J20 Statistical Analysis of Handwriting: Probabilistic Outcomes for Closed-Set Writer Identification.

Amy M. Crawford, MS, Center for Statistics and Applications in Forensic Evidence, Iowa State University, Ames, IA 50011; Alicia L. Carriquiry, PhD, Center for Statistics and Applications in Forensic Evidence, Iowa State University, Ames, IA 50011; Danica Ommen, PhD, Center for Statistics and Applications in Forensic Evidence, Iowa State University, Ames, IA 50011*

Learning Overview: The goal of this presentation is to provide insights into features of handwritten documents that are important for statistical modeling with the task of writer identification.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by introducing ongoing forensic statistics research to practitioners in the field. Collaboration and communication is fundamental to the success of emerging research. All statistical methods and code are open-source and available to the community.

The objective of this work is to develop a statistical model that takes all styles of writing (cursive, print, and connected print) and gives a probabilistic evaluation of a closed-set writer identification as well as variability estimates. Such probabilistic statements have potential to support examiner conclusions and be used by the community to provide free and open-source methods for searching a collection of documents.

When forensic examiners compare handwritten evidence, they often focus on small details of writing. Likewise, this statistical approach to the comparison of such documents begins by decomposing writing into small meaningful connected pieces of ink, often corresponding to letters. The R package handwriter, developed at the Center for Statistics and Applications in Forensic Evidence, handles this processing task. The small pieces are treated as graphical structures, and graphs are grouped using a novel dynamic method that is based on the similarity of their shapes. The frequency at which graph types appear in writing, along with measurements taken on the small graphs, serve as data for a Bayesian hierarchical model. This study considered measurements such as slopes, lengths, centroid locations, and loops and compared the measurements within groups to investigate their ability to discriminate between writers. Measurements that separate writers well in a numeric capacity will be useful for the statistical model.

There are a few notable insights that arise through the modeling process. First, the model that takes all styles of writing as inputs is able to separate print and cursive writings through a latent variable. Next, through a multivariate over-dispersion calculation, one can evaluate intra-writer variability as it relates to the model.

The statistical model is ultimately used to perform a closed set writer identification analysis. Data for the analysis come from the first 100 writers in the Center for Statistics and Applications in Forensic Evidence (CSAFE) handwriting data collection. In this data collection, study participants were ask to transcribe three writing prompts: the “London Letter,” a short excerpt from *The Wonderful Wizard of Oz*, and an even shorter common phrase. Each prompt was written three times during each of three data collection sessions. This study investigated writer identification performance under varying amounts of known and questioned writing.

This work was partially funded by the CSAFE through a Cooperative Agreement between the National Institute of Technology (NIST) and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.

Statistics, Questioned Documents, CSAFE