Duration Modeling in HMMs

Lecture Notes

Speech Communication 2, SS 2004

Erhard Rank

Signal Processing and Speech Communication Laboratory

Graz University of Technology

Inffeldgasse 16c, A-8010 Graz, Austria

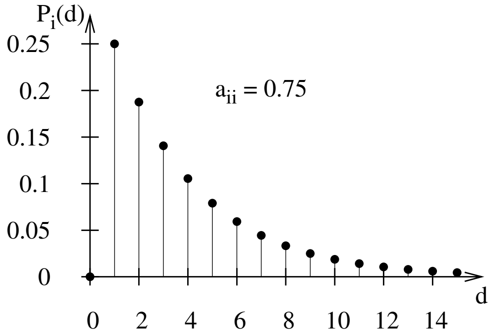
Tel.: +43 316 873 4436

E-Mail: rank@inw.tugraz.at

Mean State Duration

- Probability for duration d in state i (characteristic of Markov chain):

$$P_i(d) = P(Q|\boldsymbol{\Theta}, q_1 = s_i) = (a_{ii})^{d-1}(1 - a_{ii}), \quad d = 1, 2, \ldots$$



$$\text{with} \quad Q = \{\underbrace{s_i, s_i, \ldots, s_i}_{d\ \text{times}}, s_j\}, \quad j \neq i,$$

- Expected number of observations in a state:

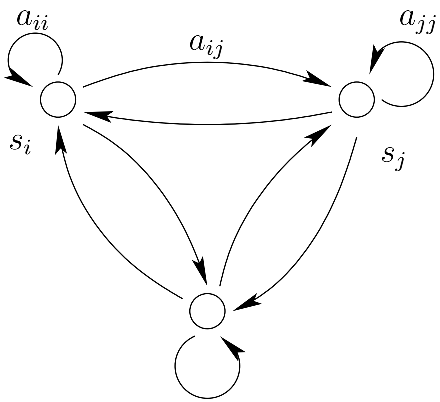
$$\bar{d}_i = \sum_{d=1}^{\infty} d\, P_i(d) = \sum_{d=1}^{\infty} d\, (a_{ii})^{d-1}(1 - a_{ii}) = \frac{1}{1 - a_{ii}}$$

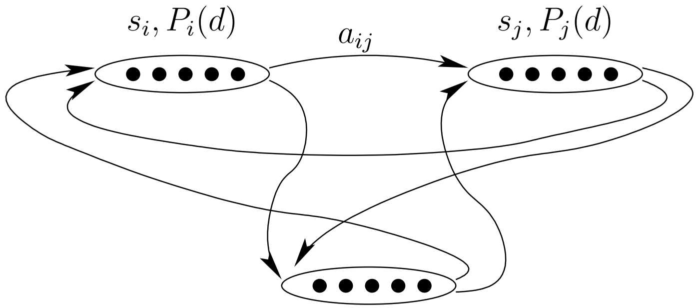
Weather example: $\bar{d}_{☀} = \frac{1}{1-0.8} = 5$ (days), $\bar{d}_{🌧} = 2.5$, $\bar{d}_{🌫} = 2$

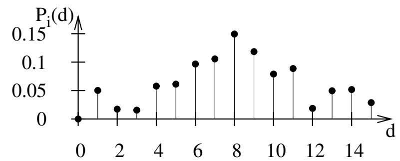
State duration probability = exponential function of d,
not appropriate for phoneme duration!

Semi-Markov Model

HMM with explicit duration modeling







Generation of an observation sequence:

- $q_1 = s_i$, acc. to π_i
- d_1 chosen acc. to $P_i(d)$
- Observations $x_1, x_2, \ldots, x_{d_1}$ generated acc. to joint emission probability $b_i(x_1, x_2, \ldots, x_{d_1})$, generally assumed independent $\Rightarrow = \prod_{n=1}^{d_1} b_i(x_n)$
- Transition to next state $q_2 = s_j, \quad j \neq i, \ (a_{ii} = 0)$

Equivalent to conventional HMM if $P_i(d) = (a_{ii})^{d-1}(1 - a_{ii})$

Problem 1 for Semi-Markov Model

HMM Problem 1: $P(X|\mathbf{\Theta}) = ?$

Assumption: entire state duration intervals are included in observation sequence

- First state begins at $n = 1$
- Last state ends at $n = N$

Definition: Forward variable $\alpha_n(j)$

- $\alpha_n(j) = P(x_1, x_2, \ldots, x_n, s_j \text{ ends at } n|\mathbf{\Theta})$

Assumption: r states are visited during n observations

- $q_1, q_2, \ldots, q_r$, with durations $d_1, d_2, \ldots, d_r$
  Constraints: $q_r = s_j, \; \sum_{l=1}^{r} d_l = n$

Problem: State durations d_l unknown

Problem 1 for Semi-Markov Model

Forward variable (sums over all states q, and all state durations d):

$$\alpha_n(j) = \sum_q \sum_d \pi_{q_1} P_{q_1}(d_1)\, P(x_1, x_2, \ldots, x_{d_1} | q_1)$$
$$\cdot\, a_{q_1 q_2}\, P_{q_2}(d_2)\, P(x_{d_1+1}, x_{d_1+2}, \ldots, x_{d_1+d_2} | q_2)$$
$$\ldots$$
$$\cdot\, a_{q_{r-1} q_r}\, P_{q_r}(d_r)\, P(x_{d_1+d_2+\ldots+d_{r-1}+1}, \ldots, x_n | q_r)$$

By induction:

$$\alpha_n(j) = \sum_{i=1}^{N_s} \sum_{d=1}^{D} \alpha_{n-d}(i)\, a_{ij}\, P_j(d) \prod_{s=n-d+1}^{n} b_j(x_s) \qquad (1)$$

$D \ldots$ maximum state duration ($P_j(d) = 0,\ d > D$)

Problem 1 for Semi-Markov Model

Initialization of forward variable:

$$\alpha_1(j) = \pi_j \, P_j(1) \, b_j(x_1)$$

$$\alpha_2(j) = \underbrace{\pi_j \, P_j(2) \prod_{s=1}^{2} b_j(x_s)}_{d_1=2,\ \text{in state } s_j} + \underbrace{\sum_{\substack{i=1 \\ i \neq j}}^{N_s} \alpha_1(i) \, a_{ij} \, P_j(1) \, b_j(x_2)}_{d_1=1,\ \text{in state } s_i \rightarrow \text{ state } s_j \text{ at } n=2}$$

$$\alpha_3(j) = \pi_j \, P_j(3) \prod_{s=1}^{3} b_j(x_s) + \sum_{d=1}^{2} \sum_{\substack{i=1 \\ i \neq j}}^{N_s} \alpha_{3-d}(i) \, a_{ij} \, P_j(d) \prod_{s=4-d}^{3} b_j(x_s)$$

$$\vdots$$

$$\alpha_D(j) = \ldots$$

For $n > D$: use eq. (1)

Parameter Update/Properties

HMM parameter update:

- Optimization of parameters (problem 3) possible, see Rabiner (1989)

Properties of Semi-Markov Models:

$+$ modeling quality significantly improves for some problems

$-$ greatly increased computational load (cf. eq. (1))

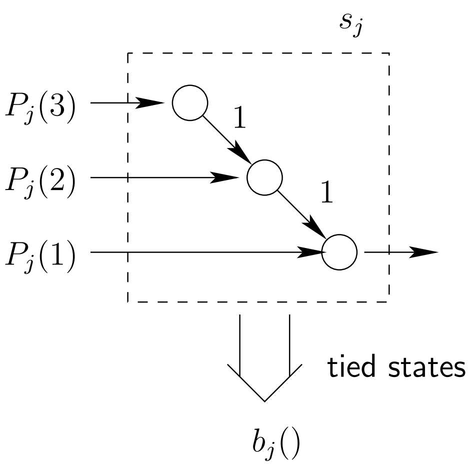
- $D\times$ storage
- $D^2/2\times$ computations (e. g., $D = 25 \rightarrow 300\times$)

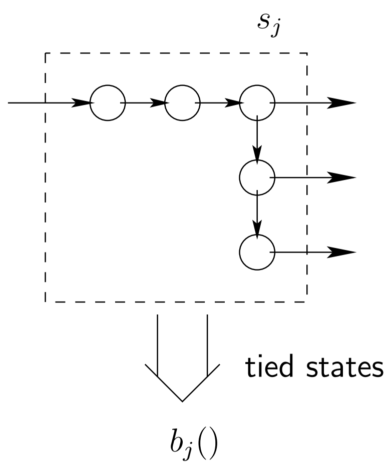
$-$ $P_j(d)$ must be estimated, D new parameters/state

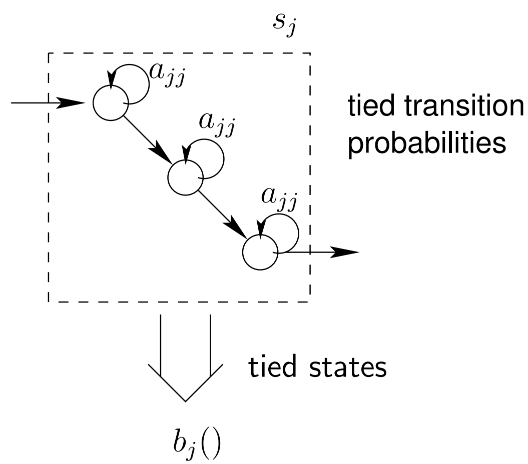
$\Rightarrow$ parametric $P_j(d)$ (Gaussian, Gamma-distr., Uniform in $[d_{\min}, d_{\max}]$)

Other duration models

Duration modeling by using tied internal states:







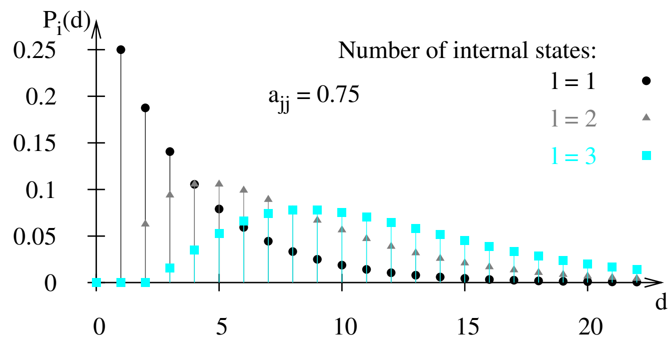
Simulation of the semi Markov model with tied emission probabilities

Duration limiting:

$$3 \leq d \leq 5$$

"Replikantenmodell", tied state transition probabilities

$$P_j(d) = a_{jj}^{d-l}\,(1-a_{jj})^l \binom{d-1}{l-1}$$



References

- L.R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proceedings of the IEEE, Vol. 77, No. 2, pp. 257-286, 1989. pdf
- E.G. Schukat-Talamazzini, *Automatische Spracherkennung*, Vieweg-Verlag, 1995.