrics to compare the similarity between handwritten documents. Formulas for the distances between handwriting samples one and two are below.

Euclidean Distance: $\sqrt{(v_{1,1}-v_{1,2})^2+(v_{2,1}-v_{1,2})^2+\cdots+(v_{40,1}-v_{40,2})^2}$.

Manhattan Distance: $|v_{1,1}-v_{1,2}|+|v_{2,1}-v_{2,2}|+\cdots+|v_{40,1}-v_{40,2}|$

Supremum Distance: $\max(|v_{1,1}-v_{1,2}|,\ |v_{2,1}-v_{2,2}|,\ldots,|v_{40,1}-v_{40,2}|)$

Diagnostics: The area under the curve (AUC) produced by the roc function was used to determine the best choice of distance metric among the three options listed above.

Discussion

Area Under the Curve: AUC can range from 0 to 1, with values closer to one indicating better performance. Supremum performed the worst with an AUC of 0.818, while Manhattan did the best with an AUC of 0.88.

Error Rate: The true association rate or TAR is the probability that handwriting samples were classified as associated when they were from the same writer. The true exclusion rate or TER is the equivalent for exclusions and different writers. We chose τ when the TER and TAR were equal (equal error rate or EER) and when the maximum of TER+TAR (optimal error rate or OER). For Manhattan, the threshold that produced EER (RMP=RNMP=1-0.784) is τ=0.708, and the threshold that resulted in OER (RMP=1-0.737, RNMP=1-0.839) is τ=0.739.

Future Directions

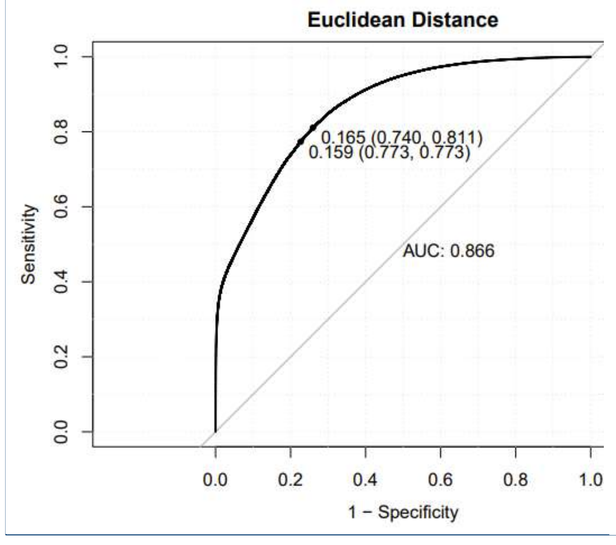
Distance Metrics: The Manhattan distance metric performed the best at 88% accuracy, but ideally, the AUC would be larger. These distance metrics were simple compared to sophisticated scores developed using machine learning algorithms.

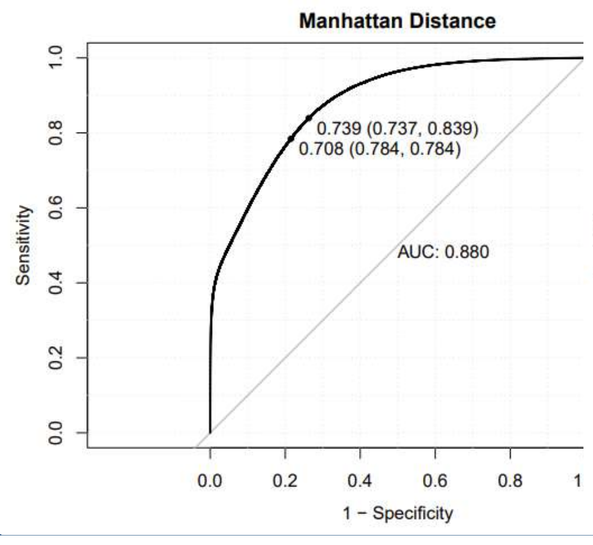
Two-Stage Approach: For a given threshold, τ, and relevant population of background comparisons, the RMP and RNMP are fixed, meaning they are the same regardless of whether the comparison was easy or hard. The binary outcome of association or exclusion results in the fall-off-the-cliff effect; there is nothing in the middle to say you are unsure. Researchers can address this drawback by combining it with a second method, such as a score-based likelihood ratio (see NIJ grant #2018-DU-BX-0228).

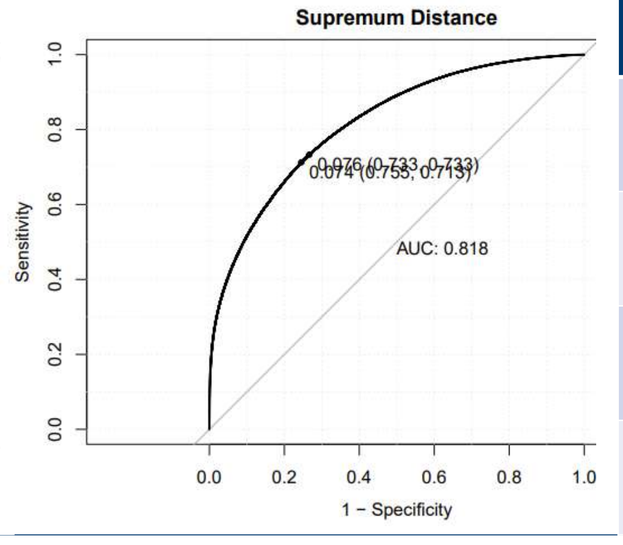
References

1. Crawford (2020); Bayesian hierarchical modeling for forensic evaluation of handwritten documents. Iowa State Theses & Dissertations. https://doi.org/10.31274/etd-20200624-257
2. Johnson, M. Q., Ommen, D. M., Handwriting identification using random forests & score-based likelihood ratios, Stat. Anal. Data Min.: ASA Data Sci. J.. 15 (2022), 357–375. https://doi.org/10.1002/sam.11566
3. Fuglsby, Cami, "U-Statistics for Characterizing Forensic Sufficiency Studies" (2017). Electronic Theses & Dissertations. 1715. https://openprairie.sdstate.edu/etd/1715

Results







Type for Manhattan	Threshold	Stage 1	Stage 2
Easy Match	EER OPT	Associate Associate	RMP=0.216 RMP=0.263
Hard Match	EER OPT	Exclusion Exclusion	RNMP=0.216 RNMP=0.161
Easy Non-Match	EER OPT	Exclusion Exclusion	RNMP=0.216 RNMP =0.263
Hard Non-Match	EER OPT	Associate Associate	RMP=0.216 RMP=0.161

This work was funded (or partially funded) by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreements 70NANB15H176 and 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, University of Pennsylvania, Swarthmore College and University of Nebraska, Lincoln.