Is “Fake” a Writing Style? Detecting Stylistic Alteration

Patrick Juola, Evaluating Variation in Language Lab, Duquesne University

http://www.evllabs.com, juola@mathcs.duq.edu

Sometimes we need to know who wrote a document (like a ransom note). But sometimes, the writer needs *not* to be known (like a whistle-blower). “Stylometry,” the analysis of writing style, is a powerful tool for unmasking writers based on how they write. Sometimes too powerful, as many criminals and forgers have discovered to their grief.

Are there ways to mask a person’s writing style? How can whistleblowers mask their identities and protect themselves? In collaboration with Allen Riddell (U. Indiana), our two-year project is to develop tested and reliable methods to defend against stylometric analysis.

Brennan & Greenstadt developed a corpus of AMT-written essays in “normal,” obfuscative, and imitative styles, and showed that it was very difficult for conventional stylometric techniques to identify the correct author. We extend this work here to show that it may be possible to detect “fake” as a consistent, detectable attribute of writing.

Materials & Methods

- Java Graphical Authorship Attribution Program (JGAAP)
- Real and parody writings by Faulkner and Hemingway

Experimental Questions

- RH v RF (actual authors)
- RH v FH (Hemingway variants)
- RF v FF (Faulkner variants)
- (RF & RH) v (FF & FH)
- *F training, *H testing (planned)
- *H training, *F testing (planned)
- Classify B&G corpus documents as real/fake (planned)

The results show JGAAP-based analysis is a fair approach to discerning imitation in writing (by contrast, word use and punctuation use does not seem to work), and identifies natural writing as imitative only 10-15% of the time. However, it is prone to identifying imitative writing as natural, especially in the case of longer character n-grams.

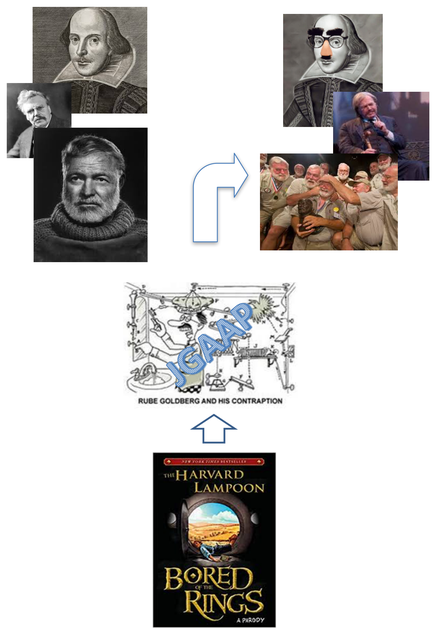
Additionally, as Hemingway's imitators were incorrectly identified significantly more often both in the individual author experiments and the combined experiments, the nature of a writing style may significantly affect our ability to recognize imitation of it.

Why does this matter? If the authenticity of a work is open to question, a detectable “fake” style may be grounds for suspicion or further detailed inquiry. It could even be taken as evidence of harmful intent, putting the author at risk. This type of analysis may also be a useful first step in a general authorship attribution attack.

Stylometry provides a way to infer identity online. Stylometric defense provides badly needed defense against this inference.

It can protect whistleblowers, journalists’ sources, human rights investigators,….

Ultimately, a good defense against authorship attribution attacks should be invisible. This work is a preliminary Red Team analysis in preparation for fuller work with Riddell’s data.



This work has significant impact not only on the forensics community (is this document genuine?) but also for any application where a person has reason to welcome (or fear) that someone will be validating the origins of a disputed document.

Examples include education, journalism, intelligence/national security, history, computer security, and law.

By extension, this type of analysis can enhance society’s trust in e-documents.