

Group member:

Name: PEIGUO GUAN

zID: z5143964

Name: JIE SHANG

zID: z5153884

Project Report:

Firstly, we create the HMM class, and it is easy to set default value

And there are some functions like

readSS_file: use for read symbol file and state file, and return the content in list type

ReadQuery_file: use for read Query file and parse the address as the question required and return the list type

create_transaction_matrix: the content return from *readSS_file* will be processed to create *transaction_matrix*, the parameter *q3=False* and when it comes to *q3* it will be true to do some special change

create_emission_matrix: : the content return from *readSS_file* will be processed to create *transaction_matrix*, the parameter *q3=False* and when it comes to *q3* it will be true to do some special change

viterbi_path: use for Q1 and Q3. find the highest value of point and backtrack to find the optimal path.

TOP_KViterbi: use for Q2. input the np.array of the transaction matrix and emission matrix and pi, and observation list. Finding the topK value of the every point and set the rank of them. When we get to the last point, we need to backtrack the path and return the topK result.

Q1:

Set all the value and create a HMM model as *model*, and using *viterbi_path* function to return the optimal result.

Q2:

Set all the value and create a HMM model as *model*, it is more different and complex from Q1, since I found that it may be more easy to use np.array type instead of and using dict type. Using ***TOP_KViterbi*** to get the result. Finding the topK value of the every point and set the rank of them. When we get to the last point, we need to backtrace the path and return the topK result.

Q3:

The *create_transaction_matrix* and *create_emission_matrix* are a little different which is that set the parameter *q3* as True and change the probability of some of the value in the *transaction_matrix* and *emission_matrix* by going through the *Query_label*. And it is easy to find that some of the compose are frequently appeared although they are not appeared or relatively low frequency in test file. So we can increase some of their value to some level so that to meet the question required. Then set all the value and create a HMM model as *model*, and using *viterbi_path* function to return the optimal result.