Hidden Markov Models

Rafael E. Banchs (rbanchs@gps.tsc.upc.edu)

Session 2: Dynamic Programming & HMMs

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Review on HMMs

Last day we defined HMMs, and recalled that they can be formally specified by a set of five elements:

$$\{ \boldsymbol{S}, \ \boldsymbol{O}, \pi, \mathbf{P}, \mathbf{Q} \}$$

Review on HMMs

Last day we defined HMMs, and recalled that they can be formally specified by a set of five elements:

State Space → $$\{ S, O, \pi, P, Q \}$$

- where, ***S*** is the set of states: $S = \{s_1, s_2, s_3 \ldots s_m\}$;

Review on HMMs

Last day we defined HMMs, and recalled that they can be formally specified by a set of five elements:

Observation Space

$$\{ \boldsymbol{S}, \boxed{\boldsymbol{O},} \pi, \mathbf{P}, \mathbf{Q} \}$$

- where, $\boldsymbol{S}$ is the set of states: $\boldsymbol{S} = \{s_1, s_2, s_3 \ldots s_m\}$;
- $\boldsymbol{O}$ is the observation alphabet: $\boldsymbol{O} = \{o_1, o_2, o_3 \ldots o_h\}$;

Review on HMMs

Last day we defined HMMs, and recalled that they can be formally specified by a set of five elements:

Model Space: λ

$$\{\, \boldsymbol{S},\ \boldsymbol{O},\ \pi,\ \mathbf{P},\ \mathbf{Q} \,\}$$

- where, ***S*** is the set of states: $\boldsymbol{S} = \{s_1, s_2, s_3 \ldots s_m\}$;
- ***O*** is the observation alphabet: $\boldsymbol{O} = \{o_1, o_2, o_3 \ldots o_h\}$;
- and, π, **P** and **Q** are the HMM associated probabilities.

Review on HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_N$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

Review on HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_N$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

- Observation sequence: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_N$
- Observation alphabet (symbols): $\boldsymbol{O} = \{o_1 , o_2 , o_3 \ldots o_h\}$

Review on HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_N$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

- Observation sequence: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_N$
- Observation alphabet (symbols): $\boldsymbol{O} = \{o_1 , o_2 , o_3 \ldots o_h\}$

- Initial state probabilities: $\pi = \{\pi_i\}$ with $i \in \boldsymbol{S}$
- State transition probabilities: $\mathbf{P} = \{p_{ij}\}$ with $i, j \in \boldsymbol{S}$

Review on HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_N$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

- Observation sequence: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_N$
- Observation alphabet (symbols): $\boldsymbol{O} = \{o_1 , o_2 , o_3 \ldots o_h\}$

- Initial state probabilities: $\pi = \{\pi_i\}$ with $i \in \boldsymbol{S}$
- State transition probabilities: $\mathbf{P} = \{p_{ij}\}$ with $i, j \in \boldsymbol{S}$

- Symbol emission probabilities: $\mathbf{Q} = \{q_{ijk}\}$ with $i, j \in \boldsymbol{S}$, $k \in \boldsymbol{O}$

Review on HMMs

- π is a vector of m elements which represent the initial probability distribution of states.

Review on HMMs

- π is a vector of m elements which represent the initial probability distribution of states.
- **P** is an $m{\times}m$ matrix with elements $p_{ij}= P(X_n=s_j|X_{n-1}=s_i)$ which represent the probability distribution of state transitions.

Review on HMMs

- π is a vector of m elements which represent the initial probability distribution of states.

- **P** is an mxm matrix with elements $p_{ij} = P(X_n = s_j | X_{n-1} = s_i)$ which represent the probability distribution of state transitions.

- **Q** constitutes the observation probability distribution that is associated to state transitions:

$$q_{ijk} = q(Y_n \mid X_n, X_{n-1}) = P(Y_n = o_k \mid X_n = s_j , X_{n-1} = s_i)$$

Probability of an observation sequence

The problem we are going to deal with today is the following:

- Consider a HMM defined as $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, what is the probability $P(Y|\lambda)$ of a given observation sequence $\boldsymbol{Y} = Y_1, Y_2, Y_3 \ldots Y_N$?

Probability of an observation sequence

The problem we are going to deal with today is the following:

- Consider a HMM defined as $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, what is the probability $P(Y|\ \lambda)$ of a given observation sequence $\boldsymbol{Y} = Y_1, Y_2, Y_3 \ldots Y_N$?
- $P(Y|\ \lambda)$ is computed by adding up all probabilities for generating the observation sequence through any possible state sequence.

Probability of an observation sequence

The problem we are going to deal with today is the following:

- Consider a HMM defined as $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, what is the probability $P(Y|\lambda)$ of a given observation sequence $\boldsymbol{Y} = Y_1, Y_2, Y_3 \ldots Y_N$?
- $P(Y|\lambda)$ is computed by adding up all probabilities for generating the observation sequence through any possible state sequence.
- Calculating $P(Y|\lambda)$ is very expensive from the computational point of view, and it can be efficiently done by using ***dynamic programming*** techniques.

Mathematical formulation of the problem

$$P(Y\,|\,\lambda) = \sum_{\mathbf{X}} P(Y, X\,|\,\lambda)$$

Mathematical formulation of the problem

observation sequence

given model

$$P(Y\,|\,\lambda) = \sum_{X} P(Y, X\,|\,\lambda)$$

Mathematical formulation of the problem

observation sequence

given model

joint probability of an observation sequence and a state sequence for the given model

$$P(Y\,|\,\lambda) = \sum_{X} \overbrace{P(Y, X\,|\,\lambda)}$$

all possible state sequences

Mathematical formulation of the problem

observation sequence

given model

joint probability of an observation sequence and a state sequence for the given model

$$P(Y \mid \lambda) = \sum_{X} \overbrace{P(Y, X \mid \lambda)} = \sum_{X} P(Y \mid X, \lambda)\, P(X \mid \lambda)$$

all possible state sequences

Mathematical formulation of the problem

observation sequence

given model

joint probability of an observation sequence and a state sequence for the given model

$$P(Y\,|\,\lambda) = \sum_{X} \overbrace{P(Y, X\,|\,\lambda)} = \sum_{X} \underbrace{P(Y\,|X, \lambda)}\, P(X\,|\,\lambda)$$

all possible state sequences

observation probability for a given state sequence

Mathematical formulation of the problem

observation sequence

given model

$$P(Y|\lambda) = \sum_{X} \overbrace{P(Y, X|\lambda)}^{\text{joint probability of an observation sequence and a state sequence for the given model}} = \sum_{X} \underbrace{P(Y|X, \lambda)}_{\text{observation probability for a given state sequence}} P(X|\lambda)$$

all possible state sequences

$$P(Y|X, \lambda) = q_{X_0X_1Y_1}\, q_{X_1X_2Y_2}\, q_{X_2X_3Y_3} \dots\; q_{X_{N-1}X_NY_N}$$

⇨ *A product of symbol emission probabilities*

Mathematical formulation of the problem

observation sequence

given model

joint probability of an observation sequence and a state sequence for the given model

$$P(Y\,|\lambda) = \sum_{X} \overbrace{P(Y, X\,|\lambda)} = \sum_{X} \underbrace{P(Y\,|X, \lambda)}\,\underbrace{P(X\,|\lambda)}$$

all possible state sequences

observation probability for a given state sequence

state sequence probability

$$P(Y\,|X, \lambda) = q_{X_0X_1Y_1}\, q_{X_1X_2Y_2}\, q_{X_2X_3Y_3} \dots\; q_{X_{N-1}X_NY_N}$$

➡ *A product of symbol emission probabilities*

Mathematical formulation of the problem

observation sequence

given model

joint probability of an observation sequence and a state sequence for the given model

$$P(Y\,|\lambda) = \sum_{X} \overbrace{P(Y, X\,|\lambda)} = \sum_{X} \underbrace{P(Y\,|X, \lambda)}\,\underbrace{P(X\,|\lambda)}$$

all possible state sequences

observation probability for a given state sequence

state sequence probability

$$P(Y\,|X, \lambda) = q_{X_0X_1Y_1}\, q_{X_1X_2Y_2}\, q_{X_2X_3Y_3} \cdots\; q_{X_{N-1}X_NY_N}$$

⇨ ***A product of symbol emission probabilities***

$$P(X\,|\lambda) = \pi\, p_{X_0X_1} p_{X_1X_2} p_{X_2X_3} p_{X_3X_4} \cdots\; P_{X_{N-1}X_N}$$

⇨ ***A markov chain evolution probability***

Exercise 2 revisited

Consider the same Markov Chain from exercise 1:

- State alphabet: $\boldsymbol{S} = \{s_1, s_2, s_3\}$
- Initial probability distribution: $\pi = [1, 0, 0]$
- Transition matrix: $\mathbf{P} = \begin{pmatrix} 0.2 & 0.8 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$

Exercise 2 revisited

But now, consider the binary observation alphabet: $\boldsymbol{O} = \{A\,,\, B\}$

And the following symbol emission probabilities:

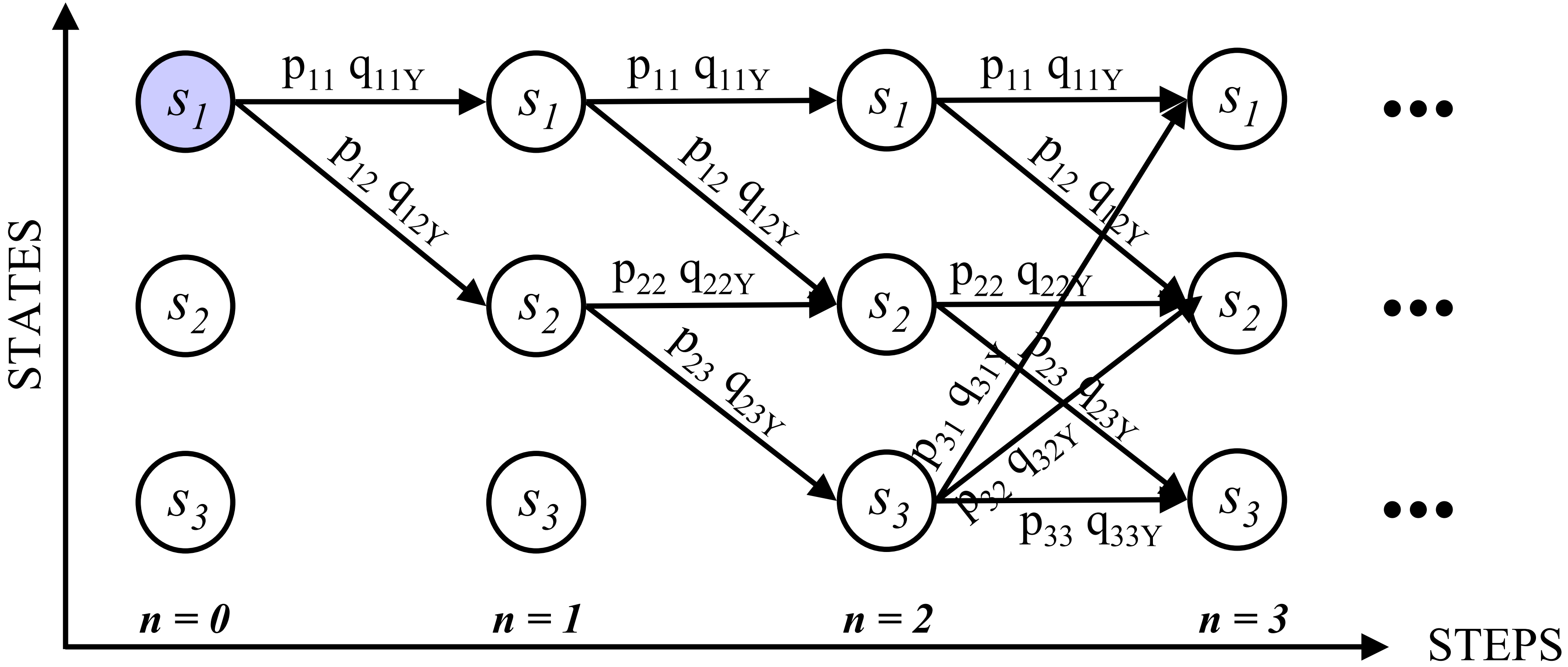
$$q_{ijA} = q(\,A \mid X_n,\, X_{n\text{-}1}) \qquad q_{ijB} = q(\,B \mid X_n,\, X_{n\text{-}1})$$

$$\begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix} \qquad \begin{pmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.0 \end{pmatrix}$$

What is the probability of observing the sequence $\boldsymbol{Y} = A,\, A,\, A$?

Lattice representation for HMMS

Consider the following representation for the HMM in exercise 2



Computation of probabilities

The probability of any observation sequence ***Y*** is computed by:

- considering the corresponding symbol emission probabilities at each step in the sequence,

Computation of probabilities

The probability of any observation sequence $\boldsymbol{Y}$ is computed by:

- considering the corresponding symbol emission probabilities at each step in the sequence,
- computing the probabilities for all possible paths starting at $n=0$ and ending at $n=N$ (being N the size of sequence $\boldsymbol{Y}$),

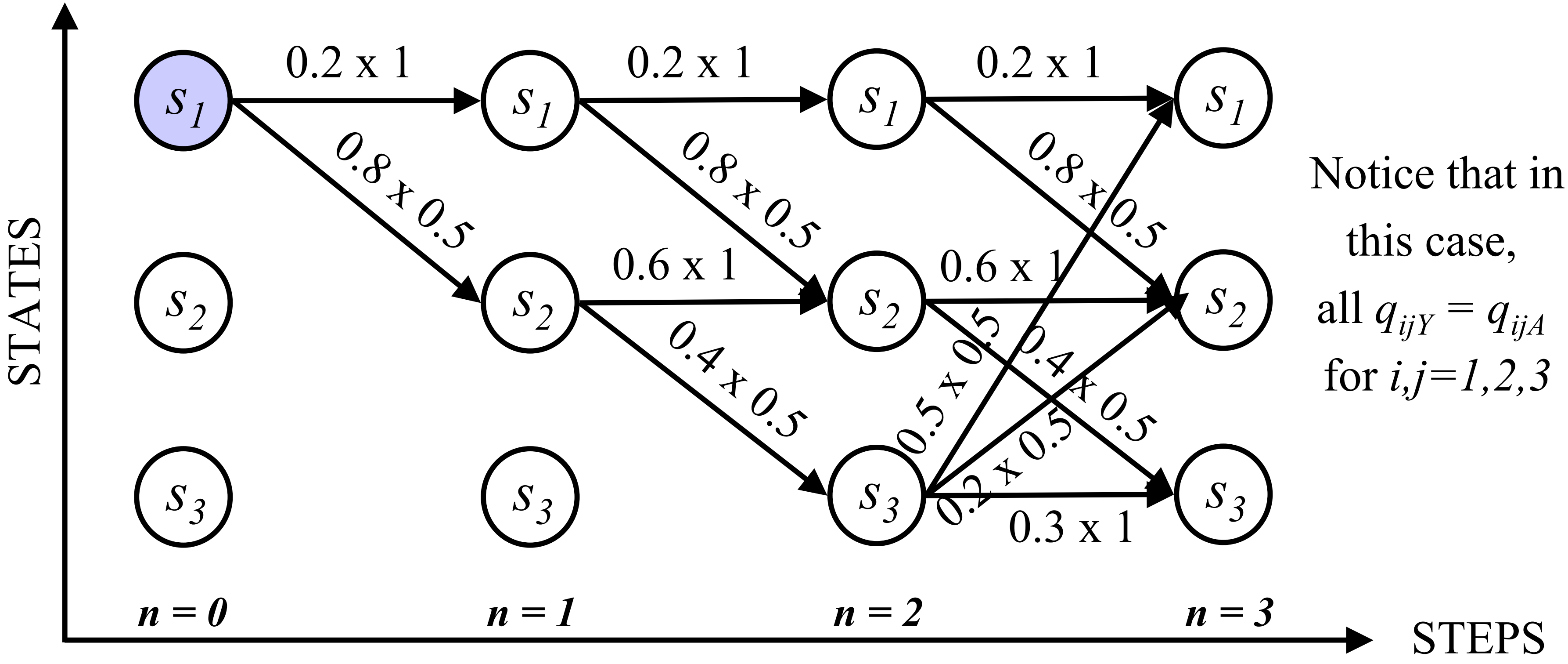
Computation of probabilities

The probability of any observation sequence Y is computed by:

- considering the corresponding symbol emission probabilities at each step in the sequence,
- computing the probabilities for all possible paths starting at $n=0$ and ending at $n=N$ (being N the size of sequence Y),
- adding up all resulting probabilities.

Computation of probabilities

So, in the particular case of $\boldsymbol{Y} = A, A, A$



Notice that in this case, all $q_{ijY} = q_{ijA}$ for $i,j=1,2,3$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 = 0.016$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 = 0.016$

$S_1\ S_1\ S_2\ S_2$: $p(path_3) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 = 0.048$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 = 0.016$

$S_1\ S_1\ S_2\ S_2$: $p(path_3) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 = 0.048$

$S_1\ S_1\ S_2\ S_3$: $p(path_4) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 = 0.016$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 = 0.016$

$S_1\ S_1\ S_2\ S_2$: $p(path_3) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 = 0.048$

$S_1\ S_1\ S_2\ S_3$: $p(path_4) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 = 0.016$

$S_1\ S_2\ S_2\ S_2$: $p(path_5) = 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 \text{ x } 0.6 \text{ x } 1.0 = 0.144$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 = 0.016$

$S_1\ S_1\ S_2\ S_2$: $p(path_3) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 = 0.048$

$S_1\ S_1\ S_2\ S_3$: $p(path_4) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 = 0.016$

$S_1\ S_2\ S_2\ S_2$: $p(path_5) = 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 \text{ x } 0.6 \text{ x } 1.0 = 0.144$

$S_1\ S_2\ S_2\ S_3$: $p(path_6) = 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 \text{ x } 0.4 \text{ x } 0.5 = 0.048$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \times 1.0 \times 0.2 \times 1.0 \times 0.2 \times 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2 \times 1.0 \times 0.2 \times 1.0 \times 0.8 \times 0.5 = 0.016$

$S_1\ S_1\ S_2\ S_2$: $p(path_3) = 0.2 \times 1.0 \times 0.8 \times 0.5 \times 0.6 \times 1.0 = 0.048$

$S_1\ S_1\ S_2\ S_3$: $p(path_4) = 0.2 \times 1.0 \times 0.8 \times 0.5 \times 0.4 \times 0.5 = 0.016$

$S_1\ S_2\ S_2\ S_2$: $p(path_5) = 0.8 \times 0.5 \times 0.6 \times 1.0 \times 0.6 \times 1.0 = 0.144$

$S_1\ S_2\ S_2\ S_3$: $p(path_6) = 0.8 \times 0.5 \times 0.6 \times 1.0 \times 0.4 \times 0.5 = 0.048$

$S_1\ S_2\ S_3\ S_1$: $p(path_7) = 0.8 \times 0.5 \times 0.4 \times 0.5 \times 0.5 \times 0.5 = 0.020$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2\ x\ 1.0\ x\ 0.2\ x\ 1.0\ x\ 0.2\ x\ 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2\ x\ 1.0\ x\ 0.2\ x\ 1.0\ x\ 0.8\ x\ 0.5 = 0.016$

$S_1\ S_1\ S_2\ S_2$: $p(path_3) = 0.2\ x\ 1.0\ x\ 0.8\ x\ 0.5\ x\ 0.6\ x\ 1.0 = 0.048$

$S_1\ S_1\ S_2\ S_3$: $p(path_4) = 0.2\ x\ 1.0\ x\ 0.8\ x\ 0.5\ x\ 0.4\ x\ 0.5 = 0.016$

$S_1\ S_2\ S_2\ S_2$: $p(path_5) = 0.8\ x\ 0.5\ x\ 0.6\ x\ 1.0\ x\ 0.6\ x\ 1.0 = 0.144$

$S_1\ S_2\ S_2\ S_3$: $p(path_6) = 0.8\ x\ 0.5\ x\ 0.6\ x\ 1.0\ x\ 0.4\ x\ 0.5 = 0.048$

$S_1\ S_2\ S_3\ S_1$: $p(path_7) = 0.8\ x\ 0.5\ x\ 0.4\ x\ 0.5\ x\ 0.5\ x\ 0.5 = 0.020$

$S_1\ S_2\ S_3\ S_2$: $p(path_8) = 0.8\ x\ 0.5\ x\ 0.4\ x\ 0.5\ x\ 0.2\ x\ 0.5 = 0.008$

Computation of probabilities

A total of nine different paths are possible:

$S_1\ S_1\ S_1\ S_1$: $p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008$

$S_1\ S_1\ S_1\ S_2$: $p(path_2) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 = 0.016$

$S_1\ S_1\ S_2\ S_2$: $p(path_3) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 = 0.048$

$S_1\ S_1\ S_2\ S_3$: $p(path_4) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 = 0.016$

$S_1\ S_2\ S_2\ S_2$: $p(path_5) = 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 \text{ x } 0.6 \text{ x } 1.0 = 0.144$

$S_1\ S_2\ S_2\ S_3$: $p(path_6) = 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 \text{ x } 0.4 \text{ x } 0.5 = 0.048$

$S_1\ S_2\ S_3\ S_1$: $p(path_7) = 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 \text{ x } 0.5 \text{ x } 0.5 = 0.020$

$S_1\ S_2\ S_3\ S_2$: $p(path_8) = 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 \text{ x } 0.2 \text{ x } 0.5 = 0.008$

$S_1\ S_2\ S_3\ S_3$: $p(path_9) = 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 \text{ x } 0.3 \text{ x } 1.0 = 0.024$

Computation of probabilities

A total of nine different paths are possible:

$$\left.\begin{array}{l}
S_1\ S_1\ S_1\ S_1:\ p(path_1) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 = 0.008 \\
S_1\ S_1\ S_1\ S_2:\ p(path_2) = 0.2 \text{ x } 1.0 \text{ x } 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 = 0.016 \\
S_1\ S_1\ S_2\ S_2:\ p(path_3) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 = 0.048 \\
S_1\ S_1\ S_2\ S_3:\ p(path_4) = 0.2 \text{ x } 1.0 \text{ x } 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 = 0.016 \\
S_1\ S_2\ S_2\ S_2:\ p(path_5) = 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 \text{ x } 0.6 \text{ x } 1.0 = 0.144 \\
S_1\ S_2\ S_2\ S_3:\ p(path_6) = 0.8 \text{ x } 0.5 \text{ x } 0.6 \text{ x } 1.0 \text{ x } 0.4 \text{ x } 0.5 = 0.048 \\
S_1\ S_2\ S_3\ S_1:\ p(path_7) = 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 \text{ x } 0.5 \text{ x } 0.5 = 0.020 \\
S_1\ S_2\ S_3\ S_2:\ p(path_8) = 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 \text{ x } 0.2 \text{ x } 0.5 = 0.008 \\
S_1\ S_2\ S_3\ S_3:\ p(path_9) = 0.8 \text{ x } 0.5 \text{ x } 0.4 \text{ x } 0.5 \text{ x } 0.3 \text{ x } 1.0 = 0.024
\end{array}\right\} \begin{array}{l} P(A,A,A) \\ = 0.332 \end{array}$$

Computational complexity

In the general case of an observation sequence of size N and a HMM of M states, the total number of multiplications to be performed is: $(2N+1)\ M^{N+1}$

Computational complexity

In the general case of an observation sequence of size N and a HMM of M states, the total number of multiplications to be performed is: $(2N+1)\ M^{N+1}$

- This problem becomes intractable for large values of N and M.

Computational complexity

In the general case of an observation sequence of size N and a HMM of M states, the total number of multiplications to be performed is: $(2N+1)\ M^{N+1}$

- This problem becomes intractable for large values of N and M.
- However, it can be noticed that in the process there are some products that are repeated again and again...

Computational complexity

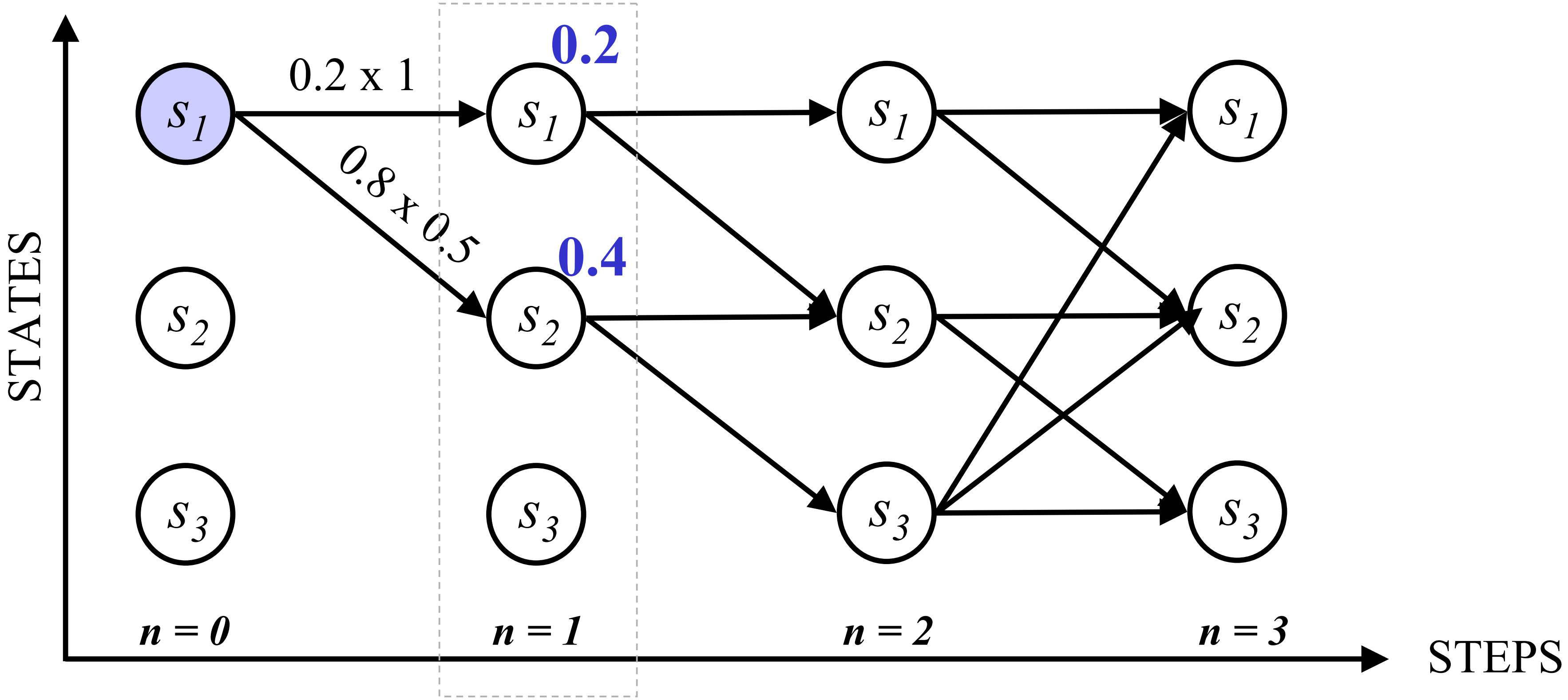
In the general case of an observation sequence of size N and a HMM of M states, the total number of multiplications to be performed is: $(2N+1)\, M^{N+1}$

- This problem becomes intractable for large values of N and M.
- However, it can be noticed that in the process there are some products that are repeated again and again...
- ***Dynamic programming*** consists in memorizing some partial results in order to avoid repeating the same computations !!!

Dynamic Programming

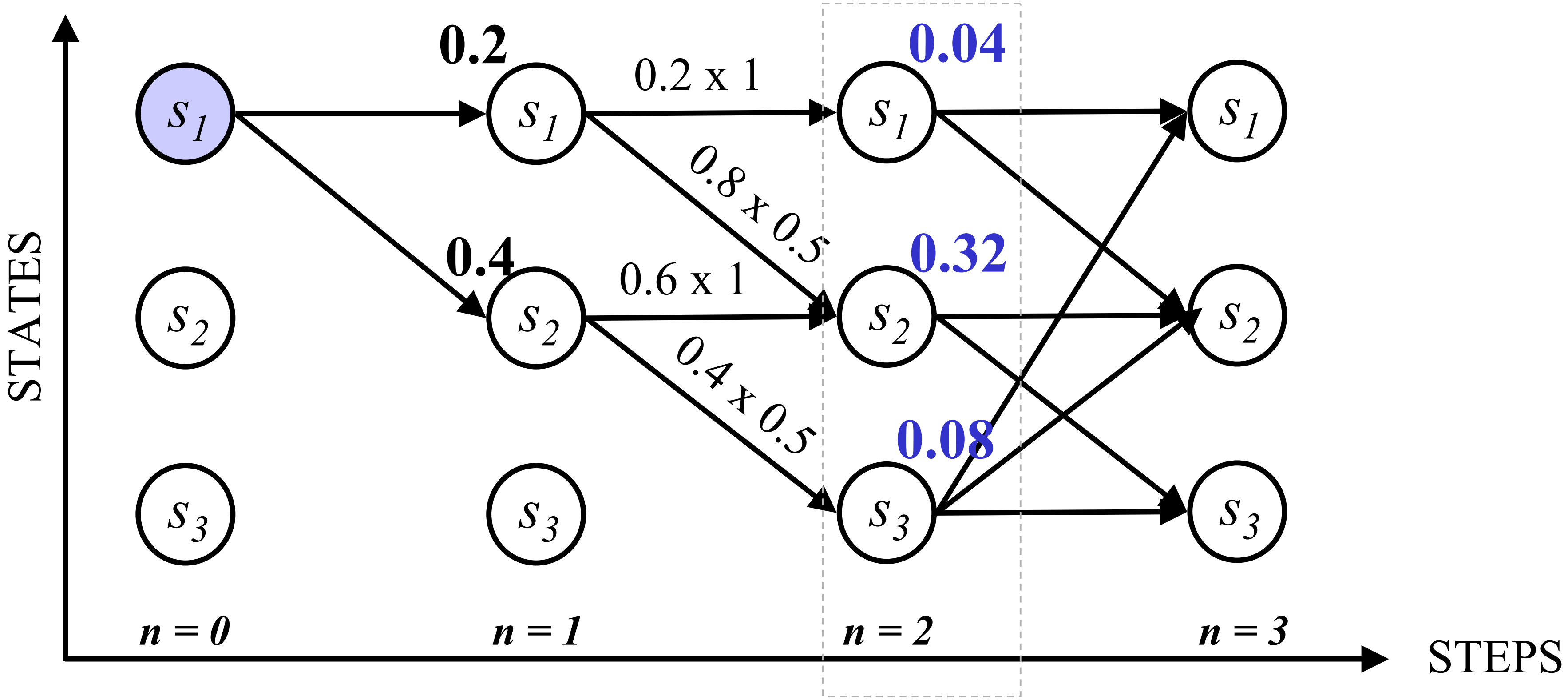
Consider exercise 2 again:

First we compute partial probabilities at $n=1$



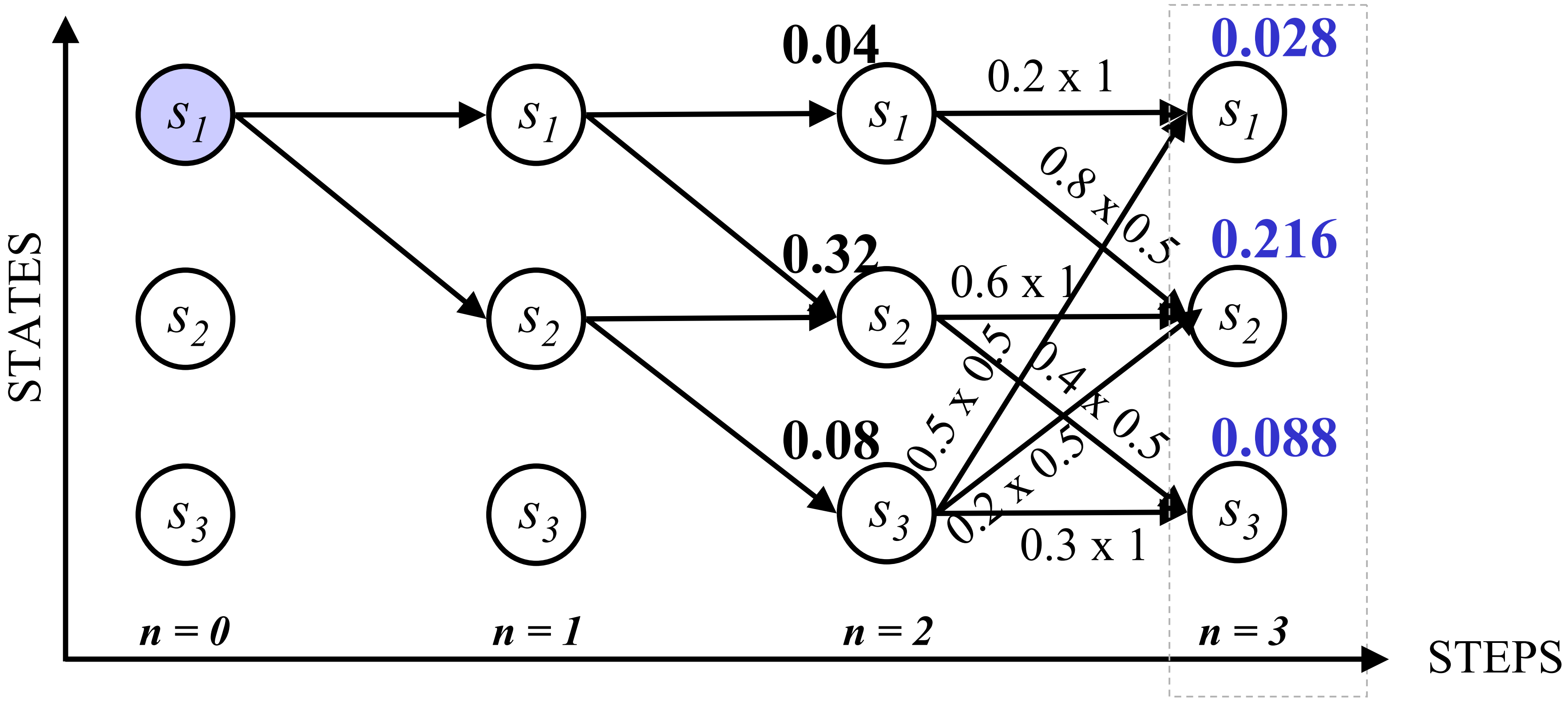
Dynamic Programming

Then, we proceed to compute partial probabilities at *n=2*



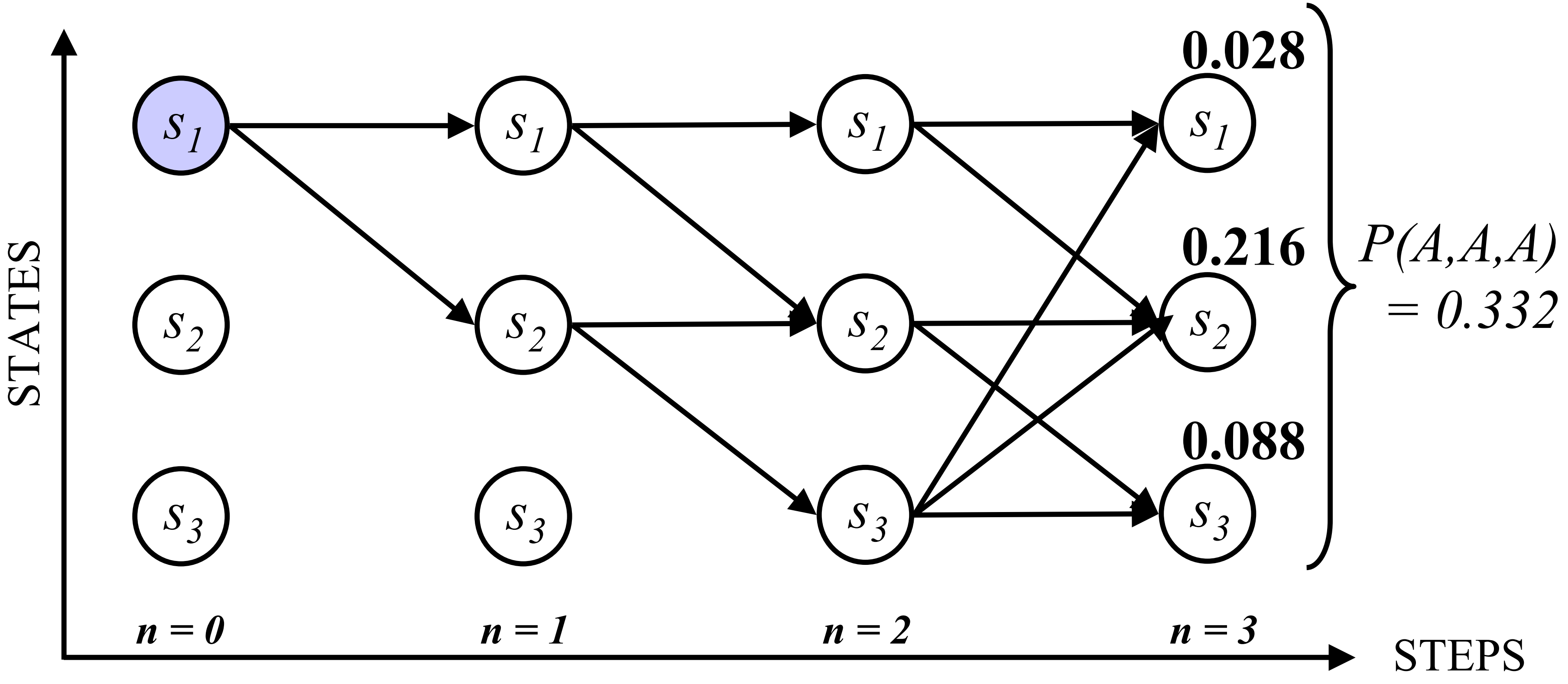
Dynamic Programming

And so on...



Dynamic Programming

Finally, the observation probability is computed by adding up probabilities at *n=3*



Dynamic Programming procedures

- The shown algorithm is referred to as the forward procedure for dynamic programming since it computes partial probabilities from the initial step at $n=0$ up to the final one at $n=N$

Dynamic Programming procedures

- The shown algorithm is referred to as the forward procedure for dynamic programming since it computes partial probabilities from the initial step at $n=0$ up to the final one at $n=N$

- It is also possible to implement a backward procedure, in which partial probabilities are computed from the final step $n=N$ down to the initial one $n=0$.

Dynamic Programming procedures

- The shown algorithm is referred to as the forward procedure for dynamic programming since it computes partial probabilities from the initial step at $n=0$ up to the final one at $n=N$
- It is also possible to implement a backward procedure, in which partial probabilities are computed from the final step $n=N$ down to the initial one $n=0$.
- Dynamic programming procedures reduce the total number of multiplications required from $(2N+1)\ M^{N+1}$ to $2N\ M^{2}$

The forward procedure

- Consider the vector f_n which contains the probability values of having observed the sequence $Y_1, Y_2 \ldots Y_n$ and reaching any of the M states at step n. (f_n is a row vector of M elements)

The forward procedure

- Consider the vector f_n which contains the probability values of having observed the sequence $Y_1, Y_2 \ldots Y_n$ and reaching any of the M states at step n. (f_n is a row vector of M elements)
- Initialization: $f_0 = \pi$
- Induction: $f_n = f_{n-1} * \; (\mathbf{P} \, .* \, \mathbf{Q}_{\boldsymbol{Y_n}})$ for $n = 1, 2, 3 \ldots. N$
- Total probability: $P(Y|\,\lambda) = f_N * \boldsymbol{1}$

The forward procedure

- Consider the vector f_n which contains the probability values of having observed the sequence $Y_1, Y_2 \ldots Y_n$ and reaching any of the M states at step n. (f_n is a row vector of M elements)

- Initialization: $f_0 = \pi$

matrix product

MATLAB's dot product (element by element)

- Induction: $f_n = f_{n-1} * \ (\mathbf{P} \,.\!* \mathbf{Q}_{Y_n})$ for $n = 1, 2, 3 \ldots. N$

Symbol emission probability matrix for given observation at step *n*

- Total probability: $P(Y|\ \lambda) = f_N * \mathbf{1}$

column vector of ones

The backward procedure

- Consider the vector b_n which contains the probability values of observing the future sequence $Y_{n+1}, Y_{n+2} \ldots Y_N$ when being at any of the M states at step n. (b_n is a column vector of M elements)

The backward procedure

- Consider the vector b_n which contains the probability values of observing the future sequence Y_{n+1}, $Y_{n+2} \dots Y_N$ when being at any of the M states at step n. (b_n is a column vector of M elements)

- Initialization: $b_N = 1$

- Induction: $b_{n-1} = (\mathbf{P} \,.\!* \mathbf{Q}_{\mathbf{Y}_n}) * b_n$ for $n = N, N\text{-}1, N\text{-}2 \dots 1$

- Total probability: $P(Y|\,\lambda) = \pi * b_0$

The backward procedure

- Consider the vector b_n which contains the probability values of observing the future sequence $Y_{n+1}, Y_{n+2} \dots Y_N$ when being at any of the M states at step n. (b_n is a column vector of M elements)

column vector of ones

- Initialization: $b_N = 1$

Symbol emission probability matrix for given observation at step *n*

- Induction: $b_{n-1} = (\mathbf{P} \; .* \; \mathbf{Q}_{\mathbf{Y}_n}) * b_n$ for $n = N, N\text{-}1, N\text{-}2 \dots. 1$

- Total probability: $P(Y|\,\lambda) = \pi * b_0$

MATLAB's dot product

Important property

Consider the row vectors f_n and the column vectors b_n of the forward and backward procedures of dynamic programming:

- it can be proved that $P(Y|\lambda) = f_n * b_n$
- and it holds for all values of $n = 0,1,2 \dots N$

Exercise 3

Consider again the HMM of exercise 2...

1.- Use MATLAB or any other language to implement the forward and backward procedures for computing the probability of the following observation sequences:

a.- $\boldsymbol{Y} = A, A, A$ b.- $\boldsymbol{Y} = B, B, B$ c.- $\boldsymbol{Y} = B, A, B, A$

2.- Verify that $P(Y|\, \lambda) = f_n * b_n$ holds for any value of $n=0,1,2,\ldots N$

3.- Repeat steps 1.- and 2.- by taking $\pi = [0.2, 0.5, 0.3]$ this time

Hidden Markov Models

Rafael E. Banchs (rbanchs@gps.tsc.upc.edu)

Session 2: Dynamic Programming & HMMs

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla

UNIVERSITAT POLITÈCNICA DE CATALUNYA