Andrew | Ilsa | Abhinil

Authorship Attribution

With Burrows Delta and Nearest Centroid Classifier

CS_410

THE MODEL

1: Applies Text Distortion

Text distortion is a means of enhancing stylization choices of the author, so the model would focus less on the content choice/genre (topic words). We implement the procedure outlined by Stamatatos, which uses a list *W_k* to specify words to not be distorted.

2: Extracts Stylometric Features

Select the top *N* most frequent words across entire collection as features, then compute for each document a vector encoding the relative frequency of each feature.

3: Builds Centroids for each Author

Compute the mean (mu) and standard deviation (sigma) for each feature across every document vector, then apply a Z-transform to each vector. For each author, the centroid is the mean across these Z-values.

4: Computes Burrow's Delta

For a new text: compute the feature vector, apply the Z-transform, measure the mean distance from each author's centroid, taking the prediction as the minimum value.
Burrow's original implementation uses a normalized Manhattan metric. We generalize Burrow's Delta to other metric functions, like Cosine Similarity.

E. Stamatatos, Authorship Attribution Using Text Distortion,
https://icsdweb.aegean.gr/stamatatos/papers/eacl2017.pdf

S. Evert, T. Proisl, F. Jannidis, S. Pielström, C. Schöch, T. Vitt,
Towards a better understanding of Burrows's Delta in literary authorship attribution,
https://aclanthology.org/W15-0709.pdf

THE DATA

- Project Gutenberg: A large library of over 75,000 free ebooks
- Web crawler: Scrapes works of the 100 most downloaded authors
    - Raw unprocessed text containing 2,036 books (totals ~180,000,000 words)

Cleaning and preprocessing

- Removed metadata, headers, and footers
- Created train/test split (80/20) stratified by author
- Discarded authors with fewer than 7 texts
- Kept punctuation and letter casing intact while standardizing numbers for each text to capture stylistic patterns

CHALLENGES TO MODEL ACCURACY

Text Distortion

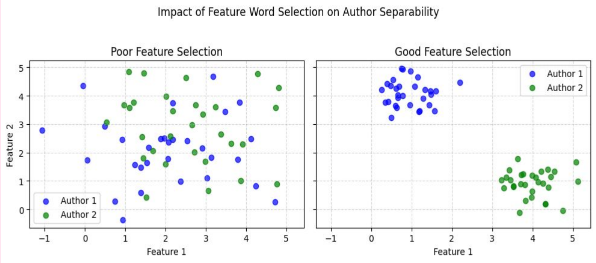
- Source texts have varying lengths and formats
- Overly simple preprocessing rules negatively affect word frequency for some authors
- Text truncation or missing sections impacts stylistic markers
- Quoted text from other authors creates noise

The quick ***** *** jumps **** the lazy ***.

W_k = { “the”, “quick”, “jumps”, “lazy” }

Burrows Computation

- Z-score normalization can amplify noise in rare words
- Manhattan distance (Delta) may not capture all stylistic differences
- Selection of feature words critically impacts accuracy



Author Ambiguity

- Authors with similar writing periods show stylistic convergence
- Authors may adopt different styles across their career
- Small sample sizes for some authors limit reliable modeling
- Translations into English, such as is with many Classical works, can lose author style

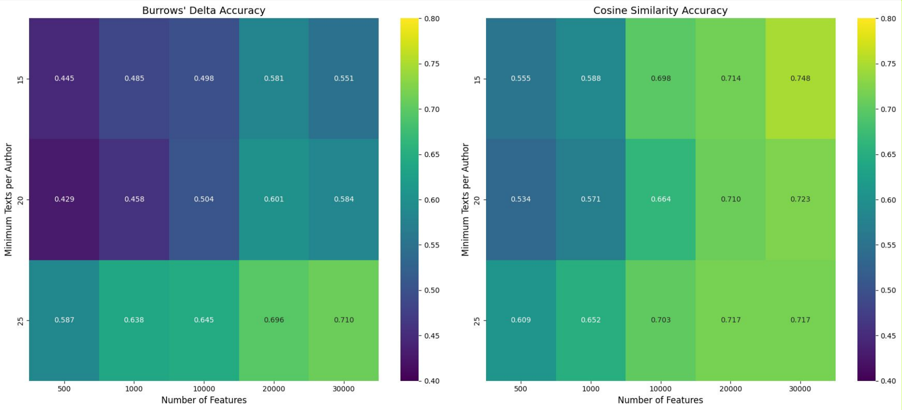
RESULTS - Classifying 100 Authors

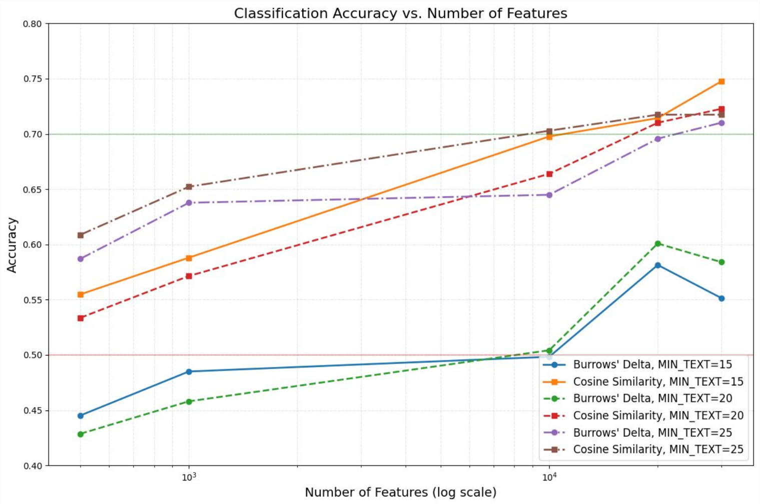
Best Burrows' Delta configuration:
MIN_TEXT: 25
NUM_FEATURES: 30000
Accuracy: 0.7101

Best Cosine Similarity configuration:
MIN_TEXT: 15
NUM_FEATURES: 30000
Accuracy: 0.7475

Overall best configuration:
Metric: Cosine Similarity
MIN_TEXT: 15
NUM_FEATURES: 30000
Accuracy: 0.7475

Accuracy of Random picking would be 1/100 = 0.01


Burrows' Delta Accuracy
Minimum Texts per Author
Number of Features
Cosine Similarity Accuracy
Minimum Texts per Author
Number of Features


Classification Accuracy vs. Number of Features
Accuracy
Number of Features (log scale)
10^3
10^4
Burrows' Delta, MIN_TEXT=15
Cosine Similarity, MIN_TEXT=15
Burrows' Delta, MIN_TEXT=20
Cosine Similarity, MIN_TEXT=20
Burrows' Delta, MIN_TEXT=25
Cosine Similarity, MIN_TEXT=25