Arab Republic of Egypt
Ministry of Communications
and Information Technology

Information
Technology
Institute

EPITA
SCHOOL OF ENGINEERING & COMPUTER SCIENCE

A.I. IN AUDIO & SIGNAL PROCESSING

Session 3: HMM for speech processing

COURSE STRUCTURE

Quick Summary

Audio processing for AI

- Signal, audio, speech encoding (4h)
- Deep learning for audio processing (4h+4h)

Automata for language modelling

- HMM for speech processing (4h)
- Automata and transducer (4h)

Towards speaking with an AI-bot

- Speech synthesis (4h)
- Automatic speech recognition (4h)
- Speaker and emotion recognition (4h)

SESSION 3: HMM FOR SPEECH PROCESSING

Quick Summary

1. **Markov models & HMM**
2. **Scoring a sentence**
3. **Decoding a sequence of phonems**
4. **Training a language model**

HMM FOR SPEECH PROCESSING.

Markov models and HMM

MARKOV MODELS & HMM

Markov property defining a Markov Model

$\forall n \geq 0, (i_0, \dots, i_{n-1}, i, j) \in \boldsymbol{E}^{n+2},$

$$P(X_{n+1} = j |\, X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j |\, X_n = i)$$

We consider homogeneous models ($p_{i,j}$ is constant over time).

E	states space
$X_0, X_1, \dots, X_{n-1}, X_n$	random variable sequence of successive states
$p_{i,j}$	transition probability from state i to state j
n	time index (noted t further)

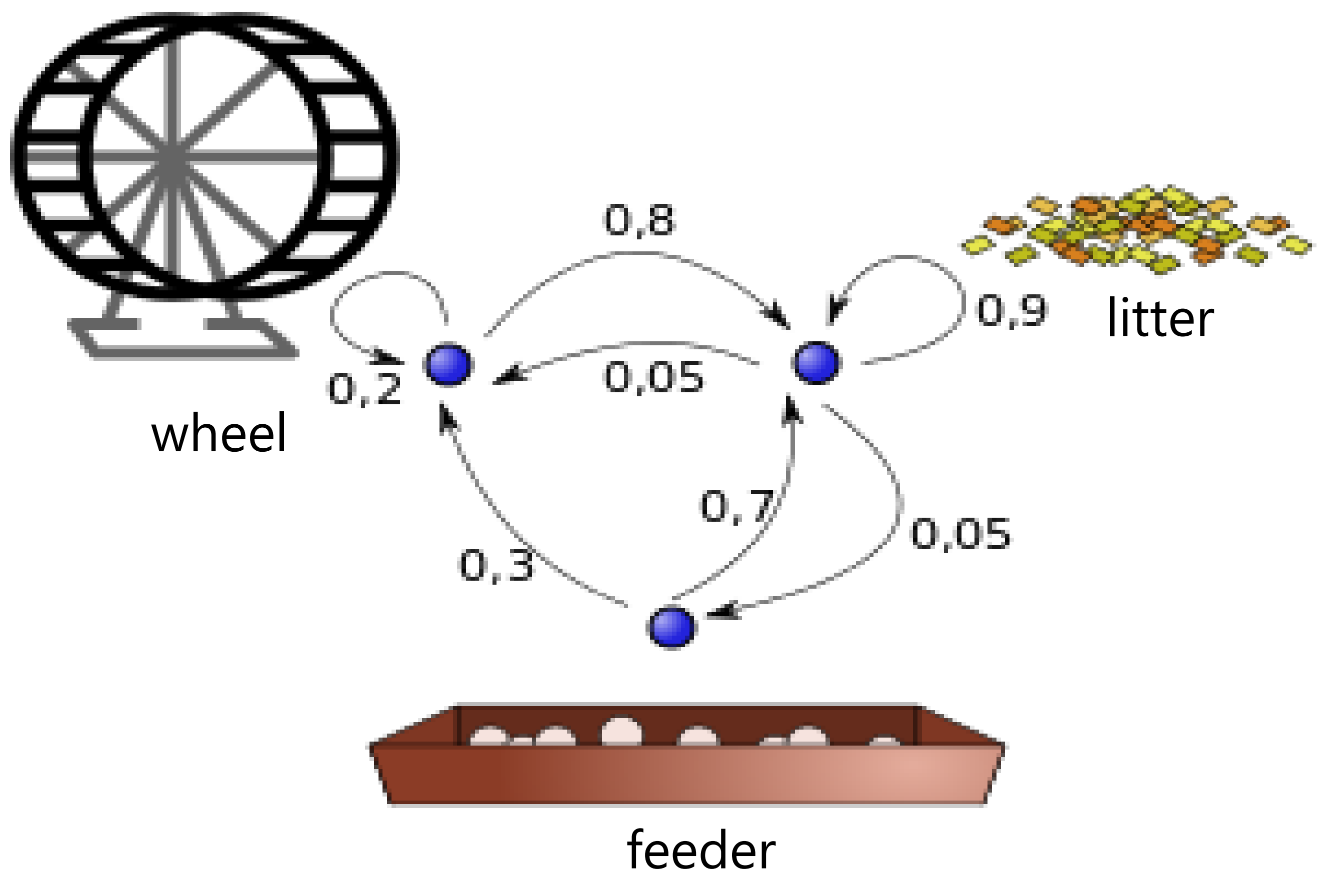
Transition probability

$$p_{i,j} = P(X_1 = j | X_0 = i) \quad \text{with} \quad \forall i \in \boldsymbol{E},\ \sum_{j \in E} p_{i,j} = 1$$

Example of Markov process

- Hamster pet

  → hamster activity at t_n is predictable,
  knowing its activity at t_0



MARKOV MODELS & HMM

Hidden Markov Model

Markov model with “partially observable” states
Usually, part only of the model is known:

→ either the sequence of observations O is unknown
→ either the sequence of states Q is unknown
→ either the transition probabilities are unknown

Elements of a discrete HMM

$S = \{S_0, S_1, \dots, S_N\}$ set of possible states
$V = \{V_0, V_1, \dots, V_M\}$ set of possible observations

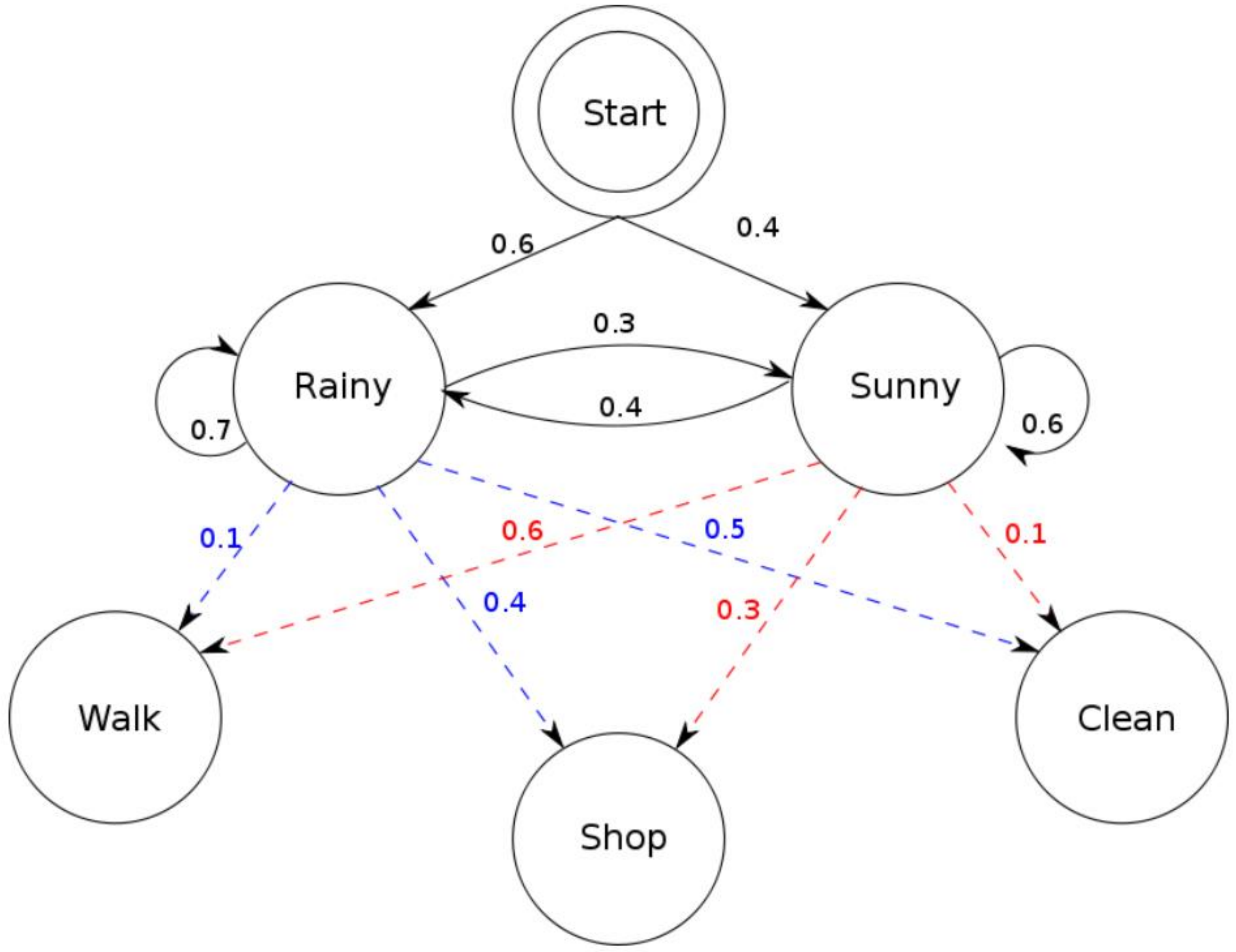
$Q = (q_0, q_1, \dots, q_T)$ sequence of states with t from 0 to T
$\mathcal{O} = (\mathcal{o}_0, \mathcal{o}_1, \dots, \mathcal{o}_T)$ sequence of observations with t from 0 to T

$a_{i,j} = P(q_{t+1} = S_j | q_t = S_i)$ state transition probability (matrix A)
$b_j(k) = P(\mathcal{o}_t = V_k | q_t = S_j)$ observation probability (matrix B)

$\pi = \{\pi_0, \pi_2, \dots, \pi_N\}$ initial state distribution, with $\pi_i = P(q_0 = S_i)$

Example of Hidden Markov Model

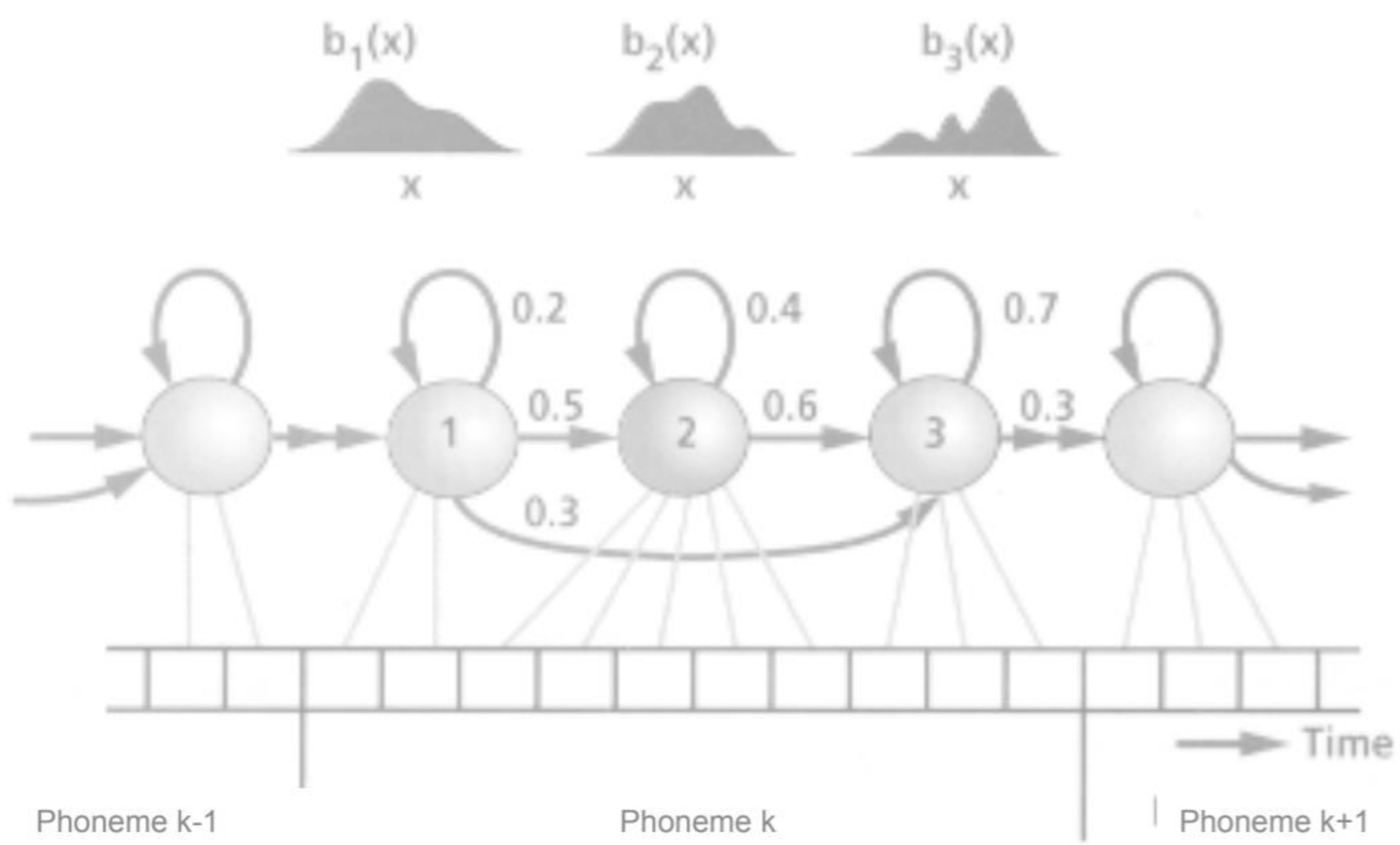
- activity depending on the weather

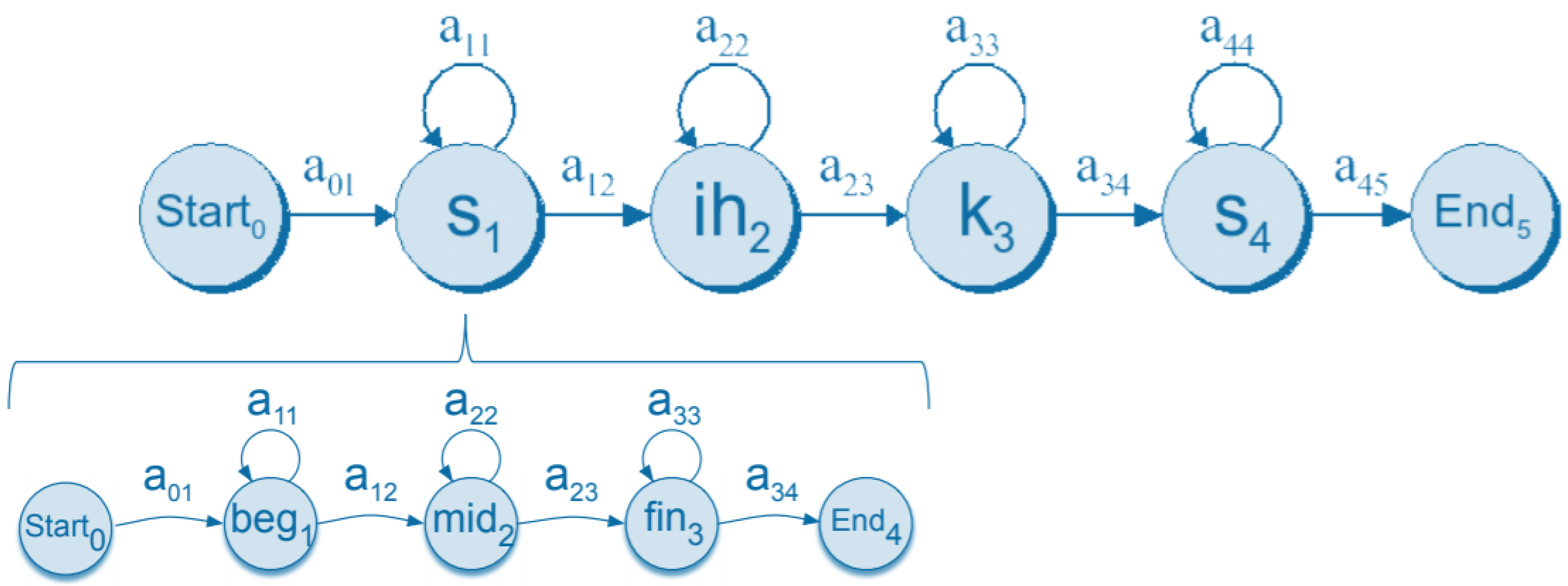


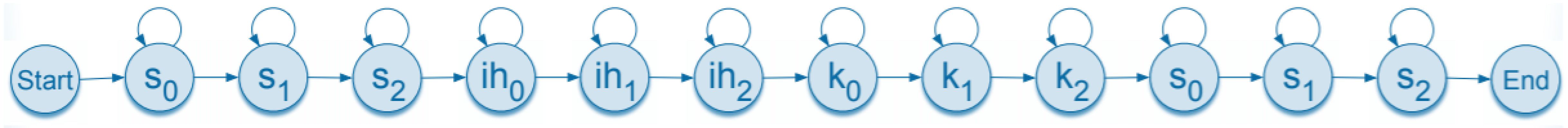
L. Baum, J. A. Eagon, T. Petrie. 1966-1970.

MARKOV MODELS & HMM

HMM application to speech







MARKOV MODELS & HMM

Basic problems for HMMs

- Scoring
  Given the state sequence $Q = (q_0, q_1, \ldots, q_T)$, and a model $\lambda = (A, B, \pi)$, how do we efficiently compute $\mathrm{P}(\mathcal{O}|\lambda, Q)$, the probability of the observation sequence, given the model?

  → Forward algorithm

- Matching/Decoding
  Given the observation sequence $\mathcal{O} = (\sigma_0, \sigma_1, \ldots, \sigma_T)$, and a model λ, how do we choose a corresponding state sequence $Q = (q_0, q_1, \ldots, q_T)$ which is optimal in some meaningful sense (i.e., best "explains" the observations). $\mathrm{P}(Q|\lambda, \mathcal{O})$?

  → Viterbi algorithm

- Training
  How do we adjust the model parameters model $\lambda = (A, B, \pi)$ to maximize $\mathrm{P}(\lambda|\mathrm{Q}, \mathcal{O})$?

  → Baum-Welch re-estimation procedures
  (known as forward-backward algorithm)

HMM FOR SPEECH PROCESSING.

Scoring a sentence

SCORING A SENTENCE

Goal

Find $\mathrm{P}(\mathcal{O}|\lambda)$,

$\mathrm{P}(\mathcal{O}|\lambda)$: probability to observe $\mathcal{O} = (\sigma_0, \sigma_1, \dots, \sigma_n)$, knowing the model $\lambda = (A, B, \pi)$

Analytical solving

law of total probability

(1) $$\mathrm{P}(\mathcal{O}|\lambda) = \sum_{all\ Q} \mathrm{P}(\mathcal{O}|Q,\lambda)\ \mathrm{P}(Q|\lambda)$$

Indépendance of observations knowing Q

(2) $$\mathrm{P}(\mathcal{O}|Q,\lambda) = \prod_{t=0}^{T} \mathrm{P}(\sigma_t|Q,\lambda)$$

σ_t depends on q_t and $q_0, q_1, \dots, q_{t-1}$

besides, as Q follow Markov property

$$\mathrm{P}(\sigma_t|Q,\lambda) = \mathrm{P}(\sigma_t|q_t,\lambda)$$
$$= b_{q_t}(\sigma_t)$$ by definition

initial state and transition probabilities

(3) $$\mathrm{P}(Q|\lambda) = \pi_{q_0} \prod_{t=1}^{T} \mathrm{a}_{\mathrm{q}_{t-1},\mathrm{q}_t}$$

(1), (2) and (3) give the result

(4) $$\mathrm{P}(\mathcal{O}|\lambda) = \sum_{all\ Q}\left[\pi_{q_0}.\, b_{q_0}(\sigma_0).\prod_{t=1}^{T} \mathrm{a}_{\mathrm{q}_{t-1},\mathrm{q}_t}.\, b_{q_t}(\sigma_t)\right]$$

MARKOV MODELS & HMM

<u>Computational solving: Forward algorithm</u>

- Initialization

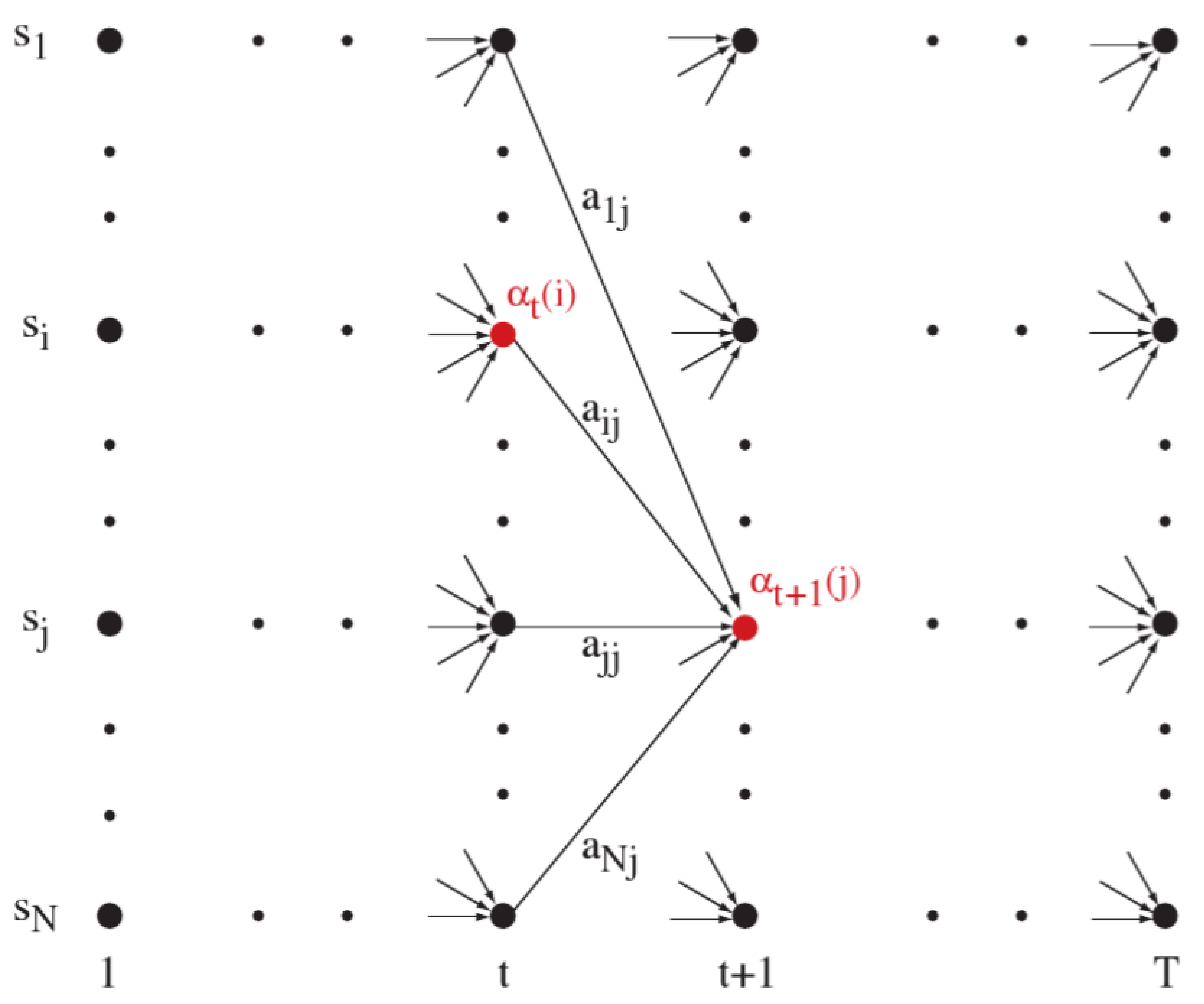
$$\alpha_0(i) = \pi_i \,.\, b_i(\mathcal{O}_0) \quad \text{for } i \in [\![0, N]\!]$$

- Induction

$$\alpha_t(j) = \left[\sum_{i=0}^{N} \alpha_{t-1}(i) \;.\, a_{i,j}\right] . \, b_j(\mathcal{O}_t) \quad \text{for } t \in [\![1, T]\!],\ j \in [\![0, N]\!]$$

- Termination

$$\mathrm{P}(\mathcal{O}|\lambda) = \sum_{i=0}^{N} \alpha_T(i)$$



$a_{i,j} = P(q_{t+1} = S_j | q_t = S_i)$ state transition probability (matrix A)
$b_j(k) = P(\mathcal{O}_t = V_k | q_t = S_j)$ observation probability (matrix B)

$\pi = \{\pi_0, \pi_2, \dots, \pi_N\}$ initial state distribution, with $\pi_i = P(q_0 = S_i)$

HMM FOR SPEECH PROCESSING.

Decoding a sequence of phonems

DECODING A SEQUENCE OF PHONEMS

Goal

Find most probable sequence of state $Q = (q_0, q_1, \ldots, q_T)$, given observations $\mathcal{O}$ and model λ.
→ find Q maximizing $\mathrm{P}(\mathrm{Q}|\mathcal{O}, \lambda)$

Forward algorithm provides a probability through all path sequence Q
→ find the optimum path sequence

Solving approaches

- Consider the path sequence maximizing successively each $a_{i,j}$
  → possibly not optimal

- Consider the path sequence maximizing $\mathrm{P}(\mathrm{Q}|\mathcal{O}, \lambda)$ with respect to the whole sequence
  → Viterbi algorithm

DECODING A SEQUENCE OF PHONEMS

Analytical solving

From equations (1) and (4), (see scoring previous chapter)

$$\mathrm{P}(\mathrm{Q}|\mathcal{O},\lambda) = \pi_{q_0}.\underbrace{b_{q_0}(\sigma_0).\prod_{t=1}^{T} \mathrm{a}_{\mathrm{q_{t-1}},\mathrm{q_t}} . b_{q_t}(\sigma_t)}_{\delta_T}$$

$$\mathrm{P}(\mathrm{Q}|\mathcal{O},\lambda) = \pi_{q_0}.b_{q_0}(\sigma_0).\prod_{t=1}^{T} \mathrm{a}_{\mathrm{q_{t-1}},\mathrm{q_t}} . b_{q_t}(\sigma_t)$$

Idea to compute iteratively overtime the probability δ_t for t$\in [\![1, T]\!]$

$$\delta_t(j) = \max_{0\leq i\leq N}\left(\delta_{t-1}\, a_{i,j}\right) . b_j(\sigma_t)$$

And thus, compute at each time step t, the most likely state transition

Viterbi algorithm assumptions

- $\mathcal{O}$ and Q are both in sequences
- $\mathcal{O}$ and Q are isomorphic (one observed event per hidden event)
- Q verifies Markov property

DECODING A SEQUENCE OF PHONEMS

Viterbi algorithm

1. Initialization:

$$\delta_1(i) = \pi_i b_i(o_1), \qquad 1 \le i \le N$$
$$\psi_1(i) = 0$$

2. Recursion:

$$\delta_t(j) = \max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(o_t), \qquad 2 \le t \le T \qquad 1 \le j \le N$$
$$\psi_t(j) = \operatorname*{argmax}_{1\le i\le N}[\delta_{t-1}(i)a_{ij}], \qquad 2 \le t \le T \qquad 1 \le j \le N$$

3. Termination:

$$P^* = \max_{1\le i\le N}[\delta_T(i)]$$
$$q_T^* = \operatorname*{argmax}_{1\le i\le N}[\delta_T(i)]$$

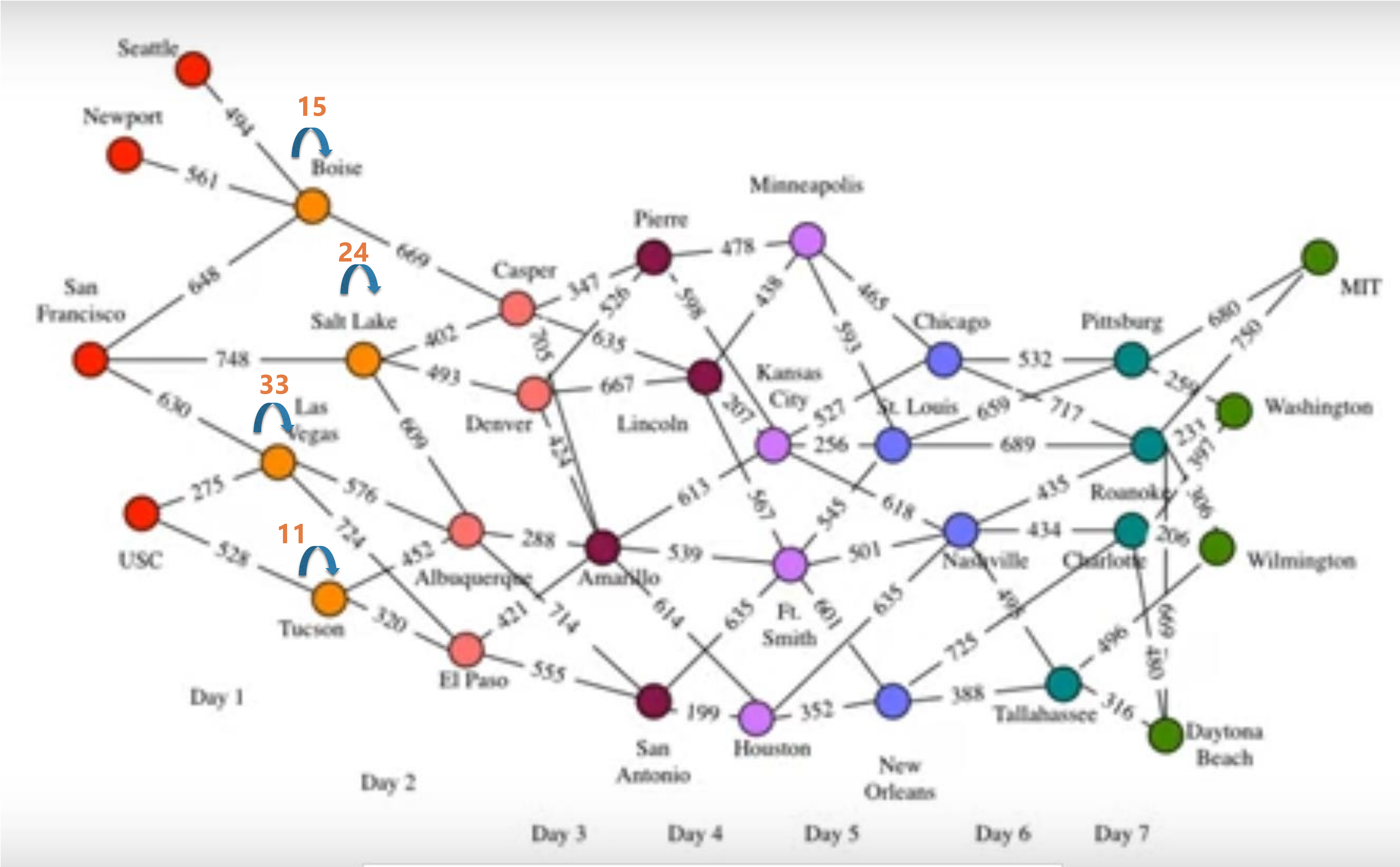
4. Path (state-sequence) backtracking:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \qquad t = T-1, T-2, \ldots, 1$$

Example

DECODING A SEQUENCE OF PHONEMS

Viterbi algorithm



nple

HMM FOR SPEECH PROCESSING.

Training a language model

TRAINING A LANGUAGE MODEL

Goal

Adjusting model parameters to maximize $\mathrm{P}(\mathrm{Q}, \mathcal{O}|\lambda)$.

$\mathcal{O} = (\mathscr{o}_0, \mathscr{o}_1, \dots, \mathscr{o}_T)$ is one of the training sequence

Analytical solving

→ none

Baum-Welch re-estimation procedures

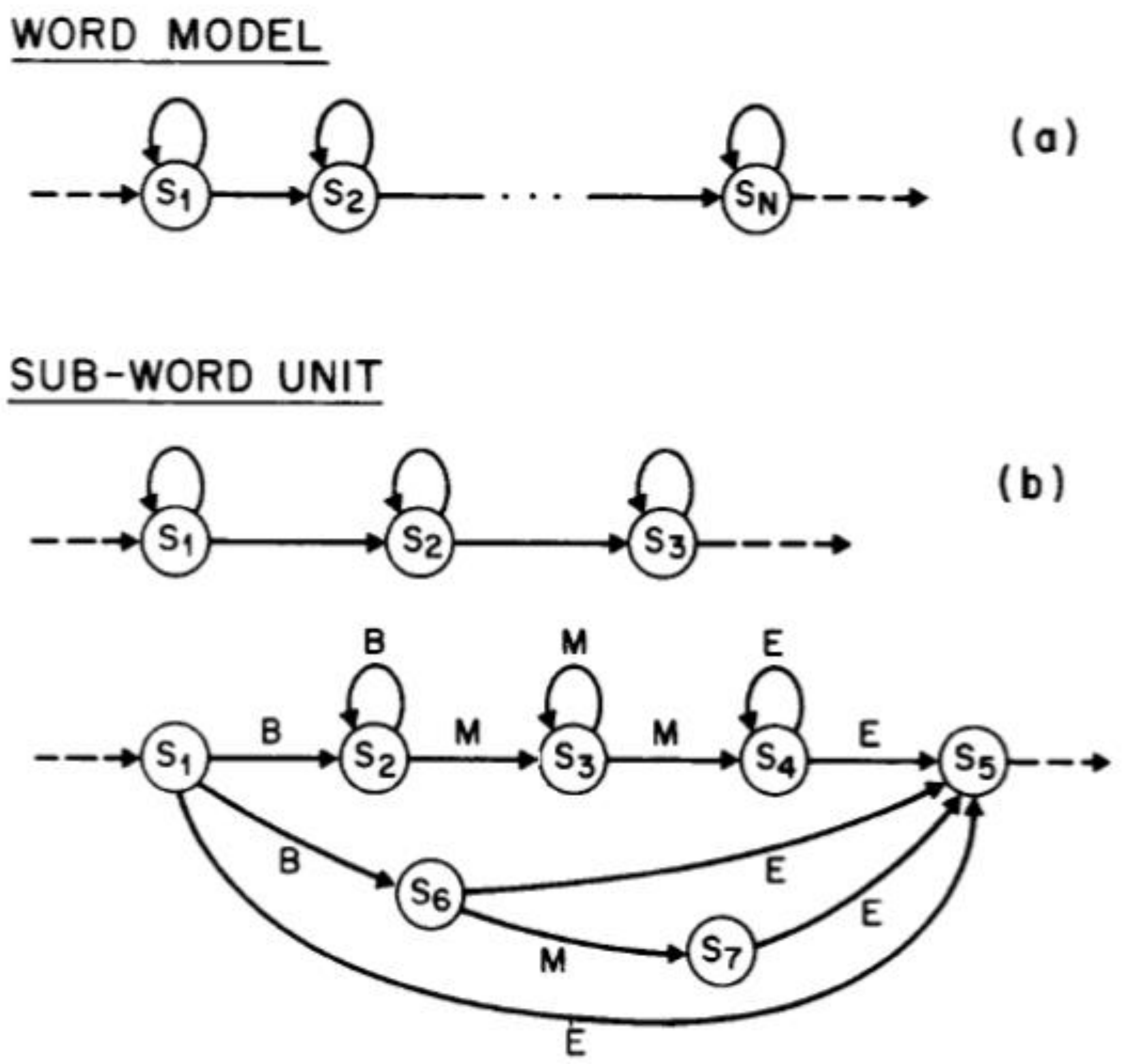
Iterative algorithm that:

- Compute statistics on the current model given the training data
- Adapt the model given the previous statistics
- Return to 1st step until convergence

Also known as forward-backward algorithm

TRAINING A LANGUAGE MODEL

Language model using HMM



HMM FOR SPEECH PROCESSING.

Thank you for your attention.

References:

- Xavier Anguera

PRACTICAL EXERCISE

1. Modelize Rainy-sunny model with hmmlearn

Use the following items:

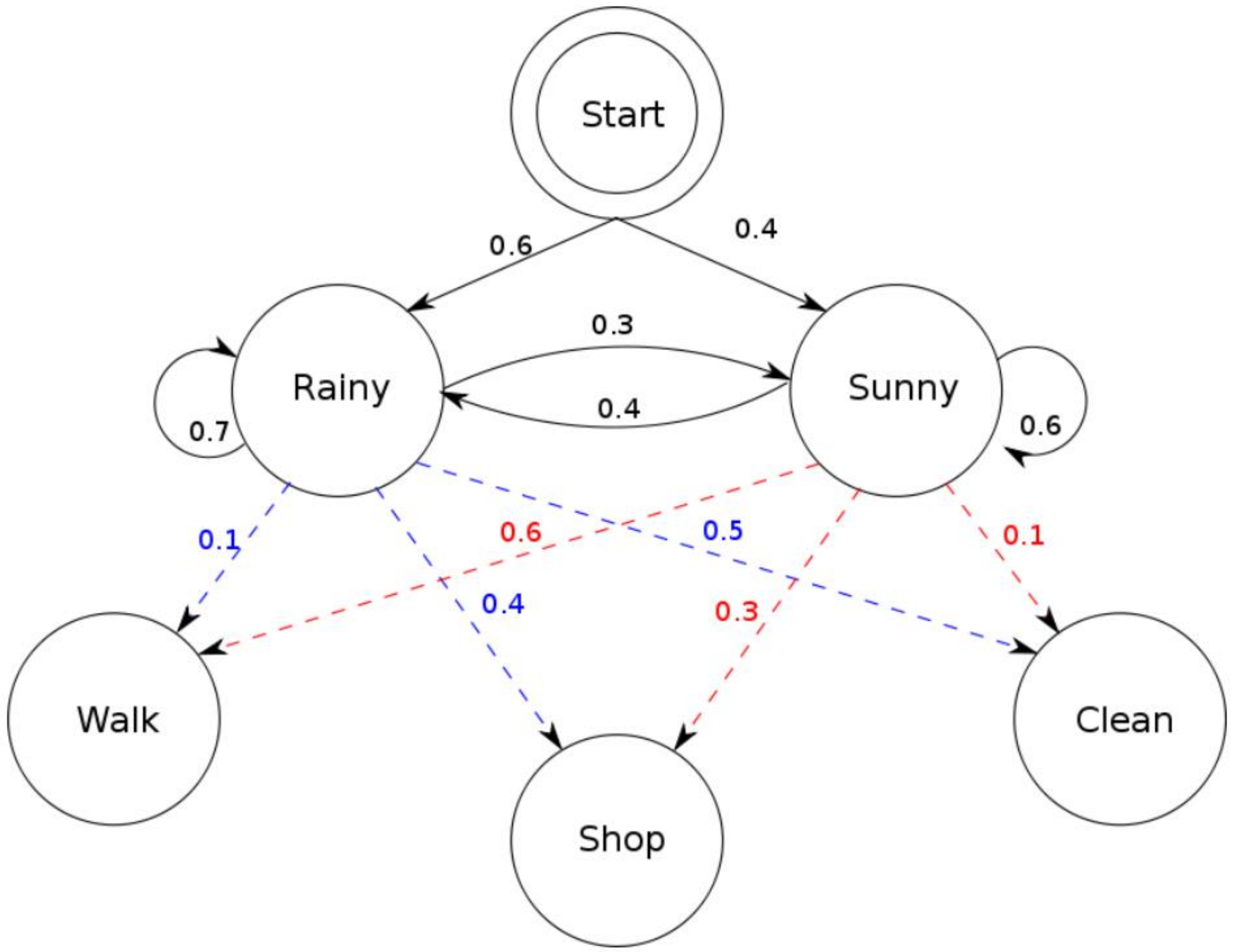
- from hmmlearn import hmm
- MultinomialHMM
- startprob_
- transmat_
- emissionprob_

TO DO:

- write starting probability
- transition matrix
- emission probability

Example of Hidden Markov Model

- activity depending on the weather



PRACTICAL EXERCISE

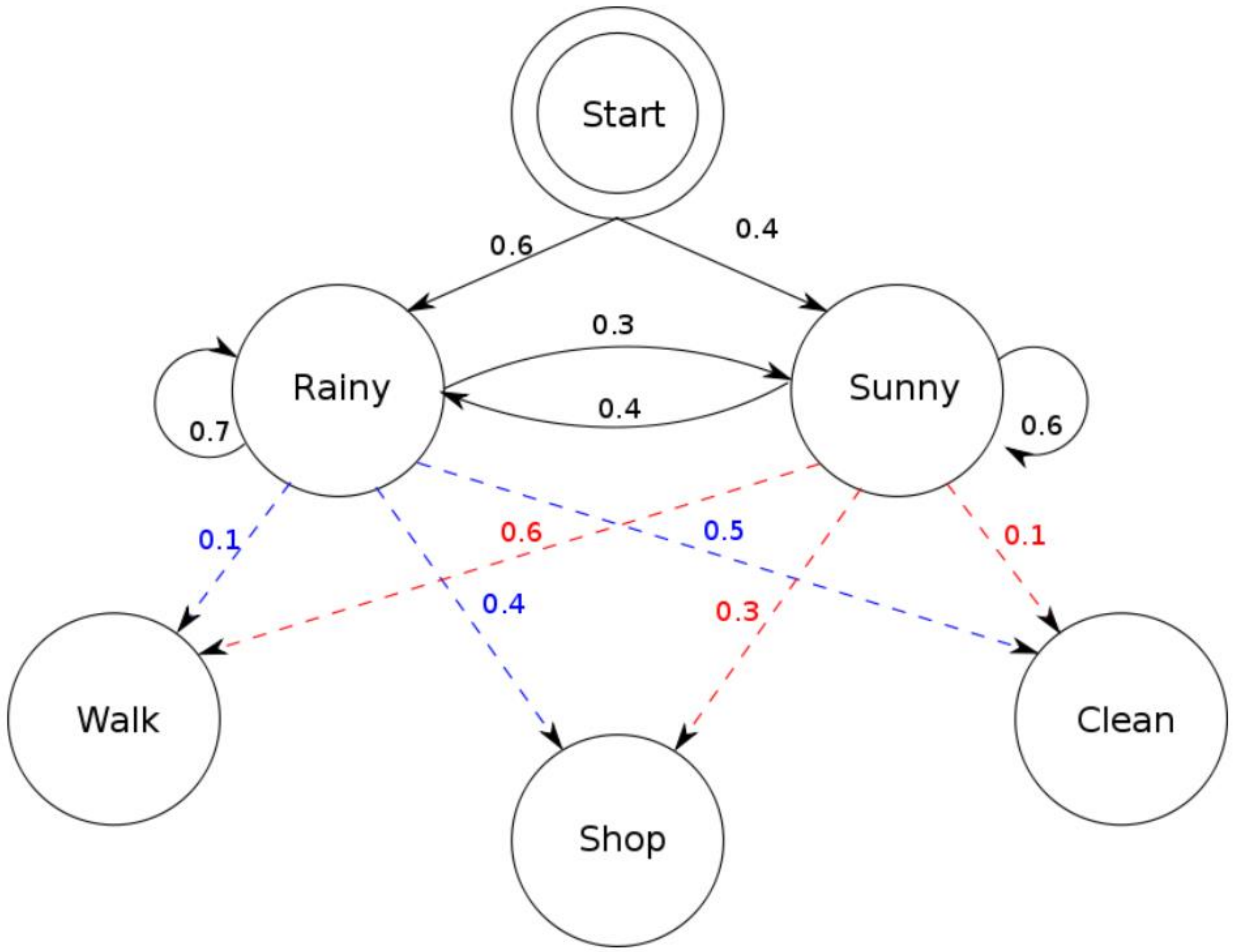
2. Solve scoring problem

Find probability of observations for the following sequences of states:
- (Start)
- (Rainy)
- (Sunny)
-(Sunny, Sunny, Sunny)

Use the following items:
- model.score

Example of Hidden Markov Model

- activity depending on the weather



PRACTICAL EXERCISE

2. Solve scoring problem

Find the seqence of states for the following observations:
- (Walk)
- (Shop)
- (Clean)
-(Clean, Clean, Clean)

Use the following items:
- model.decode

Example of Hidden Markov Model

- activity depending on the weather