Project 1 Implementation Report

Hanzhang Yin

October 19, 2024

For my implementation, I used two different strategy to address the author prediction task.

Log-likelihood Calculation:

Zhu Shuzhen's Markov Matrix:

$$\begin{bmatrix} \frac{3}{10} & \frac{47}{190} & \frac{29}{190} & \frac{3}{10} \\ \frac{53}{156} & \frac{1}{4} & \frac{11}{78} & \frac{7}{26} \\ \frac{35}{103} & \frac{27}{103} & \frac{15}{103} & \frac{26}{103} \\ \frac{55}{199} & \frac{56}{199} & \frac{27}{199} & \frac{61}{199} \end{bmatrix}$$

Du Fu's Markov Matrix:

$$\begin{bmatrix} \frac{1}{3} & \frac{29}{103} & \frac{46}{309} & \frac{73}{309} \\ \frac{88}{279} & \frac{73}{279} & \frac{49}{279} & \frac{23}{93} \\ \frac{27}{80} & \frac{41}{160} & \frac{1}{10} & \frac{49}{160} \\ \frac{69}{236} & \frac{37}{118} & \frac{21}{118} & \frac{51}{236} \end{bmatrix}$$

Total log-likelihood for Zhu Shuzhen:

$$\begin{aligned} &3\log\left(\frac{35}{103}\right) + 5\log\left(\frac{53}{156}\right) + 4\log\left(\frac{61}{199}\right) + 3\log\left(\frac{3}{10}\right) + 6\log\left(\frac{56}{199}\right) \\ &+ 3\log\left(\frac{55}{199}\right) + 2\log\left(\frac{7}{26}\right) + 5\log\left(\frac{26}{103}\right) + 5\log\left(\frac{1}{4}\right) \\ &+ 3\log\left(\frac{47}{190}\right) + 4\log\left(\frac{29}{190}\right) + 2\log\left(\frac{15}{103}\right) + 2\log\left(\frac{11}{78}\right) \\ &+ \log\left(\frac{27}{199}\right) \end{aligned}$$

Total log-likelihood for Du Fu:

$$3\log\left(\frac{27}{80}\right) + \log\left(\frac{1}{3}\right) + 5\log\left(\frac{88}{279}\right) + 6\log\left(\frac{37}{118}\right) + 5\log\left(\frac{49}{160}\right) + 3\log\left(\frac{69}{236}\right) + 3\log\left(\frac{29}{103}\right) + 5\log\left(\frac{73}{279}\right) + 2\log\left(\frac{23}{93}\right) + 2\log\left(\frac{73}{309}\right) + 4\log\left(\frac{51}{236}\right) + \log\left(\frac{21}{118}\right) + 2\log\left(\frac{49}{279}\right) + 4\log\left(\frac{46}{309}\right) + 2\log\left(\frac{1}{10}\right)$$

Zhu Shuzhen's Test Case Output:

Zhu Shuzhen

Zhu Shuzhen's & Du Fu Test Case Output:

Euclidean Distance and Consine Similarity Calculation:

Zhu Shuzhen's Markov Matrix:

$$\begin{bmatrix} \frac{57}{200} & \frac{47}{169} & \frac{29}{93} & \frac{19}{62} \\ \frac{53}{200} & \frac{3}{13} & \frac{22}{93} & \frac{7}{31} \\ \frac{7}{40} & \frac{27}{169} & \frac{5}{31} & \frac{13}{93} \\ \frac{11}{40} & \frac{56}{169} & \frac{9}{31} & \frac{61}{186} \end{bmatrix}$$

Column sums: [1, 1, 1, 1, 0]

Du Fu's Markov Matrix:

$$\begin{bmatrix} \frac{103}{314} & \frac{87}{275} & \frac{46}{153} & \frac{73}{242} \\ \frac{44}{157} & \frac{73}{275} & \frac{49}{153} & \frac{69}{242} \\ \frac{27}{157} & \frac{41}{275} & \frac{16}{153} & \frac{49}{242} \\ \frac{69}{314} & \frac{74}{275} & \frac{14}{51} & \frac{51}{242} \end{bmatrix}$$

Column sums: [1, 1, 1, 1, 0]

Predicted author for Zhu Shuzhen's test tones:

Cosine Similarity with Zhu Shuzhen: 0.5728528390325099
Cosine Similarity with Du Fu: 0.612084042064026
Euclidean Distance to Zhu Shuzhen: 0.8218014556290012
Euclidean Distance to Du Fu: 0.7969871938136044
Cosine Similarity Prediction: Du Fu
Euclidean Distance Prediction: Du Fu

Baseline Method - Infinity Vector Norm Calculation:

Baseline Infinity Norm Prediction for Zhu Shuzhen's test tones:

- Infinity Norm Difference with Zhu Shuzhen: 0.07692068226362686
- Infinity Norm Difference with Du Fu: 0.09668192553969868
- Baseline Infinity Norm Prediction: Zhu Shuzhen

Baseline Infinity Norm Prediction for Du Fu's test tones:

- Infinity Norm Difference with Zhu Shuzhen:
- Infinity Norm Difference with Du Fu:
- Baseline Infinity Norm Prediction: