

Hidden Markov Models

Parameter estimation

Deborah Thomas

Bioinformatics Computing

Computer Science & Engineering

October 13, 2005

Recap

- Hidden Markov Models: used to predict characteristics about samples
- Algorithms:
 - ▶ Forward algorithm: based on sequence up to now
 - ▶ Backward algorithm: based on sequence till the end
 - ▶ Viterbi algorithm: Most likely location based (path with highest probability)
- Assumption: we had the model with the probabilities associated with it

Where we are now

- Parameters are unknown and need to be estimated
- Unknowns:
 - ▶ A_{kl} = Transition probability
 - ▶ E_{kl} = Symbol probability
- Define:
 - ▶ Θ = Vector combining probabilities
 - ▶ Goal: Max $P(x|\Theta)$ over all Θ

Estimation - known sequence

- Example: CpG islands already labelled
- Number of each transition and emission can be counted

- Calculations:

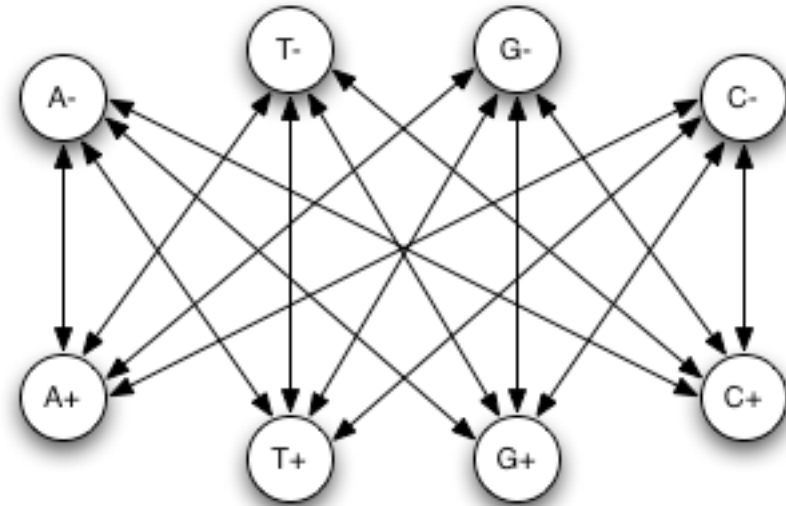
- ▶ Transitions:

$$a_{kl} = A_{kl} / (\sum_{q \in Q} A_{kq})$$

- ▶ Emissions:

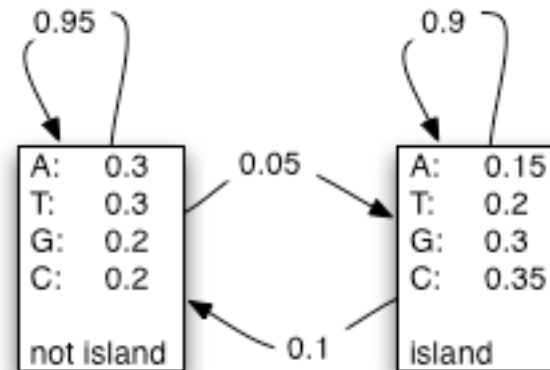
$$e_k(b) = E_k(b) / (\sum_{\sigma \in Q} E_k(\sigma))$$

- However, states aren't known - use Baum Welch algorithm



Baum-Welch training

- Iterative method
 - ▶ Uses current information and updates it with training sequences
- Calculates *expected* transitions and emissions
 - ▶ A_{kl} = number of transitions k to l in training data
 - ▶ $E_k(b)$ = number of emissions of b from k
- Forward and backward algorithm

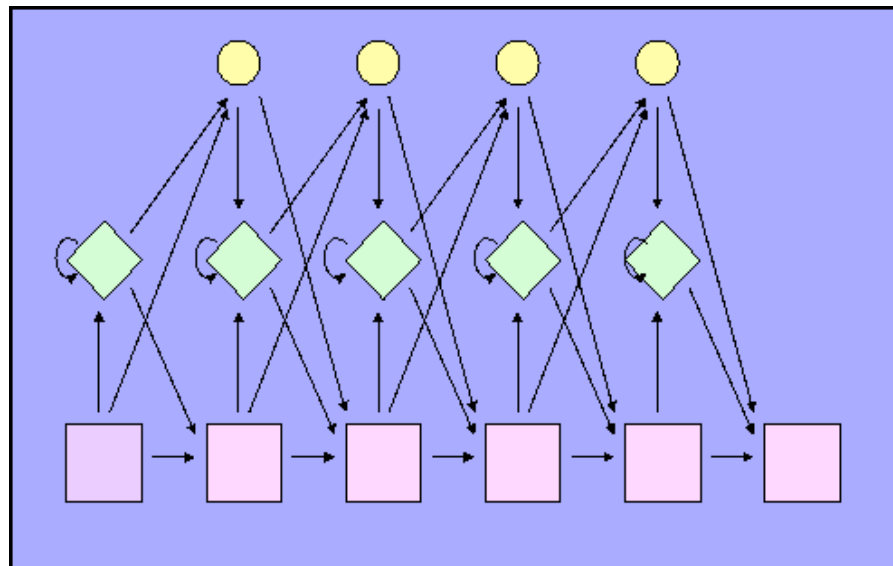


Algorithm

- Initialization: pick arbitrary parameters
- For each sequence $j = 1 \dots n$
 - ▶ Calculate $f_k(i)$ for sequence j - forward algorithm
 - ▶ Calculate $b_k(i)$ for sequence j - backward algorithm
 - ▶ Add the contribution of j to A and E
- Calculate the new model parameters
- Calculate the new log likelihood of model
- Termination:
 - ▶ When likelihood reaches a threshold
 - ▶ After certain number of iterations

Profile HMM alignment

- Pairwise alignments:
 - ▶ Fail to identify distant relations
- Model:
 - ▶ Include insertions and deletions
 - ▶ Transition b/w matching and insertion states:
Gap penalty



Demonstration

- Demonstration of Viterbi algorithm
- http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html_dev/viterbi_algorithm/s3_pg3.html
- Examples:
 - ▶ Damp, Soggy, Dryish, Damp
 - ▶ Dry, Damp, Soggy, Damp, Damp, Damp
 - ▶ Dry, Soggy, Dry, Soggy, Dry, Soggy

Resources

- SAM - Sequence Alignment and Modeling software system
 - ▶ Krogh et al. 1994, Hughey and Krogh 1996
 - ▶ <http://www.cse.ucsc.edu/research/compbio/sam.html>
- HMMER - Sequence analysis using profile hidden Markov Models
 - ▶ S. Eddy 1998
 - ▶ <http://hmmer.wustl.edu/>