Quantifying Common Word Variance Through Rainbow Triangle Graph Decomposition

Alexandra Arabio, Annie DiFrank, Alicia Carriquiry, Danica Ommen

Background and Goals

Writing exemplars: Handwritten samples were collected by CSAFE and added to an open repository of over 90 writers ranging in ages 18-60+. This project uses the common phrase prompt from this repository for analysis. These writing samples are done in triplicate at each session and each writer sat for two other sessions that were at least three weeks from one another in attempt at gaining an understanding of the variance of each writer.

Decomposition Process: The idea of node/point decomposition stems from the concept the graph theory question surrounding Kneser Graph decomposition of set 15. This decomposition process uses a piece wise matching process to gain quantitative information from an exemplar using distances, lengths, and slopes of vectors.

Rainbow Triangulation: Rainbow triangles can be formed using the same principles of Kneser Graphs if all edges and vertices have different colors. This process can provide information on the variation of one writer's repetition through comparing the similarity of each triangle formed and forming a threshold of variance.

Methods

Process of Analysis: CSAFE has been working on packaging a software to aid in handwriting analysis entitled 'handwriter'. This software takes the original writing exemplar and thins it into its pixilated components so that nodes can be placed on the sample to indicate specific characteristics. Once these nodes are placed the sample is then broken down into it's individual 'words'. For this project, the 6th, 9th, and 13th words being 'the' are extracted for the common word comparison. This is where the nodes are color coded as shown below in figure 1.

Once the nodes are places onto the sample, the blue node is indicated through the software and completely decomposed to form vectors. These vectors become the edges of the rainbow triangles that will be formed through the connecting of three unique vectors with three uniquely colored nodes.

Comparison: After the formal analysis described above is performed for each exemplar, the samples are combined to form an average triangle of each color in order to gain a quantitative and visual average triangle based on the clustering of each node and edge. This average will be used to discover the variance of each writer and form the threshold that will be used to include or exclude writers when attempting to match an unknown writer with the known writers in the repository. A data frame is also formed containing other graphical features for added comparison.

Node Color Codes:

- Purple = lowest most pixel
- Orange = right most pixel
- Blue = left most pixel
- Pink = 2+ intersecting or overlapping pixels
- Green = highest most pixel

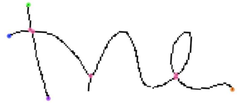

Figure 1. Example of writer X common word with node placement

Data and Results

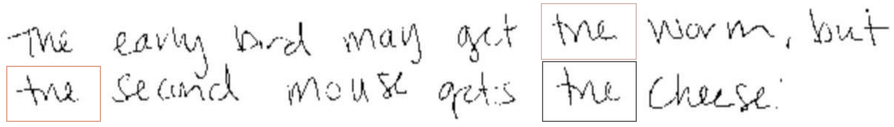

Figure 2. Example of writer X common phase indicating extracted common words

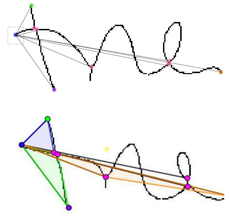

Figure 3. Decomposition of blue node and triangulation process vectors

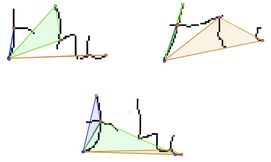

Figure 4. Triangulation of common word from the common phrase prompt from writer Y repetition 1

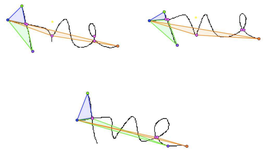

Figure 5. Triangulation of common word from the common phrase prompt from writer X repetition 1

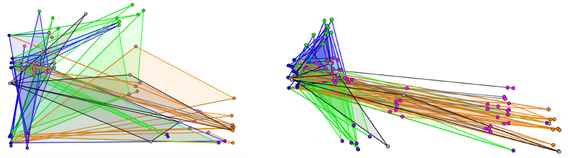

Figure 6. Comparison of writer Y and writer X compilations of all 27 common words exemplars from each writer

Discussion and Future Directions

Discussion: Through the rainbow triangulation process we have been able to obtain ranges of possible distances, lengths, areas, and other components of triangles for each writer. Using these ranges, we can form thresholds of likelihood of an unknown exemplar being written by a certain known writer. With these thresholds, we exclude known writers as possible suspects when the unknown sample does not fit with the variance. In a case where there is a closed set of writers, being able to exclude all but one is just as efficient as matching one writer with an unknown. Being able to examine and compare components of a full word such as in figure 6 will help provide a visual for a lay juror as well as be able to simplify the understanding of the comparison.

Future Directions: As this project is further developed, we hope to investigate different common words as well as how this method may be applicable for signatures. We also hope to apply more statistical analysis techniques to this method moving forward to show the uniqueness of each writer.

Resources and References

1. *Crawford, Amy; Ray, Anyesha; Carriquiry, Alicia; Kruse, James; Peterson, Marc (2019): CSAFE Handwriting Database. Iowa State University.Dataset.* https://doi.org/10.25380/iastate.10062203.v1
2. . Ehard, Stefan, and Elena Mohr. "Rainbow Triangles and Cliques in Edge-Colored Graphs." *European Journal of Combinatorics*, vol. 84, 2020, p. 103037., doi:10.1016/j.ejc.2019.103037.
3. Whitt, T.r., and C.a. Rodger. "Decomposition of the Kneser Graph into Paths of Length Four." *Discrete Mathematics*, vol. 338, no. 8, 2015, pp. 1284–1288., doi:10.1016/j.disc.2015.02.011.

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.