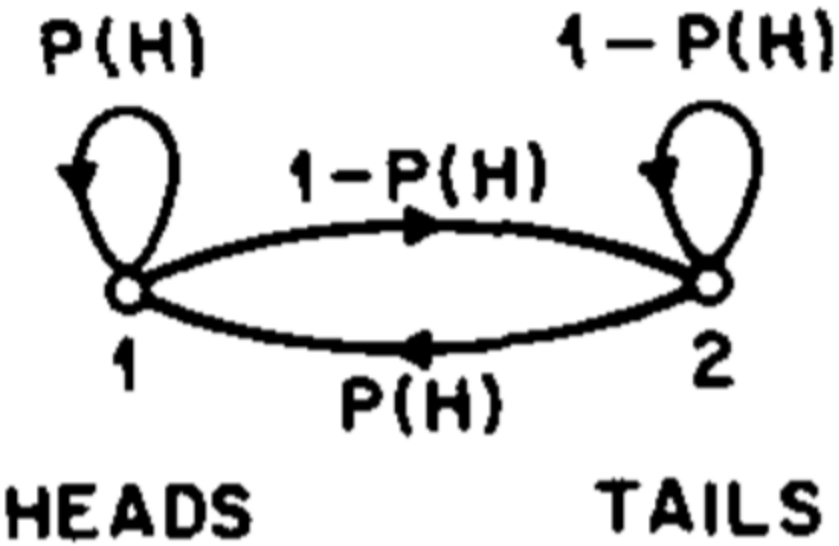
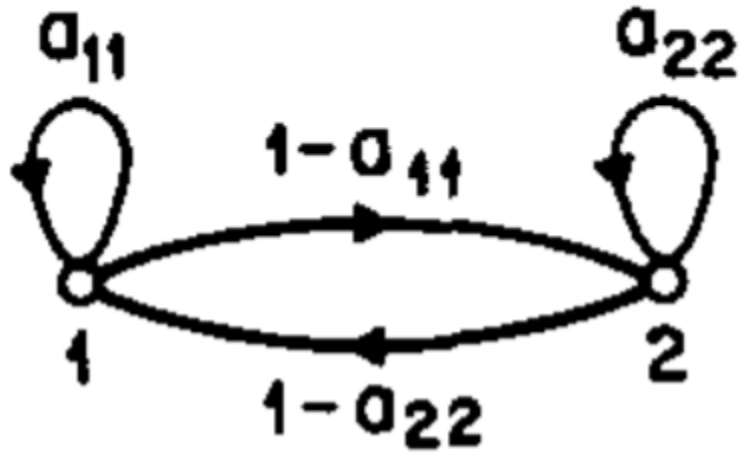


O = H H T T H T H H T T H . . .

S = 1 1 2 2 1 2 1 1 2 2 1 . . .

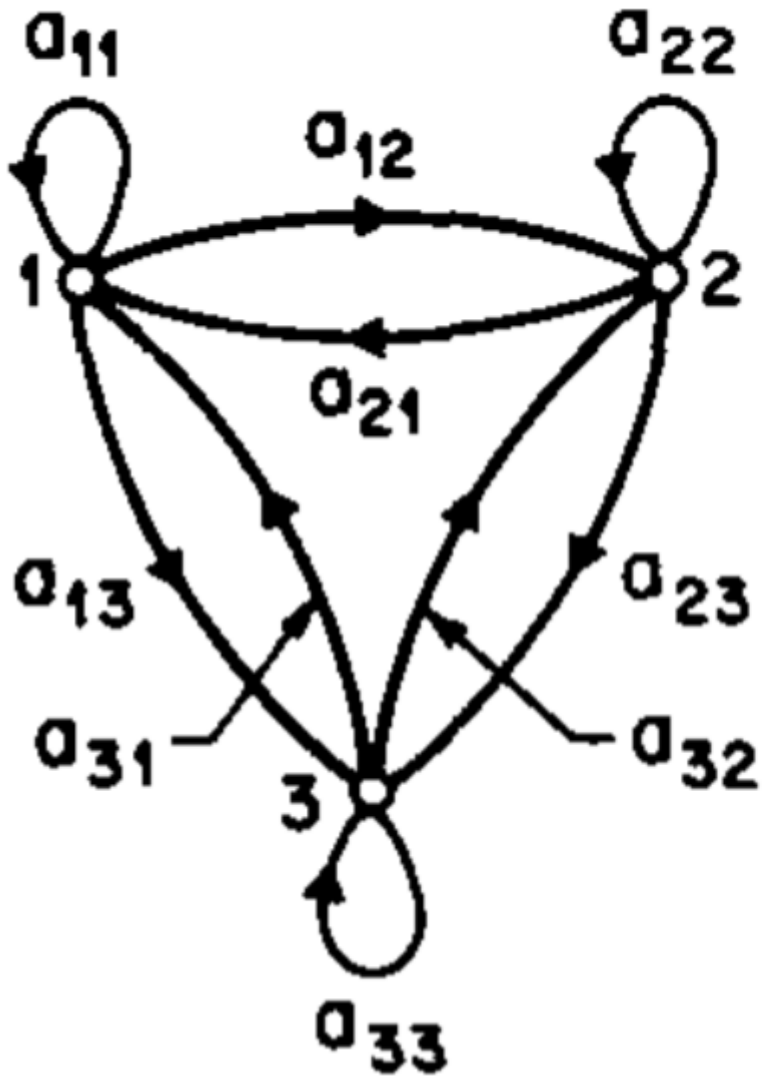


$P(H) = P_1$ $\quad$ $P(H) = P_2$

$P(T) = 1-P_1$ $\quad$ $P(T) = 1-P_2$

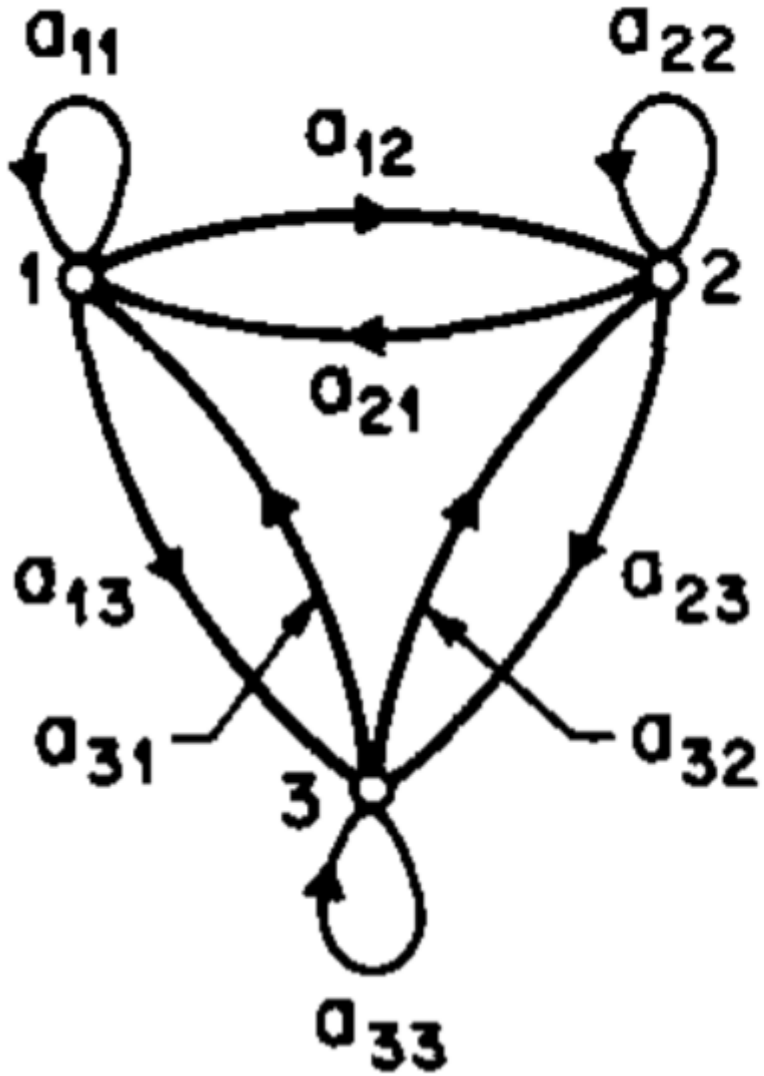
O = H H T T H T H H T T H ...

S = 2 1 1 2 2 2 1 2 2 1 2 ...



O = H H T T H T H H T T H ...

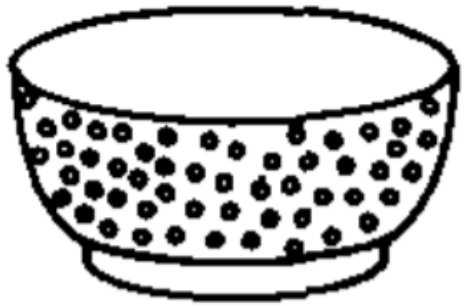
S = 3 1 2 3 3 1 1 2 3 1 3 ...



O = H H T T H T H H T T H ...

S = 3 1 2 3 3 1 1 2 3 1 3 ...

	STATE		
	1	2	3
P(H)	P_1	P_2	P_3
P(T)	$1-P_1$	$1-P_2$	$1-P_3$



. . .

URN 1

$P(RED) = b_1(1)$
$P(BLUE) = b_1(2)$
$P(GREEN) = b_1(3)$
$P(YELLOW) = b_1(4)$
$\vdots$
$P(ORANGE) = b_1(M)$

URN 2

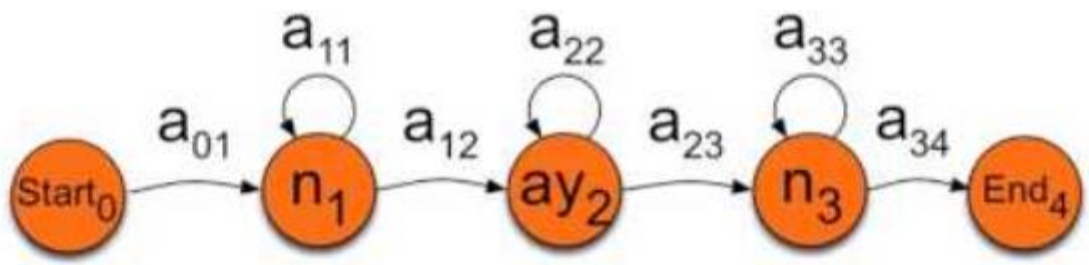
$P(RED) = b_2(1)$
$P(BLUE) = b_2(2)$
$P(GREEN) = b_2(3)$
$P(YELLOW) = b_2(4)$
$\vdots$
$P(ORANGE) = b_2(M)$

URN N

$P(RED) = b_N(1)$
$P(BLUE) = b_N(2)$
$P(GREEN) = b_N(3)$
$P(YELLOW) = b_N(4)$
$\vdots$
$P(ORANGE) = b_N(M)$

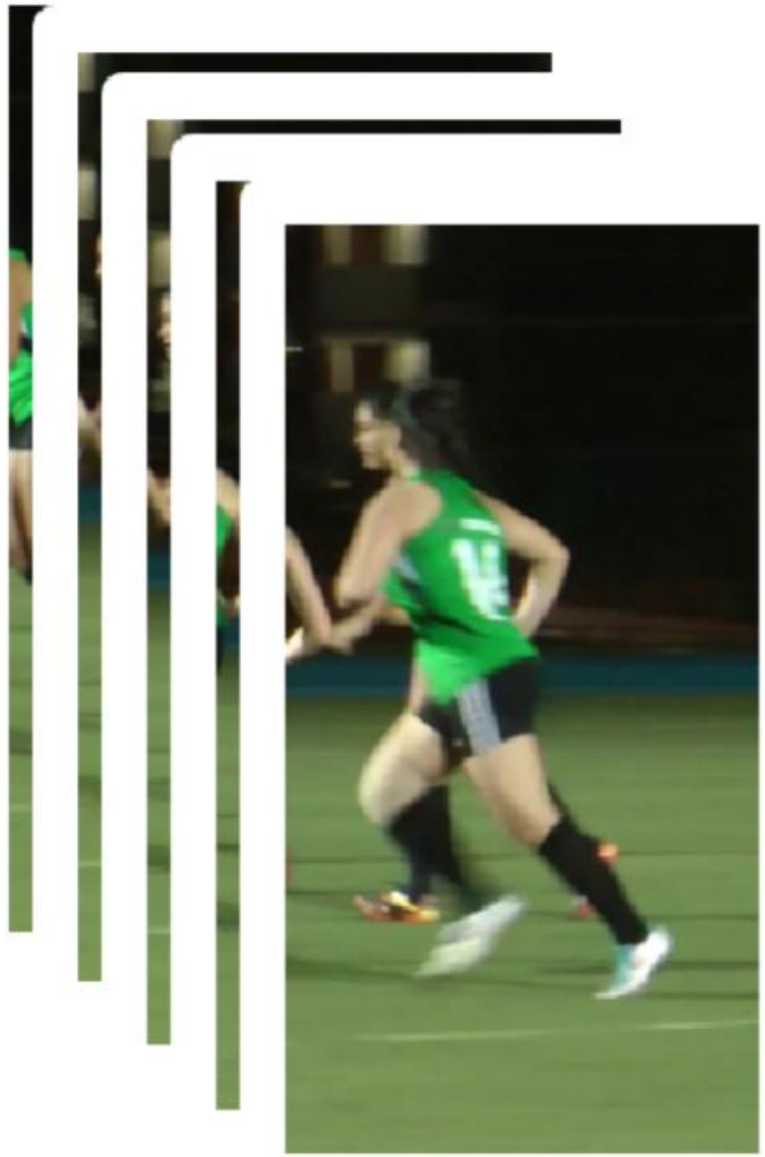
$O = \{GREEN, GREEN, BLUE, RED, YELLOW, RED, \ldots\ldots, BLUE\}$

Word HMM

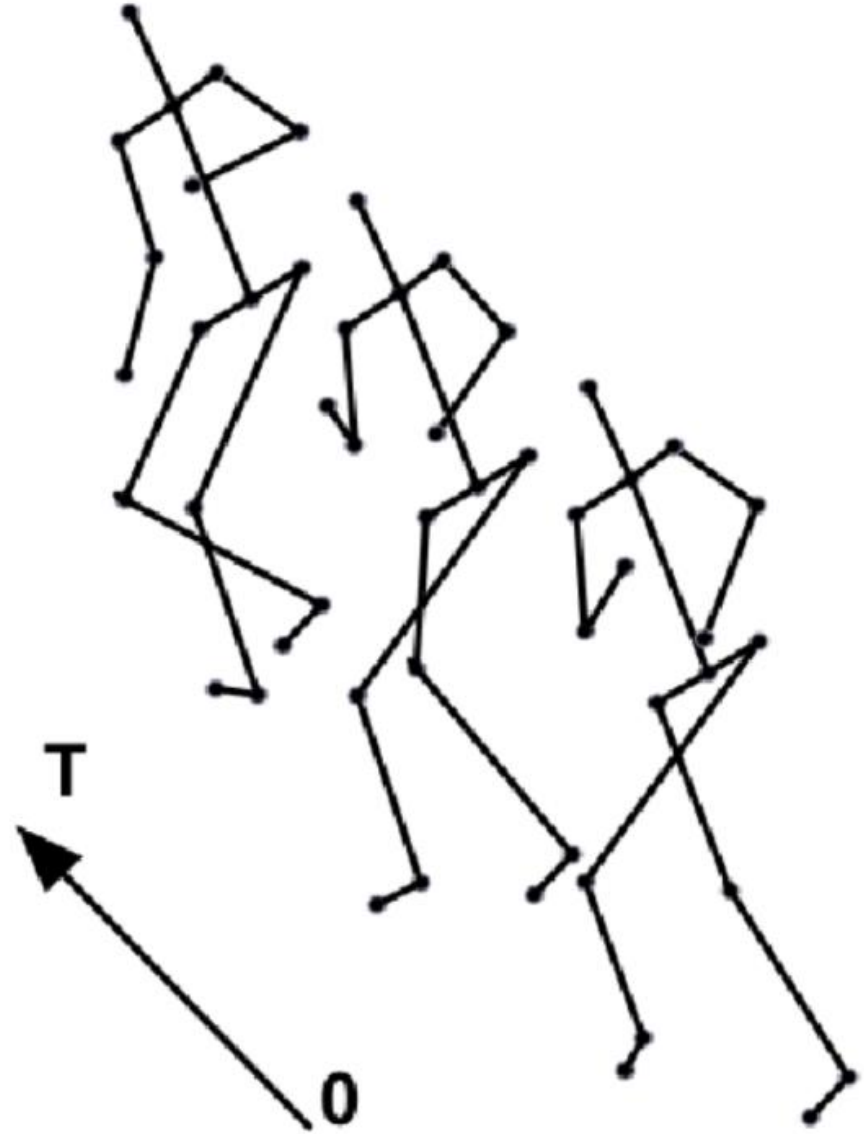


Dan Jurafsky, Stanford

Daphne Koller



A. Image sequence



B. Skeleton sequence

Human Body Pose Tracking Based on Spatio-Temporal Joints Dependency Learning

by

Rodrigo Barrita Zebadúa

Thesis submitted in partial fulfillment of the requirements for the degree of:

MSc. in Computer Science

Finding Genes in DNA with a Hidden Markov Model

John Henderson* Steven Salzberg† Kenneth H. Fasman‡

Abstract

This study describes a new Hidden Markov Model (HMM) system for segmenting uncharacterized genomic DNA sequences into exons, introns, and intergenic regions. Separate HMM modules were designed and trained for specific regions of DNA: exons, introns, intergenic regions, and splice sites. The models were then tied together to form a biologically feasible topology. The integrated HMM was trained further on a set of eukaryotic DNA sequences, and tested by using it to segment a separate set of sequences. The resulting HMM system, which is called VEIL (Viterbi Exon-Intron Locator), obtains an overall accuracy on test data of 92% of total bases correctly labelled, with a correlation coefficient of 0.73. Using the more stringent test of exact exon prediction, VEIL correctly located both ends of 53% of the coding exons, and 49% of the exons it predicts are exactly correct. These results compare favorably to the best previous results for gene structure prediction, and demonstrate the benefits of using HMMs for this problem.

International Journal of Bio- Science and Bio- Technology
Vol. 1, No. 1, December, 2009

HMM based approach for classifying protein structures

Georgina Mirceva[1] and Danco Davcev[2]

Faculty of Electrical Engineering and Information Technologies
University Ss. Cyril and Methodius, Skopje, Macedonia
[1]georgina@feit.ukim.edu.mk, [2]etfdav@feit.ukim.edu.mk

Abstract

To understand the structure-to-function relationship, life sciences researchers and biologists need to retrieve similar structures from protein databases and classify them into the same protein fold. With the technology innovation the number of protein structures increases every day, so, retrieving structurally similar proteins using current structural alignment algorithms may take hours or even days. Therefore, improving the efficiency of protein structure retrieval and classification becomes an important research issue. In this paper we propose novel approach which provides faster classification (minutes) of protein structures. We build separate Hidden Markov Model (HMM) for each class. In our approach we align tertiary structures of proteins. Viterbi algorithm is used to find the most probable path to the model. We have compared our approach against an existing approach named 3D HMM, which also performs alignment of tertiary structures of proteins by using HMM. The results show that our approach is more accurate than 3D HMM.

Keywords: *Protein Data Bank (PDB), protein classification, Structural Classification of Proteins (SCOP), Hidden Markov Model (HMM), 3D HMM.*

Consider the forward variable $\alpha_t(i)$ defined as

$$\alpha_t(i) = P(O_1\, O_2 \cdots O_t,\, q_t = S_i|\lambda)$$

i.e., the probability of the partial observation sequence, $O_1\, O_2 \cdots O_t$, (until time t) *and* state S_i at time t, given the model λ. We can solve for $\alpha_t(i)$ inductively, as follows:

1) Initialization:

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N.$$

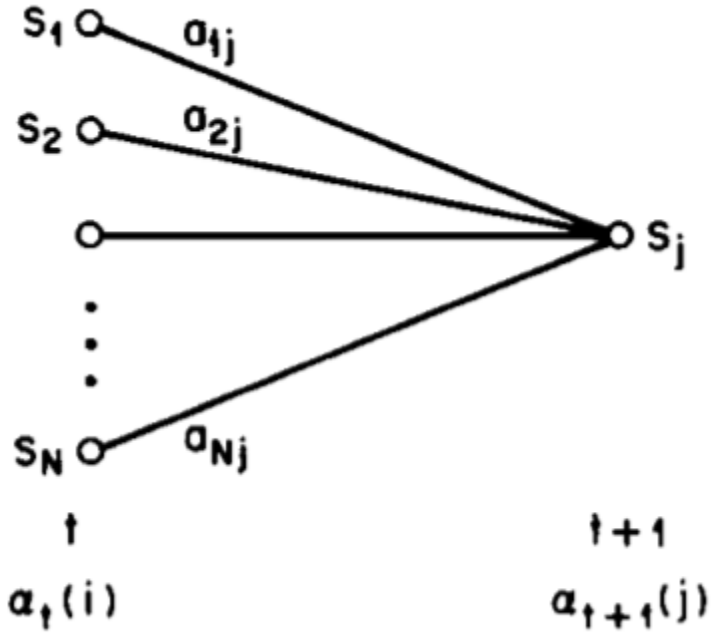
2) Induction:

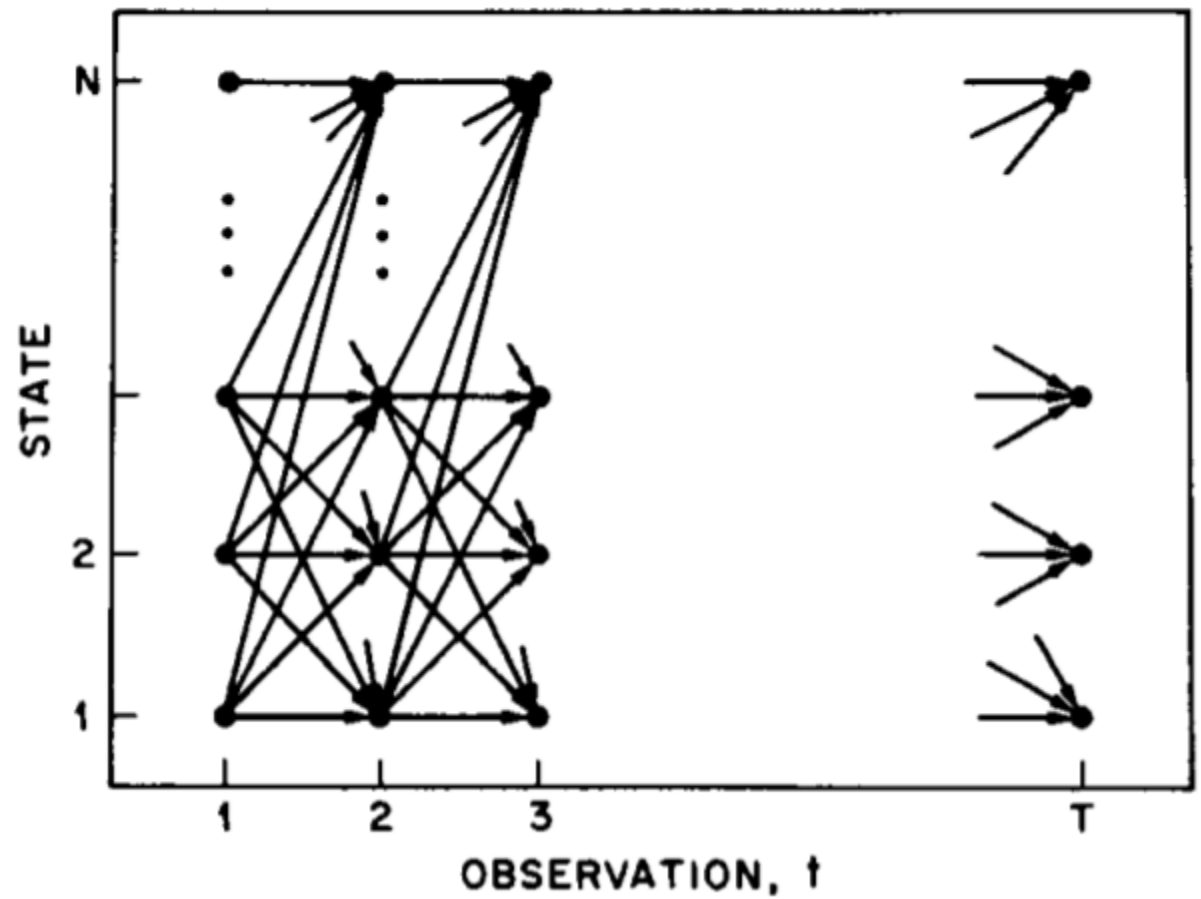
$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \qquad 1 \le t \le T-1$$

$$1 \le j \le N.$$

3) Termination:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i).$$


S_1
a_{1j}
S_2
a_{2j}
S_j
S_N
a_{Nj}
t
t+1
$\alpha_t(i)$
$\alpha_{t+1}(j)$


N
2
1
STATE
1
2
3
T
OBSERVATION, t