Project 1 Pre-port

Hanzhang Yin

October 2, 2024

The project aims to use Markov matrices to distinguish the styles of Zhu Shuzhen and Du Fu by modeling 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 sparate Markov matrix M_{Zhu} adn M_{Du} from their corpus, s.t.:

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

NOTE: denote even as "U" and oblique as "Z" in the matrix.

p_{ij} and q_{ij} are probabilities of transitioning from tone i to tone j for Zhu Shuzhen and Du Fu. respectively. We hypothesize that trhese amtrices will differe significantly, capturing unique stylistic patterns.

Validation and Classification:

For validation, given a test poem's sequecen of tones $x = (x_1, x_2, ..., x_n)$, we compute the *log-likelihood* under each matrix:

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

If the likelihood is significantly higher for one matrix, it indicates that the new text aligns more closely with that author's style. Moreover, the vector v_{Zhu}, v_{Du} represent the long-term tonal proportions for each authors. We can compare the equilibrium vector v of the test text's transition probabilities with the known equilibrium vectors of each author using similarity measures like **cosine similarity** or **Euclidean distance**.

Identifying Authors or Genres:

To identify an author, similar to the approach we have done in classification, we compute the distance between the test text's equilibrium distribution v_{test} and the equilibrium vectors v_{Zhu} and v_{Du}. The closest match suggests the likely author. Furthermore, differences in matrix structure or transition probabilities (flexible vs. rigid) can indicate genre differences as well by observation.

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 om 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.