Lecture 7

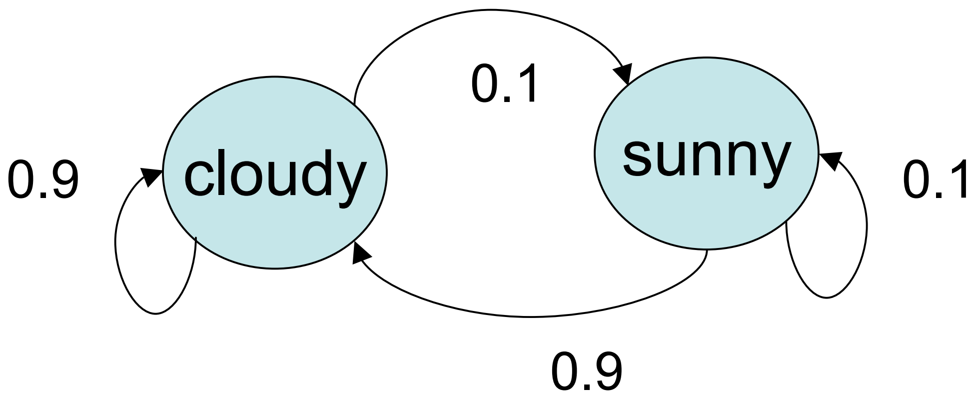
Intro to Hidden Markov Models

Markov models, revisited

- We can also view Markov models as a discrete (finite) system:
  - *N* total states
  - Start at some initial state (t = 1) based on multinomial model
  - System proceeds to the next state based on probabilities given current state
  - Also called a *probabilistic finite automata*

Simple example: South Bend winters

Tomorrow's weather



Another example

- Suppose we played a game as follows:
  - I give you $10 to start
  - We flip a coin; if it is heads you lose one dollar, if it is tails you gain a dollar
  - We stop when either you make $100 or lose all $10
- How could we model this using a simple Markov model? What does the program look like?

Barbecue begging

- A puppy smells a number of neighbors barbecuing. One unsupervised grill is two houses downhill from his yard, and another unsupervised grill is three houses uphill from his yard. Because so many people are barbecuing, he goes randomly from house to house in search of food, going downhill with twice the probability that he goes uphill. We record his progress from house to house, using 0 to stand for one unsupervised grill, 2 to stand for his yard, and 5 to stand for the other unsupervised grill.

www.mathforum.org

Hidden Markov Models

- Used when states can not directly be observed, good for noisy data
- Requirements:
  - A finite number of states, each with an output probability distribution
  - State transition probabilities
  - Observed phenomenon, which can be randomly generated given state-associated probabilities.

Example: play calling

- Suppose we simplify the ND offensive playbook into three plays:
  - Run
  - Pass short
  - Long pass
- Further, lets suppose there are two at most two offensive coaches:
  - Coach Weiss
  - Offensive coordinator Haywood

Play calling

- Coach Weiss:
  - P(run) = 0.1
  - P(short pass) = 0.1
  - P(long pass) = 0.8

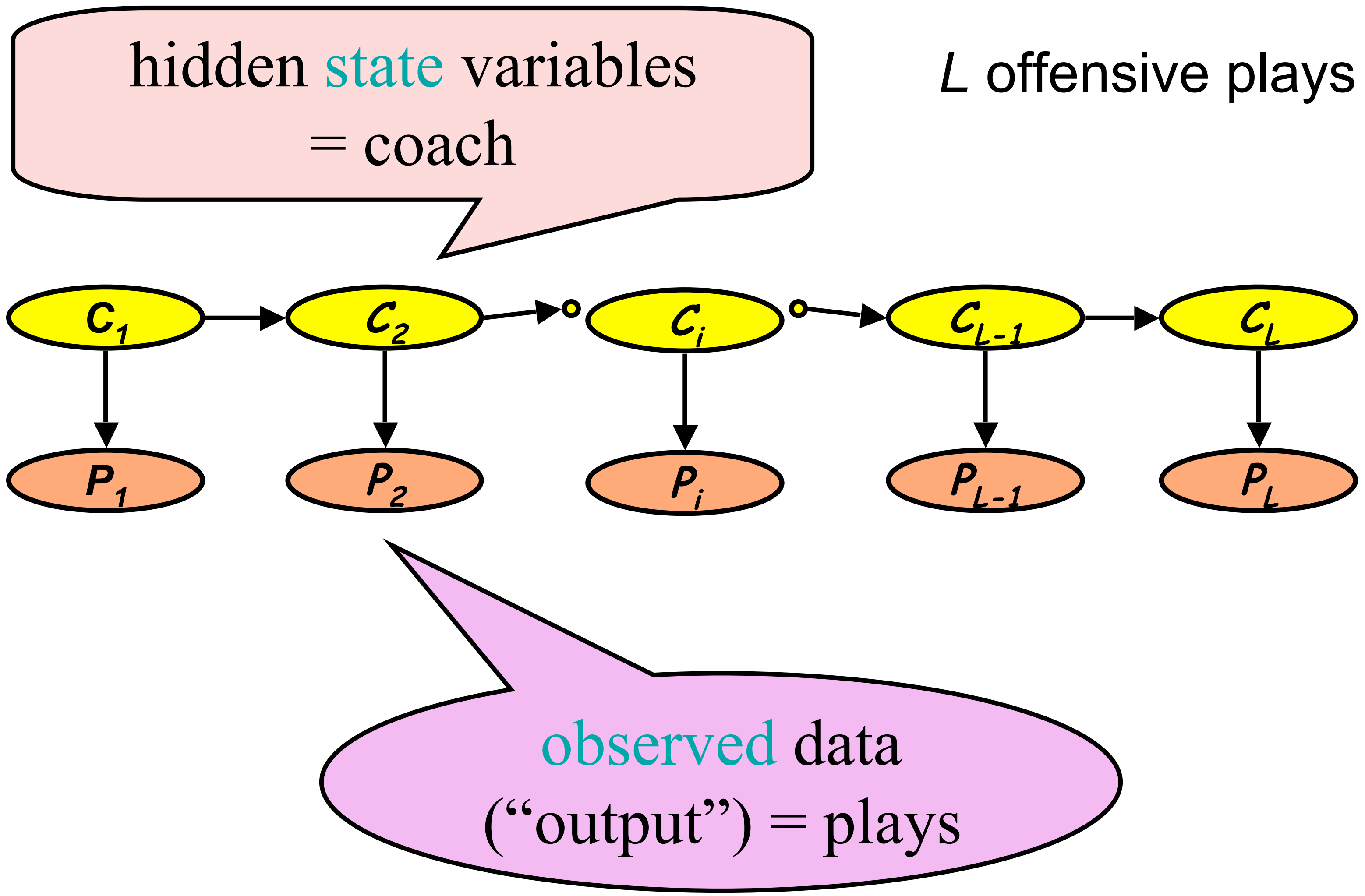
- Offensive coordinator Haywood:
  - P(run) = 0.8:
  - P(short pass) = 0.15
  - P(long pass) = 0.05

Hidden model

- As fans, we can not tell who is calling the plays, all we can observe is the play called
- We assume play calls are based on coach tendencies (output) probabilities

ND Football Game

L offensive plays

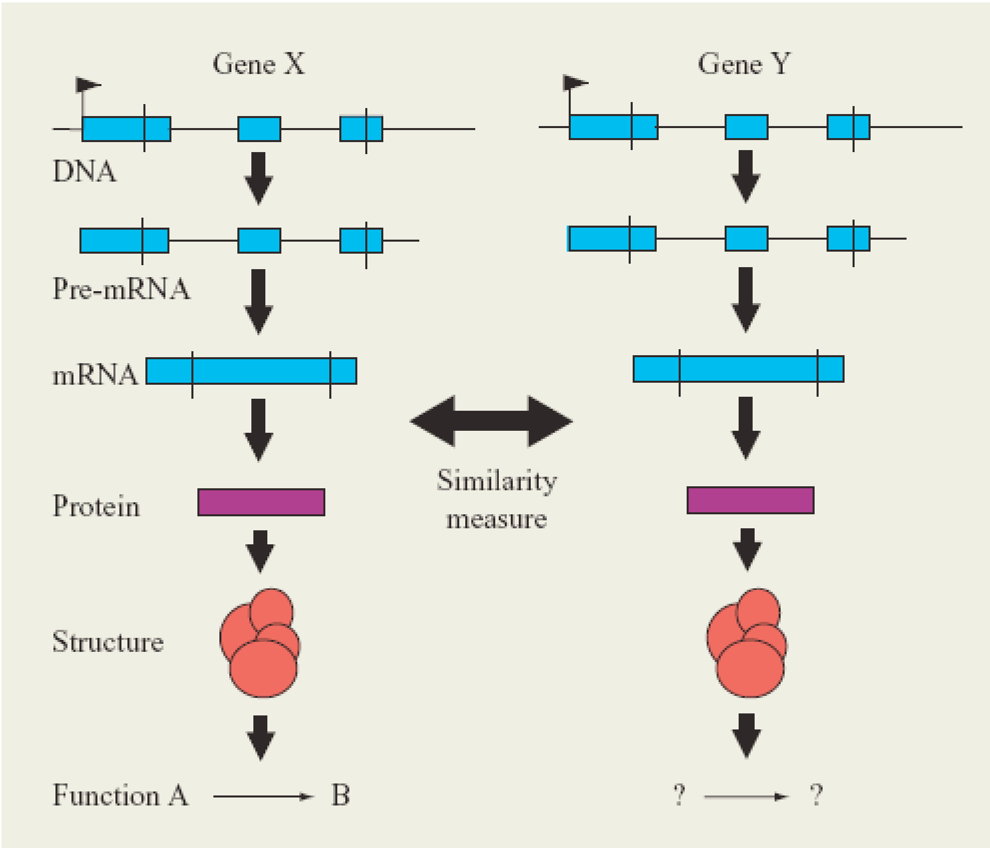


Other example goals

- Suppose we have a text written by Dante and a monkey. Can we tell who wrote what?

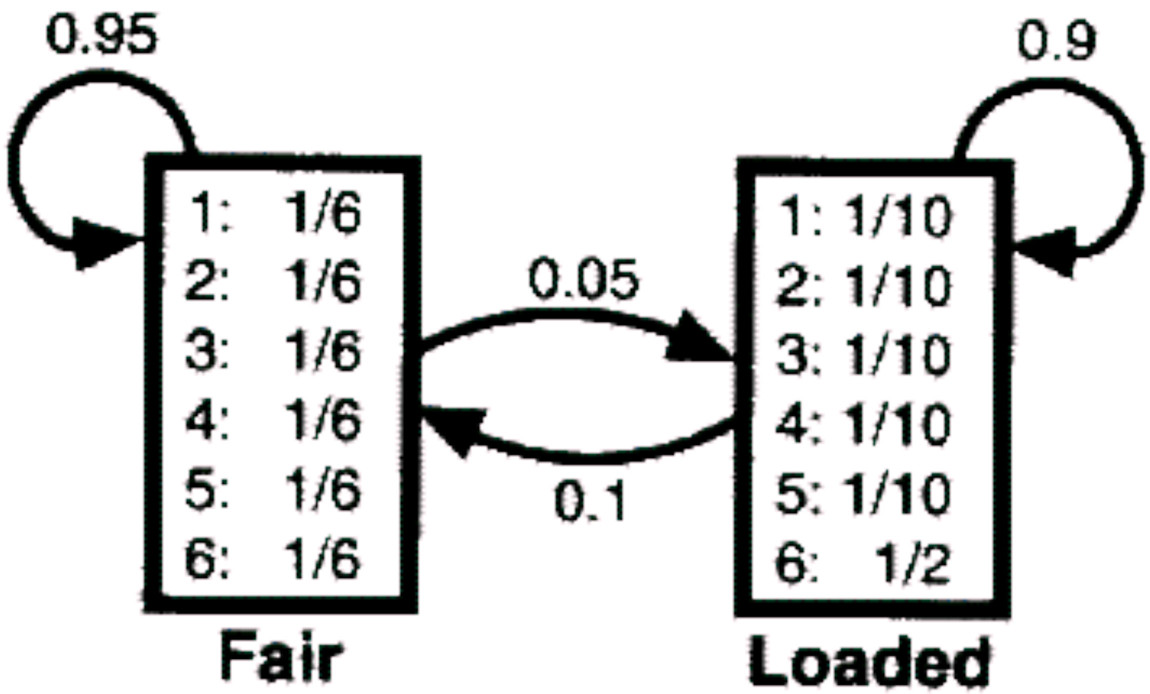
- More important: DNA sequences with coding and non-coding sequences. Can we discriminate boundaries?

Goal



(E. Birney, 2001)

Example: dishonest casino



From Durbin

Cases

Example	**Observations**	**Hidden state**
Football	Plays	Coach
Text	Words	Dante/ monkey
Casino	Rolled numbers	Fair/loaded
DNA	ACGT	Coding/not

Overview

- Structure
  - Number of states Q_1 .. Q_N
  - M output symbols
- Parameters:
  - Transition probability matrix a_{ij}
  - Emission probabilities $b_i(a)$, which is the probability state i emits character a
  - Initial distribution vector π_i

Notation (Rabiner)

- Let T be the number of observations
- Note T is also the number of states visited

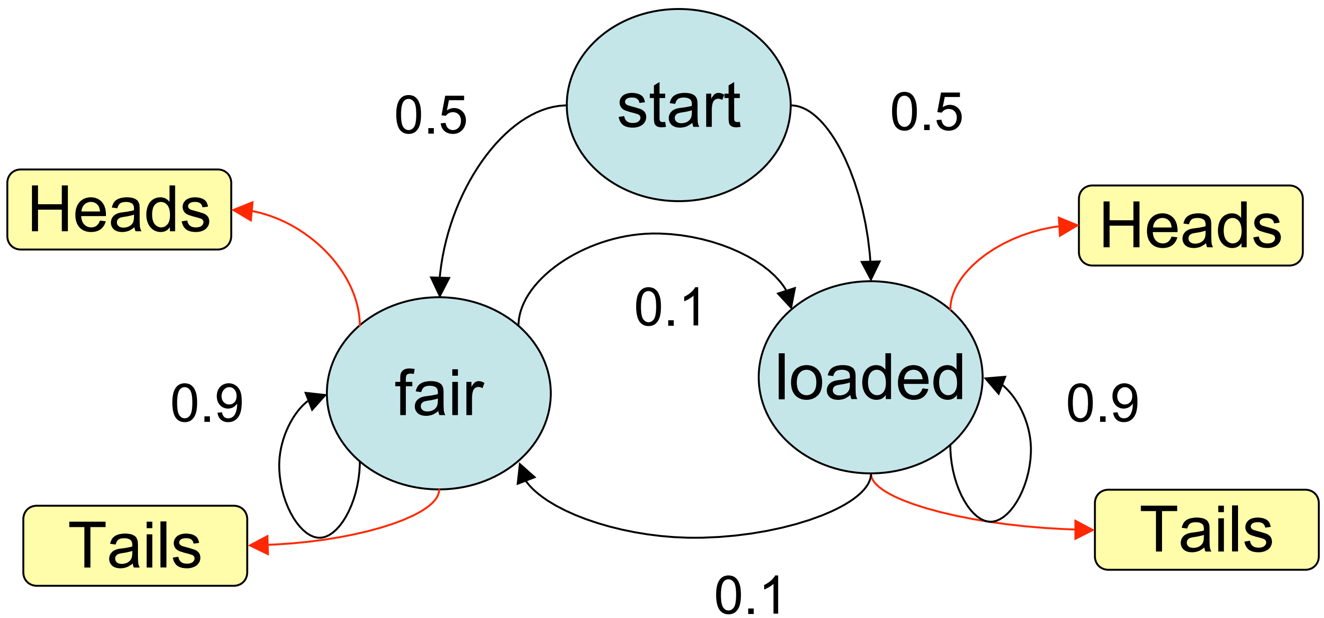
- Sequence of visited states:
  – $Q = q_1 q_2 q_3 q_4 \ldots q_T$

- Sequence of emitted symbols:
  – $O = O_1 O_2 O_3 O_4 \ldots O_T$

Model = $\lambda = \left\langle N, M, \{\pi_i\}, \{a_{ij}\}, \{b_i(j)\} \right\rangle$

Loaded coin

- Suppose in the earlier example in class I was also dishonest and used fair and loaded coins.
- How could we model this?


start
0.5
0.5
Heads
Heads
0.1
fair
loaded
0.9
0.9
Tails
Tails
0.1

Football example

- What is the HMM for the ND example mentioned previously?
- In-class

Assumptions

- Markov assumption
  - States depend on previous states
- Stationary assumption
  - Transition probabilities are independent of time (“memoryless”)
- Output independence
  - Observations are independent of previous observations

Basic problems

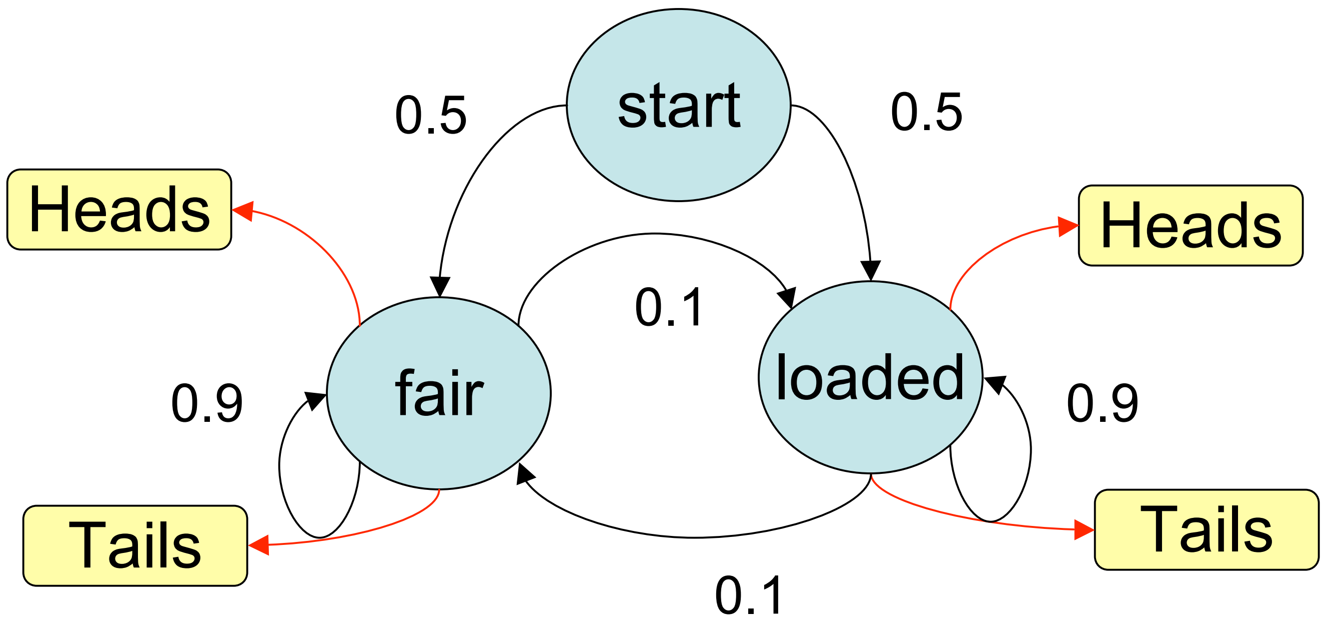
- Evaluation
  - What is the probability that the observations were generated by a given model?
- Decoding
  - Given a model and a sequence of observations, what is the most likely state observations?
- Learning:
  - Given a model and a sequence of observations, how should we modify the model parameters to maximize p{observe|model)

Decoding

- Text: Dante or Monkey?

- Case 1:
  - Fehwufhweuromeojulietpoisonjigjreijge
- Case 2:
  - mmmmbananammmmmmmbananammm

Question: Suppose the sequence of our game is:
HHHTHHHTTHHTH?



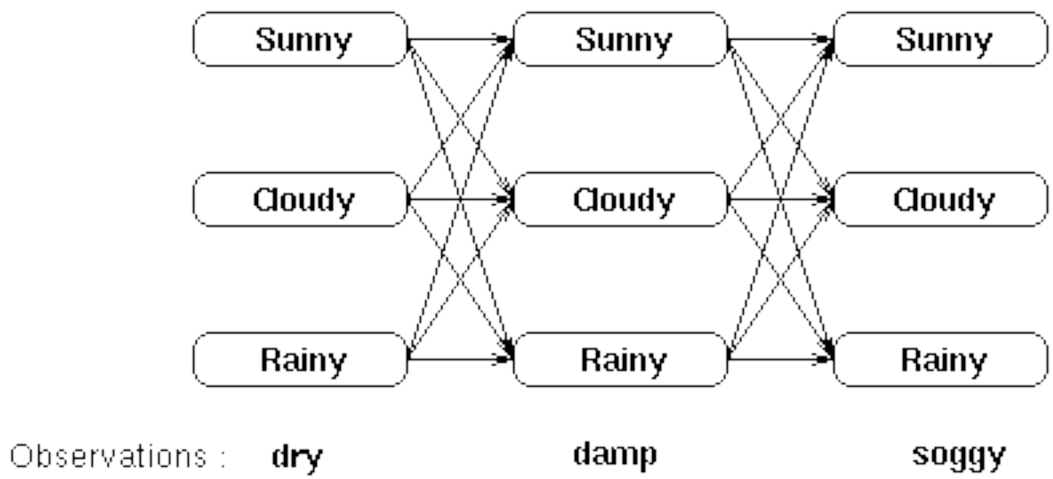
What is the probability of the sequence given the model?

Solutions

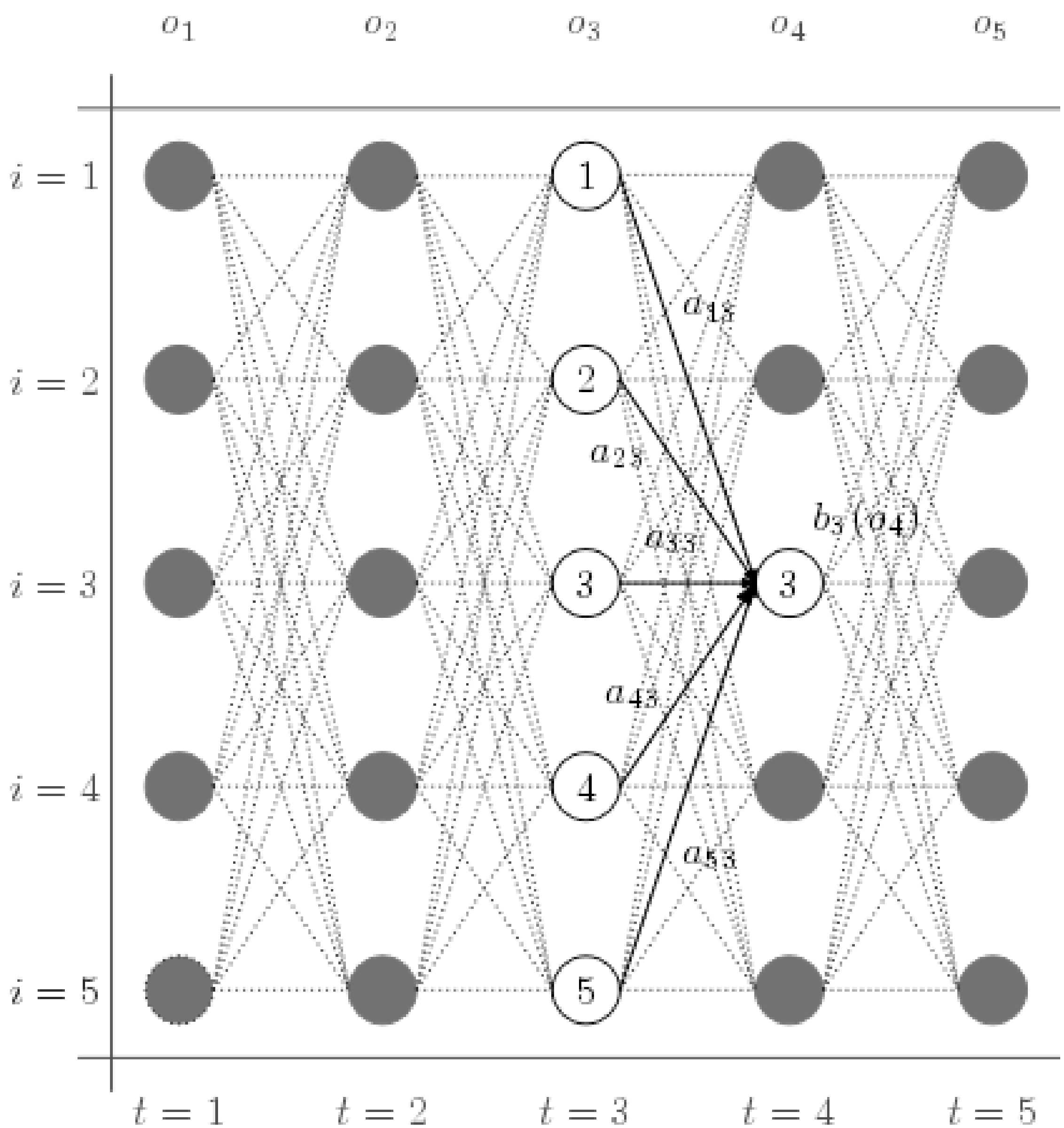
Problem	**Algorithm**	**Complexity**
Evaluation	Forward/ Backward	$O(TN^2)$
Decoding	Viterbi	$O(TN^2)$
Learning	Baum-Welch (EM)	$O(TN^2)$

T is # timesteps (or observations)   *N* = # states

Forward algorithm



The Trelis



Recursion

$$\alpha_i(t) = \begin{cases} 0 & : \quad t = 0 \wedge i \neq S_I \\ 1 & : \quad t = 0 \wedge i = S_I \\ \sum_j \alpha_j(t-1) a_{ji} b_{ji}(y) & : \quad t > 0 \end{cases}$$