Statistical Analysis of Letter Importance for Forensic Document Examination

Amy Crawford

A particularly difficult problem facing the field of forensic document examination is whether handwriting found on two different documents originated from the same person. In recent years researchers have begun to uncover the process that talented document examiners follow when conducting a visual analysis of documents. This has been done using several approaches, one of the most fascinating is eye movement tracking, where a device tracks where the eyes are drawn to and how long they stay fixated at certain locations on writing samples. This gives insight to which features are seemingly important to examiners when making decisions of authorship. There has also been a significant emphasis on the use of data and statistical methodologies to assist in comparison of handwritten documents. In a statistical context, data are measurements taken from ink on paper that has been scanned and processed. In this research we take an intuitive approach and focus on data derived from the important pieces of the writing that examiners focus on - letters.

Handwriting is a very complex set of information, so focusing on meaningful bits of it at a time lends nicely to an algorithmic approach to writer identification. Naturally, one might be curious about which letters are the most important for differentiating writers from one another in a statistical context. Along with visual tools, we used a machine learning approach to conduct an analysis of letter importance. The results of this importance analysis were then compared to the procedure of a forensic document examiner from the Iowa Division of Criminal Investigation Laboratory. Following this analysis we used letters found to be important to investigate differences between writers.

A common belief amongst forensic document examiners is that environmental factors can impact an individual's handwriting. Thus, in addition to letter importance and between writer comparisons, we present an investigation of handwriting variability within single writers. Results from comparison of writing samples collected on different days for a single writer are presented. Additionally, we present interesting patterns that show up from writing samples taken sequentially and samples with differing content.