COMP 550 Assignment 2

Jingyuan Wang 260860682

October 18, 2019

Question 1

1. The statement is true. Viterbi algorithm gives the global optimal solution to this problem of finding $argmaxQ_{1:T}P\{Q_{1:T}, O_{1:T}\}$.

$$\begin{aligned} P(Q_{1:T}|O_{1:T}) &= \frac{P(Q_{1:T}, O_{1:T})}{P(O_{1:T})} \\ &= \frac{P(O_{1:T}|Q_{1:T}) * P(Q_{1:T})}{P(O_{1:T})} \end{aligned} \tag{1}$$

Since each O_t only depends on Q_t and each Q_t depends on Q_{t-1}. Equation 1 is equivalent to:

$$\begin{aligned} &= \frac{\{\prod_{t=1}^{T} P(O_t|Q_t)\} * P(Q_1) * \{\prod_{t=1}^{T} P(Q_{t+1}|Q_t)\}}{P(O_{1:T})} \\ &\propto \{\prod_{t=1}^{T} P(O_t|Q_t)\} * P(Q_1) * \{\prod_{t=1}^{T} P(Q_{t+1}|Q_t)\} \\ &= P(Q_1) * (\prod_{t=1}^{T}\{P(O_t|Q_t)P(Q_t|Q_{t-1})\}) \end{aligned} \tag{2}$$

When performing Viterbi algorithm, we will find the maximum value of $P(O_t|Q_t)P(Q_t|Q_{t-1})$ at each time step t, the trace of which together contributes to a optimal Q_t path and thus optimal solution.

2.

$$\begin{aligned} -\sum_i logP(Y^i|X^i) &= log(\frac{1}{Z(x)} \exp \sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t)) \\ &= -\sum_i (\sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t) - log(Z(x)) \\ &= -\sum_i \Big(\sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t) - \sum_y \sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t)\Big) \end{aligned} \tag{3}$$

$$\begin{aligned} -\sum_i \sum_t log(p_t^i) &= -\sum_i \sum_t P(Y^t|x^t) \\ &= -\sum_i \sum_t log(\frac{1}{Z^t(x)} \exp \sum_k \theta_k f_k(y_t, y_{t-1}, x_t)) \\ &= -\sum_i (\sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t) - \sum_t log(Z^t(x))) \end{aligned} \tag{4}$$

where $Z^t(x) = \sum_y \exp \sum_k \theta_k f_k(y_t, y_{t-1}, x_t)$
Thus,

$$-\sum_i \sum_t log(p_t^i) = -\sum_i (\sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t) - \sum_t log(\sum_y \exp \sum_k \theta_k f_k(y_t, y_{t-1}, x_t))) \tag{5}$$

And formula 3 and 5 are obviously not equal, thus it is proved that $-\sum_i \sum_t log(p_t^i) \neq -\sum_i logP(Y^i|X^i)$

Question 2

1. Use CFG to model French is rather quick and convenient. It gives precise certain rules which we can code into programs and use it to examine if a sentence is grammatical with computations. The tree structures also provides a vivid representation of the grammar.

2. CFG could produce ambiguity in sentence explanations, i.e. it will probably generate two different trees give the same statement since French grammar does not ensure every string has a unique leftmost derivation or parse tree(also some French words such as 'Bon' has multiple lexical interpretations). Also when parsing languages like French which has many rules, it may undergo different attempts before successfully being parsed and will need more time to finish the task.

3. My CFG does not handle grammar groups such as adverbs, prepositions, does not support clauses, nor check with verb tenses or multiple adjectives. Moreover, it will not verify if the use of pronoun is ambiguous, etc.

Question 3

After implementing and experimenting with HMM model and other improvement methods, the following accuracy were achieved. When using a rather short pair of train text to establish a HMM model, it is very likely that the initial of a sentence with some characters and some transition combinations have never been observed and thus the according probabilities are zero. Consequently when adding laplace smoothing to cipher 1 and 2, less constraints were made(i.e. In the standard model, transitions did not occur in the training set can never happen in the test tags) and the performance were enhanced. After supplementing English character transitions in normal texts, it is surprising that the accuracy on cipher 2 increased significantly whereas decipher on the first set was not improved much. For the first two cipher set, I have achieved a fairly good model to decipher. By contrast, for the third set, none of the changes were helping improve the model with an accuracy of merely 20%, since in the third cipher, a ciphered character were no longer solely depending on a single original character, but depending on three adjacent words together, which makes it not a good idea to use HMM for such a deciphering task.

	Standard	With Laplace	With LM	With Laplace + LM
cipher1	9.88%	**97.66%**	19.58%	80.94%
cipher2	14.99%	83.28%	71.84%	**84.66%**
cipher3	21.40%	21.32%	21.14%	21.49%

Table 1: accuracy of HMM deciphering models on characters