

A4 Heuristics & Techniques

1. Smoothing in Training

According to this paper by Boodidhi, “smoothing can be used to improve the probability estimates” in case of “sparse training data” and “unseen words”. The paper also proved improved performance using laplace smoothing and absolute discounting smoothing methods.

Maximum Likelihood Estimation (MLE) in transition and emission probability calculations, we can smooth zero probability for unseen emission state combination. My implementation used laplace smoothing (or “add one smoothing”), to both transition and emission probabilities.

Result: accuracy is improved from basic HMM model (-d training1.txt -t test1.txt)

- Simple MLE: 197 out of 202
- MLE with Lapace smoothing: 198 out of 202

2. Improve Viterbi Implementation

- i). Use Numpy instead of for loop to implement Viterbi algorithm
 - a. Numpy is much faster due to its vectorized implementation
- ii). Save probabilities at time t instead of a matrix for all valid time t
 - a. We only need last probability sequence to calculate the new probability sequence.

Citation:

Boodidhi, S. (n.d.). Using smoothing techniques to improve the performance of Hidden Markov's Model.

<https://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=2008&context=thesesdissertations>