Homework 3 of Genomics 2019

Dehe Wang

today

Homework

Question 1: Hidden Markov model

The hidden Markov model (HMM) is a common model used in biostatistics. In this example, we constructed a model for predicting gene splicing site using HMM (see attachment "model.pdf"). The genomic sequence and status sequence are in the file "Seq.txt". And you should save all probability number in log10 form with two significant digits.

1. Please calculate the probability of this sequence for given state path (10').
2. Please use the Viterbi algorithm to calculate the maximum probability matrix (10'), pointer matrix (10') of this model. And please calculate the most likely hidden state path (10') and the maximum probability of this state path (5').

Question 2: Sequencing technology

Please refer to the courseware Lec6 for an introduction to the research methods of alternative polyadenylation (10'), RNA structure (10') and m6A modification (10'). And choose one direction for the experimental design (25').

Please save the maximum probability matrix in "MaxProbMatrix.tsv", save the pointer matrix and most likely hidden state path in "PrinterMatrix.tsv", answer the probability of question one in the the given "StudentID_Answer.tsv" file, and write the answers for question two in "StudentID_Answer.txt". And please package the 4 files as "StudentID_HW3.zip".

You should answer in the format as in the given example. And you should submit your homework to Course website before Nov. 15, 2019.

Grading

The score S of this homework is calculated by the following formula:

$$S = (S_0 - P) * 0.97^d$$

where S_0 is the total score from all questions, P is a penalty and is obtained from the following table, and d is the late time (in day) calculated by the e-mail event stamp.

Table 1: Penalty list

Contents	Score
File naming error	-5' / file
Packaging error	-5'

Reference

1. **Hidden Markov model**: `https://doi.org/10.1093/bioinformatics/14.9.755`
2. **Viterbi algorithm**: `https://doi.org/10.1109/PROC.1973.9030`
3. **Wikipedia: Viterbi algorithm**: `https://en.wikipedia.org/wiki/Viterbi_algorithm`