Linguistics 696f
Hammond
Spring '03

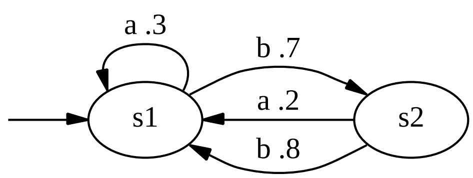
Hidden Markov Models

A. Overview

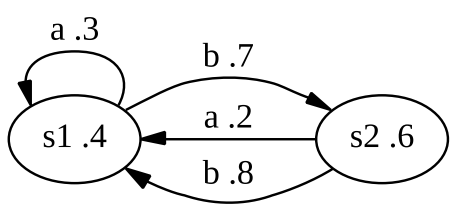
(1) a. Probabilistic FSMs
b. Formal properties of HMMs
c. N-grams and HMMs

B. Probabilistic FSMs

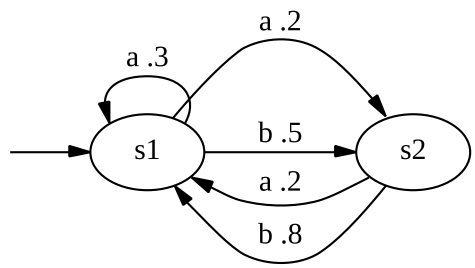
(2) A simple Markov chain:



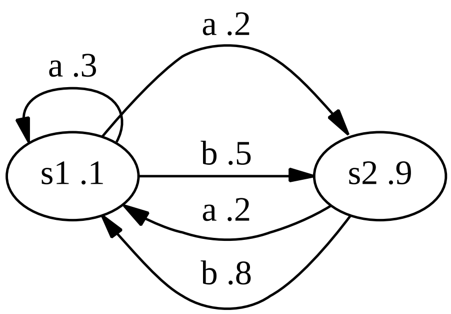
(3) Multiple start states:



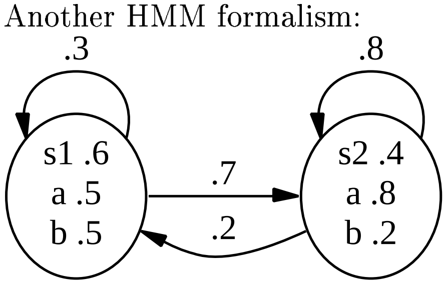
(4) Multiple arcs:



(5) Multiple arcs & start states:



(6) Another HMM formalism:



C. Formal properties of HMMs

(7) Limited Horizon:
$\mathrm{p}(X_{t+1} = s_k | X_1, \ldots, X_t) = \mathrm{p}(X_{t+1} = s_k | X_t)$

(8) Stochastic transition matrix A:
$a_{ij} = \mathrm{p}(X_{t+1} = s_j | X_t = s_i)$

(9) $a_{ij} \geq 0, \forall i, j$ and $\sum_{j=1}^{N} a_{ij} = 1, \forall i$

(10) $\pi_i = \mathrm{p}(X_1 = s_i)$

(11) $\sum_{i=1}^{N} \pi_i = 1$

(12)
$$\begin{aligned}
\mathrm{p}(X_i, \ldots, X_T) &= \mathrm{p}(X_1)\mathrm{p}(X_2|X_1)\mathrm{p}(X_3|X_1, X_2)\ldots \\
&\quad \mathrm{p}(X_T|X_1, \ldots, X_{T-1}) \\
&= \mathrm{p}(X_1)\mathrm{p}(X_2|X_1)\mathrm{p}(X_3|X_2)\ldots\mathrm{p}(X_T|X_{T-1}) \\
&= \pi_{X_1} \prod_{T=1}^{T-1} a_{X_t X_{t+1}}
\end{aligned}$$

(13) $\mathrm{p}(O_t = k | X_t = s_i, X_{t+1} = s_j) = b_{ijk}$

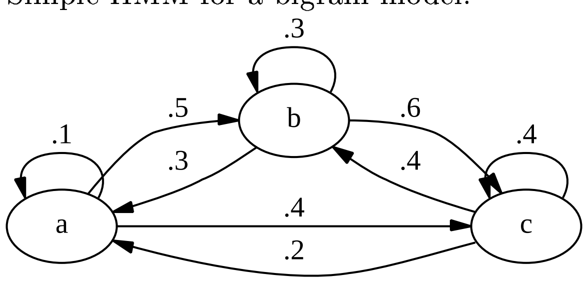
(14)

Set of states	$S = s_1, \ldots, s_N$
Output alphabet	$K = \{k_1, \ldots, k_M\} = \{1, \ldots, M\}$
Initial state probabilities	$\Pi = \{\pi_i\}, i \in S$
State transition probabilities	$A = \{a_{ij}\}, i, j \in S$
Symbol emission probabilities	$B = \{b_{ijk}\}, i, j \in S, k \in K$
State sequence	$X = (X_1, \ldots, X_{T+1})$
	$X_t : S \to \{1, \ldots, N\}$

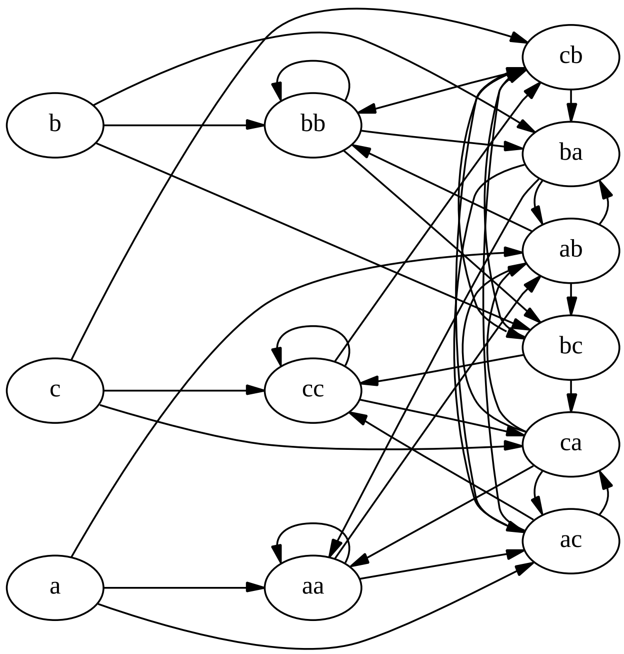
(15) $\forall n \sum_{w_{1n}} \mathrm{p}(w_{1n}) = 1$

D. N-grams and HMMs

(16) Simple HMM for a bigram model:



(17) HMM for a trigram model:



References

CHARNIAK, EUGENE. 1993. *Statistical Language Learning*. Cambridge: MIT Press.

HAMMOND, MICHAEL, 2003. Statistical natural language processing. U. of Arizona.

HOPCROFT, J.E., & J.D. ULLMAN. 1979. *Introduction to Automata Theory, Languages, and Computation*. Reading: Addison-Wesley.