

Project 1 Pre-port

Hanzhang Yin

October 2, 2024

The project aims to distinguish the styles of Zhu Shuzhen and Du Fu using Markov matrices by modeling the tonal patterns in their poetry. Each poem is encoded as a sequence of *even* or *oblique* tonal states, with transition probabilities captured in a matrix M , where M_{ij} denotes the probability of transitioning from state i to state j .

Using Markov Matrices to Distinguish Authors

For each poet, we construct a separate Markov matrix M_{Zhu} and M_{Du} from their corpus, defined as:

$$M_{\text{Zhu}} = \begin{bmatrix} p_{UU} & p_{UZ} \\ p_{ZU} & p_{ZZ} \end{bmatrix}, \quad M_{\text{Du}} = \begin{bmatrix} q_{UU} & q_{UZ} \\ q_{ZU} & q_{ZZ} \end{bmatrix}$$

where p_{ij} and q_{ij} are the probabilities of transitioning from tone i to tone j for Zhu Shuzhen and Du Fu, respectively. These matrices are expected to differ due to each poet's stylistic choices in adhering to classical tonal rules. The equilibrium vector \mathbf{v} for each matrix, satisfying $M\mathbf{v} = \mathbf{v}$, represents the long-term distribution of tonal states.

Validation and Classification

Given a test poem's tonal sequence $x = (x_1, x_2, \dots, x_n)$, we compute the *log-likelihood* under each author's matrix:

$$\mathcal{L}_{\text{Zhu}} = \sum_{i=1}^{n-1} \log M_{\text{Zhu}}[x_i, x_{i+1}], \quad \mathcal{L}_{\text{Du}} = \sum_{i=1}^{n-1} \log M_{\text{Du}}[x_i, x_{i+1}]$$

If $\mathcal{L}_{\text{Zhu}} > \mathcal{L}_{\text{Du}}$, we classify the poem as Zhu Shuzhen's. Alternatively, we can compare the test poem's equilibrium vector \mathbf{v}_{test} with \mathbf{v}_{Zhu} and \mathbf{v}_{Du} using similarity measures such as *cosine similarity* or *Euclidean distance*.

Identifying Authors or Genres

To identify an author, we compute the Euclidean distance:

$$d_{\text{Zhu}} = \|\mathbf{v}_{\text{test}} - \mathbf{v}_{\text{Zhu}}\|, \quad d_{\text{Du}} = \|\mathbf{v}_{\text{test}} - \mathbf{v}_{\text{Du}}\|$$

The author is identified as the one corresponding to the minimum distance. Moreover, differences in matrix structure or equilibrium distributions may also indicate genre differences, as classical genres often impose distinct tonal patterns.

Expectations

Given the strict tonal constraints in classical poetry and known stylistic differences, we expect to distinguish Du Fu and Zhu Shuzhen with reasonable accuracy. Referencing on their own style of writing, Du Fu's matrix should show a more balanced equilibrium, reflecting stricter adherence to tonal rules, while Zhu Shuzhen's may exhibit greater flexibility, creating distinct equilibrium patterns.