COMP9318 Project Report

Implementation of Viterbi algorithm (or Q1)

For achieving Q1, I define two classes named ‘TransitionProbability’ (be referred to as TP below) and ‘EmissionProbability’ (be referred to as EP below) and two functions named ‘tokens’ and ‘tool_read_file’. When TP and EP initialize, they will analyse the data file through calling tool_read_file function for obtaining f_1, f_2, f_3, *ID*s, *M* and *N*. Tokens would be called by tool_read_file function for divided tokens.

After analyze data file, we use Viterbi algorithm for final result of Q1. Viterbi algorithm referring from lecture receives five parameters sequence of observations, state spaces, initial probabilities, transition probabilities and emission probabilities. We consider the initial probabilities is the transition probabilities from BEGIN to each states.

Extension of Viterbi algorithm (Q2)

We define a list contain top k probabilities instead of float class. As we can store the top k probabilities of reaching each symbol. When sort the probabilities and two probabilities is the same, we prefer the smaller state. As we can satisfy ties.

Approach for advanced decoding (Q3)

Using k-nearest neighbors algorithm and add a variable in the into the formula of probability calculate.