

Hidden Markov Models

A selection of slides taken from the following:

Chris Bystroff – Protein Folding Initiation Site Motifs

Iosif Vaisman – Bioinformatics and Gene Discovery

Colin Cherry – Hidden Markov Models in Bioinformatics

Sequence = Structure

Structure = Function

Function = Life

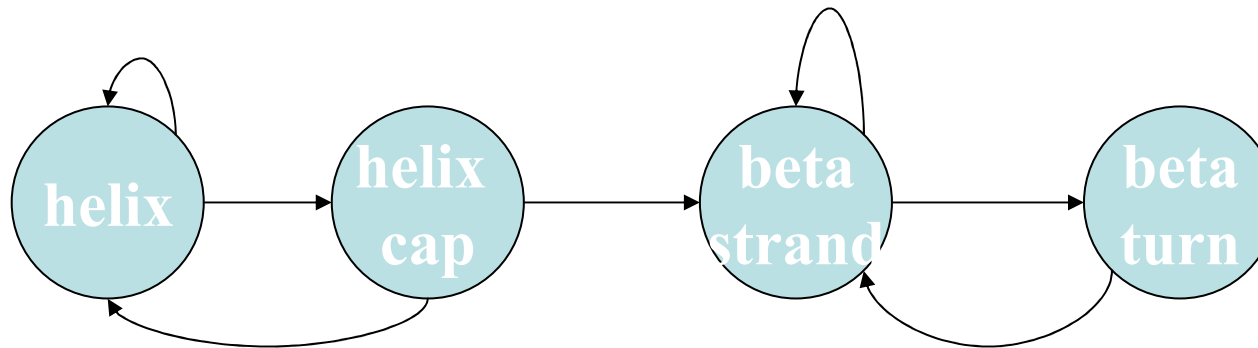
\therefore Sequence = Life

Protein sequence grammar

1. **Letters:** Amino acid profiles
2. **Words:** I-sites motifs
3. **Phrases:** a hidden Markov model
4. **Sentences**

Motif “grammar”?

Arrangement of I-sites motifs in proteins is highly non-random



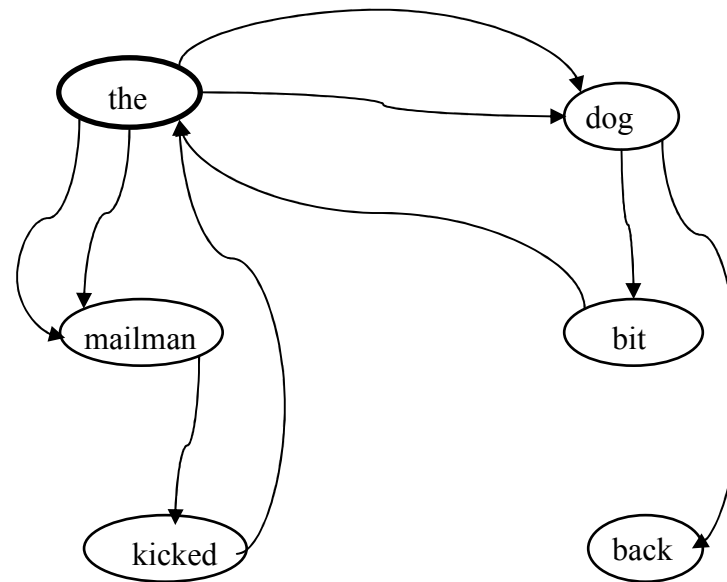
The dependencies can be modeled as a Markov chain

How to make a Markov chain

Sequence data

The dog bit the mailman. The mailman kicked the dog back.

Markov model



Stochastic output

The dog back. The mailman kicked the
mailman kicked the dog bit the dog bit
the dog bit the mailman kicked the
dog.

...

A "hidden" Markov model

What's "hidden" about it?

An HMM is a Markov chain where the meaning of the Markov state is *probabilistic*.

Basics of Hidden Markov Models

Probabilistic, generative, graphical models

HMMs consist of states and state transitions

Each state emits characters of an alphabet with characteristic probabilities

A sequence is generated by a random path through the model guided by the transition probabilities

HMMs are applied to problems where the state sequence is typically unknown.

Purpose of using an HMM:

- Sequence classification: is a given sequence generated by a given HMM ?
- Parsing: which residue of the sequence is generated by which part of the model ?

Specification of an HMM

- *N* - number of states
 - $Q = \{q_1; q_2; \dots; q_T\}$ - set of states
- *M* - the number of symbols (observables)
 - $O = \{o_1; o_2; \dots; o_T\}$ - set of symbols

Specification of an HMM

- *A - the state transition probability matrix*
 - $a_{ij} = P(q_{t+1} = j | q_t = i)$
- *B- observation probability distribution*
 - $b_j(k) = P(o_t = k | q_t = j) \quad i \leq k \leq M$
- *π - the initial state distribution*

Specification of an HMM

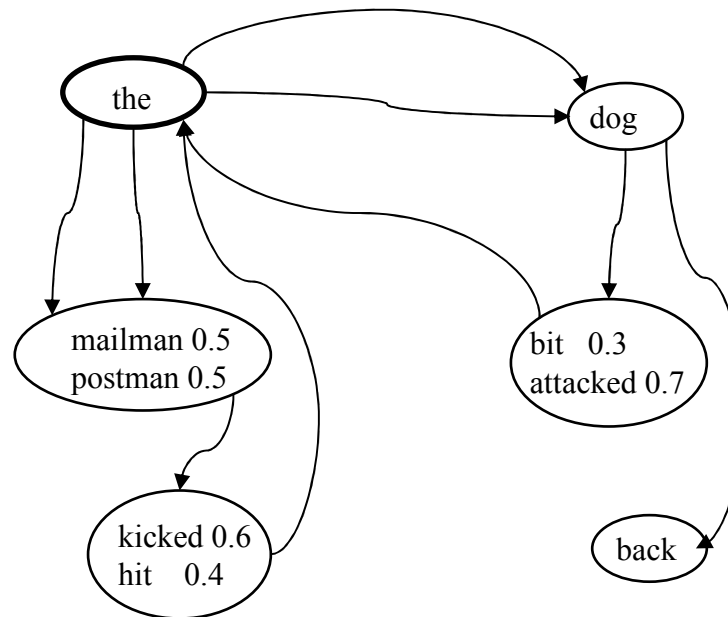
- Full HMM is thus specified as a triplet:
 - $\lambda = (A, B, \pi)$

How to make a *hidden* Markov chain

Sequence alignment data

The dog bit the mailman. The mailman kicked the dog back.
The dog attacked the postman. The postman hit the dog.

hidden Markov model

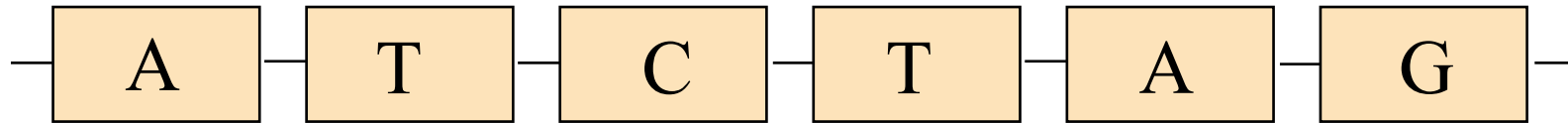


Stochastic output

The dog back. The mailman kicked the
postman kicked the dog bit the dog bit
the dog attacked the mailman kicked the
dog.

...

Markov Model (or Markov Chain)



Probability for each character based only on several preceding characters in the sequence

of preceding characters = *order* of the Markov Model

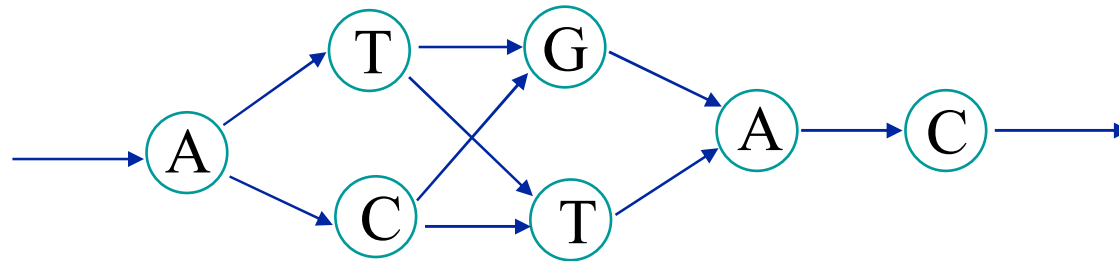
Probability of a sequence

$$P(s) = P[A] P[A,T] P[A,T,C] P[T,C,T] P[C,T,A] P[T,A,G]$$

Hidden Markov Models

States -- well defined conditions

Edges -- transitions between the states



ATGAC
ATTAC
ACGAC
ACTAC

Each transition assigned a probability.

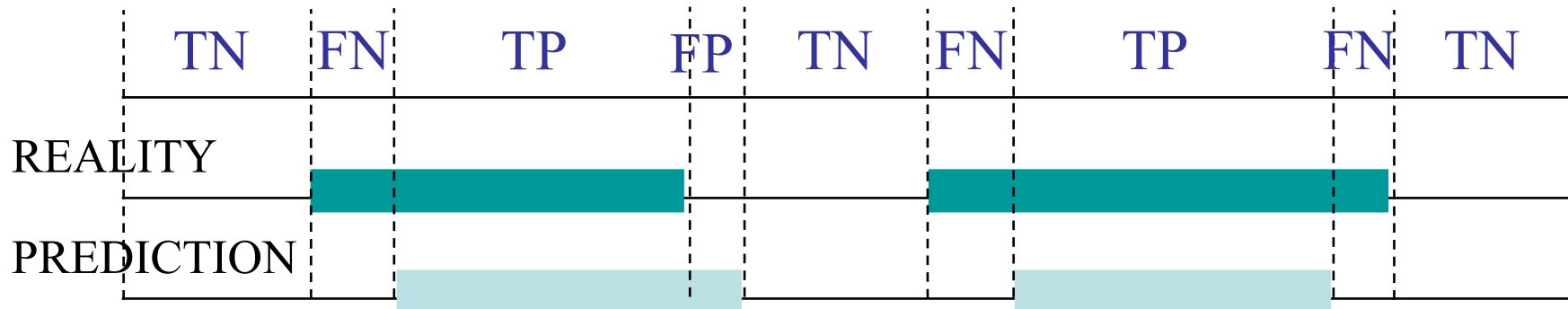
Probability of the sequence:

single path with the highest probability --- *Viterbi* path

sum of the probabilities over all paths -- *Baum-Welch* method

Measures of Prediction Accuracy

Nucleotide Level



		REALITY	
		c	nc
PREDICTION	c	TP	FP
	nc	FN	TN

Sensitivity

$$S_n = TP / (TP + FN)$$

Specificity

$$S_p = TP / (TP + FP)$$

HMM Advantages

- **Statistical Grounding**

- Statisticians are comfortable with the theory behind hidden Markov models
- Freedom to manipulate the training and verification processes
- Mathematical / theoretical analysis of the results and processes
- HMMs are still very powerful modeling tools – far more powerful than many statistical methods

HMM Advantages continued

- **Modularity**
 - HMMs can be combined into larger HMMs
- **Transparency of the Model**
 - Assuming an architecture with a good design
 - People can read the model and make sense of it
 - The model itself can help increase understanding

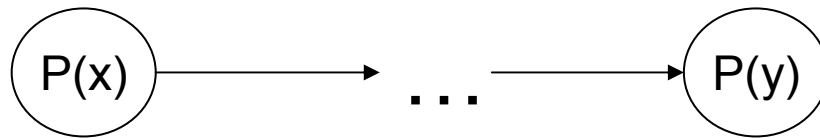
HMM Advantages continued

- **Incorporation of Prior Knowledge**
 - Incorporate prior knowledge into the architecture
 - Initialize the model close to something believed to be correct
 - Use prior knowledge to constrain training process

HMM Disadvantages

- **Markov Chains**

- States are supposed to be independent



- $P(y)$ must be independent of $P(x)$, and vice versa
- This usually isn't true
- Can get around it when relationships are local
- Not good for RNA folding problems

HMM Disadvantages continued

- **Standard Machine Learning Problems**
- Watch out for local maxima
 - Model may not converge to a truly optimal parameter set for a given training set
- Avoid over-fitting
 - You're only as good as your training set
 - More training is not always good

HMM Disadvantages continued

- **Speed!!!**

- Almost everything one does in an HMM involves: “enumerating all possible paths through the model”
- There are efficient ways to do this
- Still slow in comparison to other methods