Modelos Ocultos de Markov

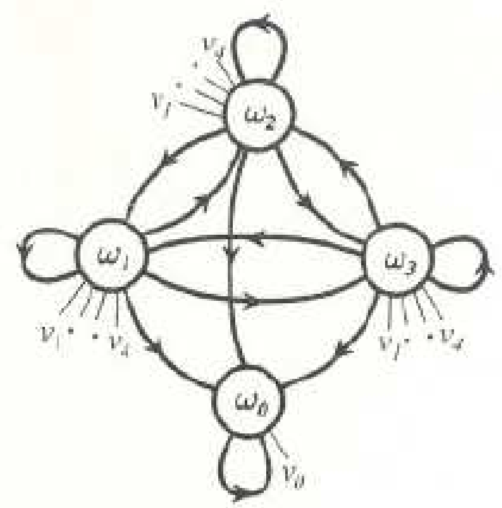
Jorge Luis Guevara Díaz

Universidad Nacional de Trujillo

Escuela de Informática

Introducción

- Son modelos muy utilizados en diversas áreas, la razón de esto es que poseen una rica estructura matemática y es esto lo que forma a base teórica para poder aplicarlos en muchas áreas, funcionando muy bien en la práctica



Introducción

- ¿Cómo caracterizar señales del mundo real en términos de modelos?
  - Por ejemplo:
    - Mejorar una señal ruidosa, para esto se debe establecer un modelo de que es una señal no ruidosa
    - Como encontrar un buen modelo?
    - Que tipo de modelo usar?

Introducción

- Modelos de señal determinísticos
  - Propiedades específicas de la señal. Ejemplo: la señal es una onda seno, amplitudes, frecuencias, etc

- Modelos de señal estocáticos
  - Modelos estadísticos de la señal, se busca solamente caracterizar las propiedades estadísticas de la señal. Ejemplo: Procesos Gaussianos, procesos Poison, Procesos de Markov, Procesos Ocultos de Markov, etc

Introducción

- Para RAH necesitaremos ambos tanto los modelos determinísticos como los modelos estocásticos

- A continuación estudiaremos un modelo estocástico los HMM (Hidden Markov Models) Modelos Ocultos de Markov

Introducción

- Estos modelos fueron introducidos en los años 1960 por Baum, y fue utilizado para aplicaciones de procesamiento de señales de habla por Baker y Jelinek a mediados de 1970

Introducción

- Sea $\mathbf{X} = X_1, X_2, \ldots X_n$ una secuencia de variables aleatorias de un albafeto discreto $O = \{o_1, o_2, \ldots, o_M\}$

  Según la regla de Bayes tenemos

  $$P(X_1, X_2, \ldots, X_n) = P(X_1)\prod_{i=2}^{n} P(X_i \mid X_1^{i-1})$$

  donde

  $$X_1^{i-1} = X_1, X_2, \ldots, X_{i-1}$$

  Se dice que las variables aleatorias ***X*** son cadenas de Markov de primer orden si

  $$P(X_i \mid X_1^{i-1}) = P(X_i \mid X_{i-1})$$

Introducción

- Entonces

$$P(X_1, X_2, \ldots, X_n) = P(X_1)\prod_{i=2}^{n} P(X_i \mid X_1^{i-1})$$

- Viene a ser

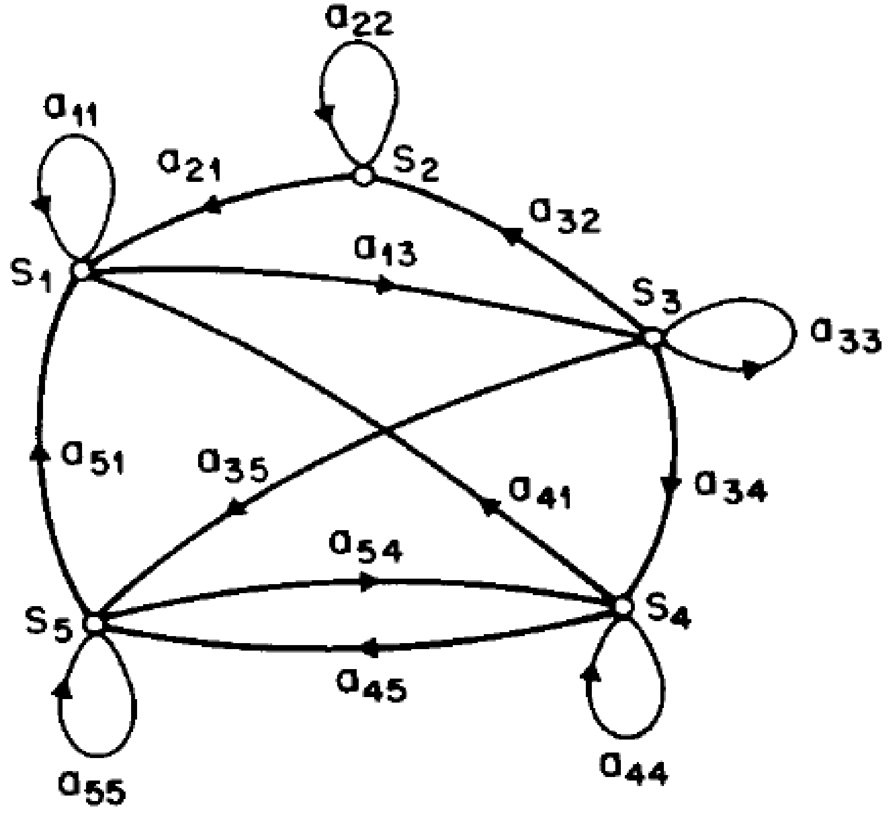
$$P(X_1, X_2, \ldots, X_n) = P(X_1)\prod_{i=2}^{n} P(X_i \mid X_{i-1})$$

Procesos Discretos de Markov

- Consideremos un sistema que puede ser descrito en cualquier instante de tiempo, empezando en cualquiera de sus N estados

Estados espaciados regularmente en el tiempo

$$S_1, S_2, \cdots, S_N$$



Cadena de Markov de N=5 estados

Procesos Discretos de Markov

- Cada estado tiene asociada una probabilidad
- Existen instantes de tiempo asociados a cada instante $\boldsymbol{t = 1,2,\ldots}$
- El estado actual en el tiempo $\boldsymbol{t}$ será denotada mediante $\boldsymbol{q_t}$
- Existen probabilidades de transición entre estados $\boldsymbol{a_{ij}}$

Procesos Discretos de Markov

- Una descripción probabilística completa del sistema requiere la especificación del actual estado en el tiempo ***t*** y todos los estados predecesores
- Las cadenas de Markov de primer orden truncan esta descripción probabilística, justo en el estado predecesor

$$
\begin{aligned}
P[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \cdots] \\
= P[q_t = S_j | q_{t-1} = S_i].
\end{aligned}
$$

Procesos Discretos de Markov

- Las probabilidades de trancición de estados $\boldsymbol{a_{ij}}$ tienen entonces la siguiente forma

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i], \qquad 1 \leq i, j \leq N$$

Con $$a_{ij} \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1$$

Modelo observable de Markov

Procesos Discretos de Markov

- En conclusión una cadena de Markov con ***N*** estados donde el estado en el tiempo ***t*** puede ser descrito como $\boldsymbol{q_t}$ puede ser descrita como

1. Un conjunto finito de estados $Q = \{q_1, q_2, \ldots q_n\}$
2. Una distribucion de probabilidad inicial en todos los estados

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N \qquad \sum_{j=1}^{N} \pi_j = 1$$

3. Una matriz de transicion de estados

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i], \quad 1 \leq i, j \leq N$$

con

$$\sum_{j=1}^{N} a_{ij} = 1 \quad 1 \leq i \leq N \qquad a_{ij} \geq 0$$

Procesos Discretos de Markov

- Teniendo esto en mente se puede decir que un modelo de cadena de Markov incorpora un monto mínimo de memoria

Ejemplo

- Modelo de 3 estados de Markov del Tiempo
  - Estado 1 lluvioso
  - Estado 2 nublado
  - Estado 3 soleado

Matriz de transición de probabilidades

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

Ejemplo

- Dado que el tiempo en el dia ***t = 1*** es soleado (estado 3), ¿Cuál es la probabilidad de acuerdo al modelo que el tiempo de los siguientes 7 días sea **soleado – soleado – lluvioso – lluvioso – soleado – nublado – soleado** ?
- Es decir, dada una secuencia de Observaciones

  $O = \{ S_3, S_3, S_1, S_1, S_3, S_2, S_3\}$ para

  $t = \{1 \ldots 8\}$

  Determinar la probabilidad de ***O*** dado el ***modelo***

Ejemplo

$$
\begin{aligned}
P(O|\text{Model}) &= P[S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3|\text{Model}] \\
&= P[S_3] \cdot P[S_3|S_3] \cdot P[S_3|S_3] \cdot P[S_1|S_3] \\
&\quad \cdot P[S_1|S_1] \cdot P[S_3|S_1] \cdot P[S_2|S_3] \cdot P[S_3|S_2] \\
&= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\
&= 1 \cdot (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\
&= 1.536 \times 10^{-4}
\end{aligned}
$$

Procesos Discretos de Markov

- Dado el modelo en un estado conocido, ¿Cuál es la probabilidad de que permanezca en el mismo estado exactamente d dias?

Procesos Discretos de Markov

- Probabilidad de la secuencia de Observación

$$O = \{S_{i_1}, S_{i_2}, S_{i_3}, \cdots, S_{i_d}, S_{j_{d+1}} \neq S_i\},$$

- Dado el modelo

$$P(O|\text{Model}, q_1 = S_i) = (a_{ii})^{d-1}(1 - a_{ii}) = p_i(d).$$

- Donde $\boldsymbol{p_i(d)}$ es la función de densidad de probabilidad discreta de duración $\boldsymbol{d}$ en el estado $\boldsymbol{i}$

Procesos Discretos de Markov

- El número esperado de observaciones en un estado basado en $\boldsymbol{p_i(d)}$ es el siguiente:

$$\bar{d}_i = \sum_{d=1}^{\infty} d p_i(d)$$

$$= \sum_{d=1}^{\infty} d(a_{ii})^{d-1}(1 - a_{ii}) = \frac{1}{1 - a_{ii}}.$$

¿Cuál es el número esperado de dias consecutivos soleados, nublados, lluviosos?

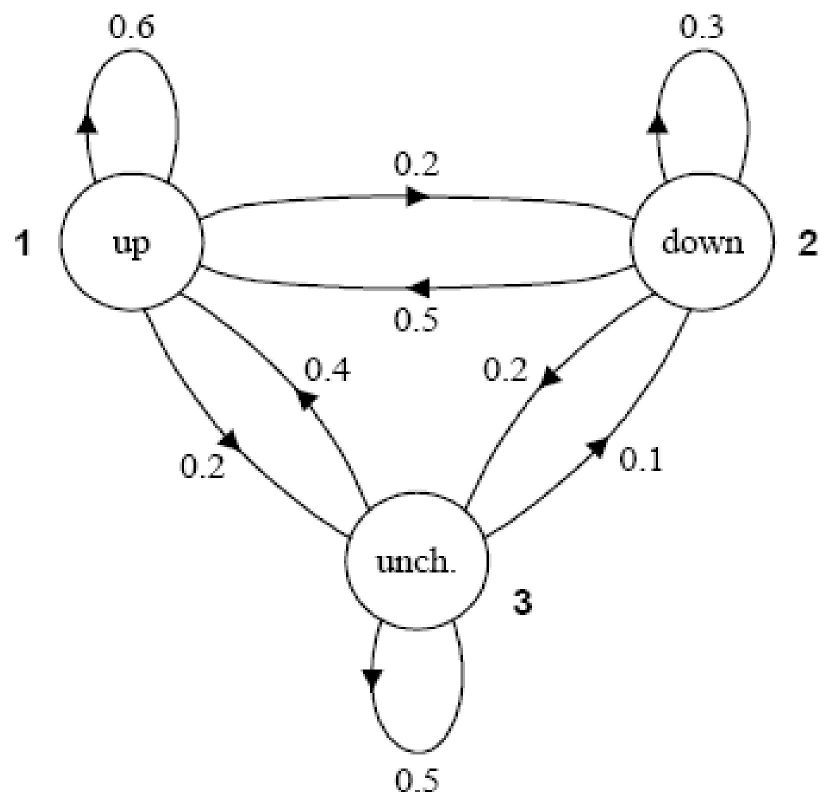
$1/(0.2) = 5;$ $\quad$ 2.5 $\quad$ 1.67

Otro Ejemplo

- ¿Cuál es la probabilidad de up-up-up-up-up?

$$A = \{a_{ij}\} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$$

$$\pi = (\pi_i)^t = \begin{pmatrix} 0.5 \\ 0.2 \\ 0.3 \end{pmatrix}$$



Modelos Ocultos de Markov

- Considere el siguiente escenario:

  Una persona hace experimentos con monedas, y solamente nos dice el resultado de sus experimentos. Es decir una secuencia ***oculta*** es realizada, por ejemplo

  Una secuencia oculta de caras y sellos toma lugar

$$\boldsymbol{O} = O_1\, O_2\, O_3 \cdots O_T$$

$$= \mathfrak{H}\ \mathfrak{H}\ \mathfrak{T}\ \mathfrak{T}\ \mathfrak{T}\ \mathfrak{H}\ \mathfrak{T}\ \mathfrak{T}\ \mathfrak{H} \cdots \mathfrak{H}$$

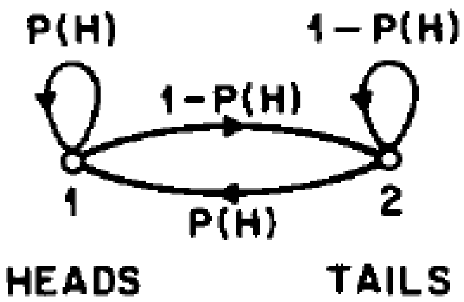
$\mathfrak{H}$ cara $\quad$ $\mathfrak{T}$ sello

Modelos Ocultos de Markov

- ¿Cómo construir un HMM que explique la secuencia observada de caras y sellos?
  - Problemas:
  1. Que estados corresponderan a caras y sellos?
  2. Cuantos estados deberá tener el modelo?
  3. Cual es el mejor modelo?

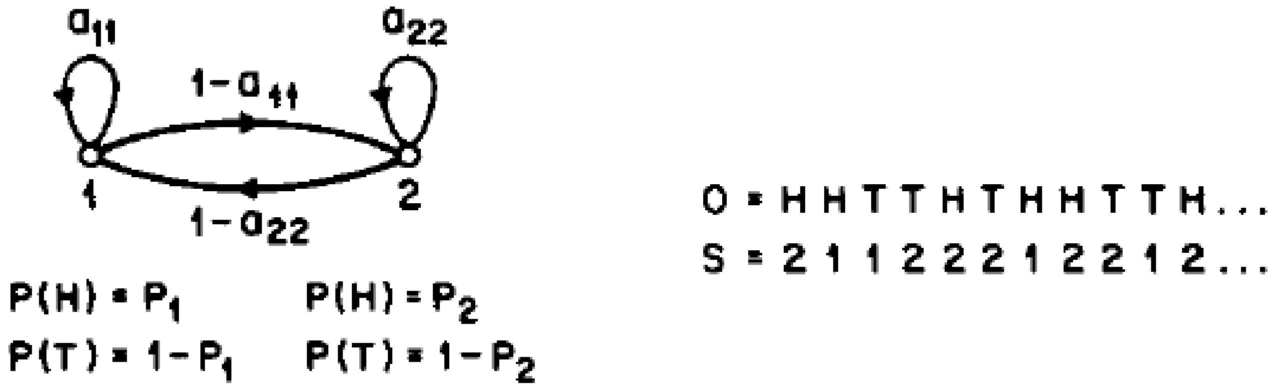
Modelos Ocultos de Markov

Si tenemos una sola moneda el modelo podría tener dos estados, cada estado corresponderá a cada lado de la moneda, se trata de un modelo observable, solo se necesita saber el valor de un bias, para completar la especificación



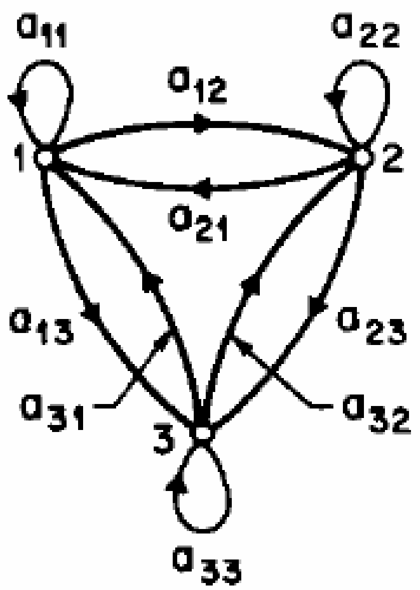
Modelos Ocultos de Markov

- Otra forma, dos estados, cada estado corresponde a dos diferentes monedas con bias, cada estado es caracterizado por la distribucion de probabilidad de cabezas y colas, las tranciciones dependen de una matriz de transicion de estados



Modelos Ocultos de Markov

- Otra manera



STATE

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

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

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

Modelos Ocultos de Markov

- Se usan tres monedas con bias y se escogen de entre las tres, basado en algún evento de probabilidad

- ¿Cuál es el mejor modelo para las observaciones actuales?

Modelos Ocultos de Markov

- Elementos
  1. ***N*** número de estados en el modelo
     1. Estados individuales $S = \{S_1 \ldots S_N\}$ç
     2. Estado en el tiempo t = q_t
  2. ***M*** numero de distintos símbolos de observación por estado, corresponde con la salida física del sistema modelado, por ejemplo para los experimentos de las monedas será cara o sello
     1. Los símbolos individuales se denotarán como

        $V = \{v_1 \ldots v_M\}$

Modelos Ocultos de Markov

3. La distribución de probabilidad de transiciones de estados ***A = {a_{ij}}*** donde

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \qquad 1 \leq i, j \leq N$$

4. La distribucion de probabilidad de las observaciones de los símbolos en el estado ***j***, ***B = { $b_j(k)$ }*** donde

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j], \qquad 1 \leq j \leq N$$

$$1 \leq k \leq M.$$

Modelos Ocultos de Markov

5. La distribucion inicial de estado $\pi = \{\pi_i\}$
   donde

$$\pi_i = P[q_1 = S_i], \qquad 1 \le i \le N.$$

Dado apropiados valores de ***N,M,A,B*** y ***π*** el HMM puede ser usado para generador para obtener la secuencia de observaciones

$$O = O_1 \, O_2 \cdots O_T$$

Donde cada observacion $\boldsymbol{O_t}$ es uno de los símbolos de ***V*** y ***T*** es el número de observaciones enla secuencia

Modelos Ocultos de Markov

- Uso del HMM como generador (algoritmo)

1. Escoger un estado inicial, $q_1 = S_i$ de acuerdo a la distribucion inicial de estado ***π***
2. t ← 1
3. Escoger $O_t = v_k$ de acuerdo a la distribucion de probabilidad de símbolos en el estado ***S_i***, $b_i(k)$
4. Transitar al nuevo estado $q_{t+1} = S_j$ conforme a la distribucion de probabilidad de transiciones de estados para el estado $\boldsymbol{S_i}$, a_{ij}

Modelos Ocultos de Markov

5. t $\leftarrow$ t +1, retornar al paso 3 si $\boldsymbol{t} < \boldsymbol{T}$, de otra manera terminar

El algoritmo anterior como un HMM puede ser usado como generador de observaciones o como un modelo, dada una secuencia de observaciones como fue generada por un HMM

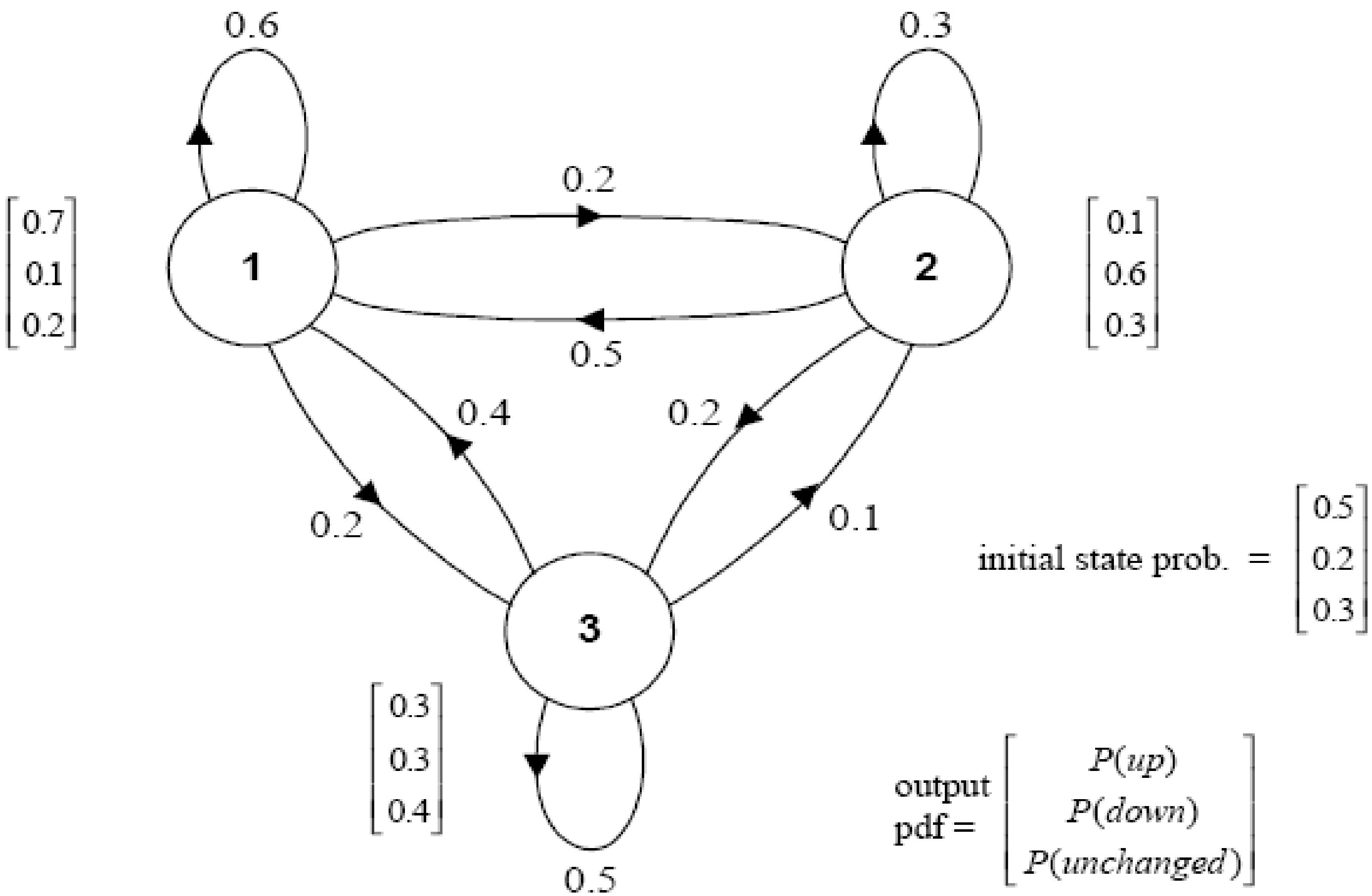
Modelos Ocultos de Markov

- Los HMM, para ser especificados necesitan los paramentros ***N*** y ***M***, tambien necesitan la especificacion de los símbolos de observacion y de las tres distribuciones de probabilidad, ***A***, ***B*** y ***π*** , Para denotar todo esto se utilizara la siguiente notación

$$\lambda = (A, B, \pi)$$

Modelos Ocultos de Markov

- Ejemplo



Modelos Ocultos de Markov

- Los tres problemas básicos

  Se deben resolver tres problemas básicos para que el modelo sea útil a aplicaciones reales

  1. Problema 1 (evaluación)
  2. Problema 2 (decodificación)
  3. Problema 3 (aprendizaje)

Modelos Ocultos de Markov

1. **Problema 1 (evaluación)** Dada una secuencia de observación $\boldsymbol{O = O_1 \ldots O_T}$ y un modelo $\boldsymbol{\lambda = (A,B,\pi)}$, como calcular eficientemente $\boldsymbol{P(O/\lambda)}$, la probabilidad de la observacion dado el modelo?
2. **Problema 2 (decodificación)** Dada una secuencia de observación $\boldsymbol{O = O_1 \ldots O_T}$ y un modelo $\boldsymbol{\lambda}$, como escoger una correspondiente secuencia de estados $\boldsymbol{Q = q_1 \ldots q_T}$ que sea óptima, es decir que mejor explique la observación
3. **Problema 3 (aprendizaje-entrenamiento)** Como ajustar los parámetros del modelo $\boldsymbol{\lambda = (A,B,\pi)}$ para maximizar $\boldsymbol{P(O/\lambda)}$?

Modelos Ocultos de Markov

- Ejemplo en RAH
  - Por cada palabra construir un HMM de N estados
  - La señal de habla de determinada palabra es una secuencia de vectores espectrales (codificados),
  - Se puede establecer un codificador espectral con ***M*** vectores espectrales unicos, siendo cada observacion el índice de este codificador más cercano a la señal

Modelos Ocultos de Markov

- Para cada palabra del vocabulario realizar un entrenamiento para construir un modelo de palabra, esto es hecho solucionando el problema numero 3
- Para saber cual es la secuencia de estados y hacer un refinamiento porsterior, se solucionará el problema numero 2
- El reconocimiento de una palabra se hará solucionando el problema 1

Solucion a los tres básicos problemas de los HMM

- Solución al problema 1
  - Calcular la probabilidad de la secuencia $\boldsymbol{O = O_1 \ldots O_T}$ dado el modelo $\boldsymbol{\lambda = (A,B,\pi)}$, es decir $\boldsymbol{P(O/\lambda)}$
  - La manera mas ingenua sería ennumerar toda la posible secuencia de estados de longitud $\boldsymbol{T}$, por ejemplo si consideramos una secuencia fija de estados

$$Q = q_1 \, q_2 \cdots q_T$$

Solucion problema 1

- La probabilidad de la secuencia de observacion para la secuencia de estados anterior es

$$P(O|Q, \lambda) = \prod_{t=1}^{T} P(O_t|q_t, \lambda)$$

$$P(O|Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdots b_{q_T}(O_T)$$

- La probabilidad de la secuencia de estados dado el modelo es

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T}$$

Solucion problema 1

- La probabilidad de que ***O*** y ***Q*** ocurran simultaneamente :

$$P(O, Q|\lambda) = P(O|Q, \lambda)\, P(Q, \lambda)$$

- La probabilidad de ***O*** dado el modelo ***P(O/ λ)*** es la sumatoria de todas las posibles combinacionesd e secuencias de estados ***q***

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q, \lambda)\, P(Q|\lambda)$$

$$= \sum_{q_1, q_2, \cdots, q_T} \pi_{q_1} b_{q_1}(O_1)\, a_{q_1 q_2} b_{q_2}(O_2)$$

$$\cdots\, a_{q_{T-1} q_T} b_{q_T}(Q_T).$$

Solucion problema 1

- La ecuación anterior se interpreta como sigue:
  1. En el tiempo t=1, estamos en el estado q_1 con probabilidad π_{q1} y generamos el símbolo O_1, con probabilidad $b_{q1}(O_1)$,
  2. En el tiempo t=t+1, hacemos una transicion al estado q_2, con probabilidad a_{q1q2}, y generamos el simbolo O_2 con probabilidad $b_{q2}(O_2)$
  3. …
  4. Hasta el tiempo ***T***
  5. Se cambia la secuencia de estados

Solucion problema 1

Complejidad $\boldsymbol{2T*N^T}$ cálculos, pues existen $\boldsymbol{N^T}$ posibles secuencias de estados y cada secuencia de estados requiere $\boldsymbol{2T}$ calculos

No es práctico pues para pequeños valores de ***N*** y ***T*** por ejemplo ***N = 5 y T = 100*** se necesitarán aproximadamente $2*100*5^{100} \sim= 10^{72}$ operaciones de cálculo

Solucion problema 1

- Afortunadamente existe un algoritmo, que realiza lo mismo en menos operaciones, este es el algoritmo ***forward-backward***

Solucion problema 1

- Algoritmo ***forward-backward***

  considerar

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

como la probabilidad de la secuencia de observacion parcial $\boldsymbol{O_1 \ldots O_t}$ hasta el tiempo $\boldsymbol{t}$ y el estado $\boldsymbol{S_i}$ en el tiempo $\boldsymbol{t}$ dado el modelo $\boldsymbol{\lambda}$.

Algoritmo *forward-backward*

1. Inicialización

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

2. Inducción

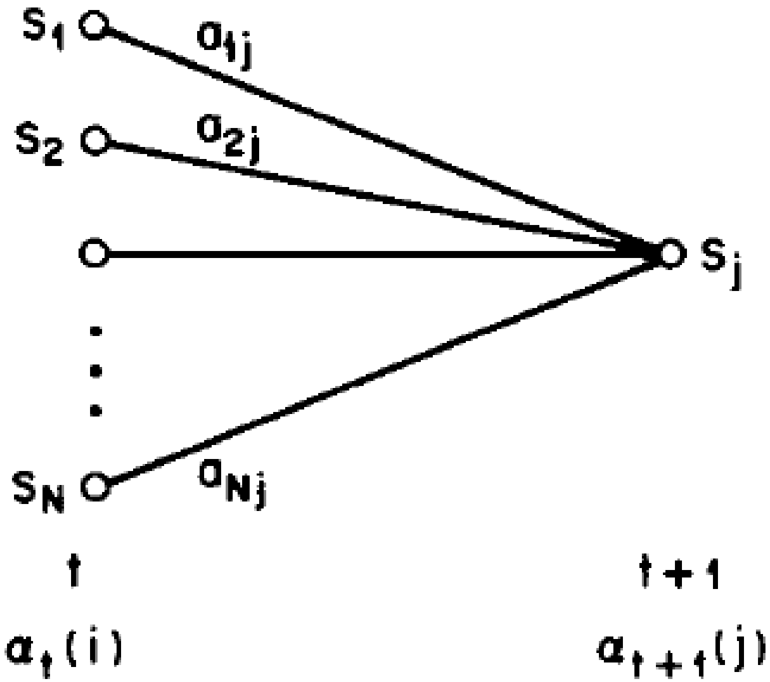
$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \qquad 1 \le t \le T-1$$

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

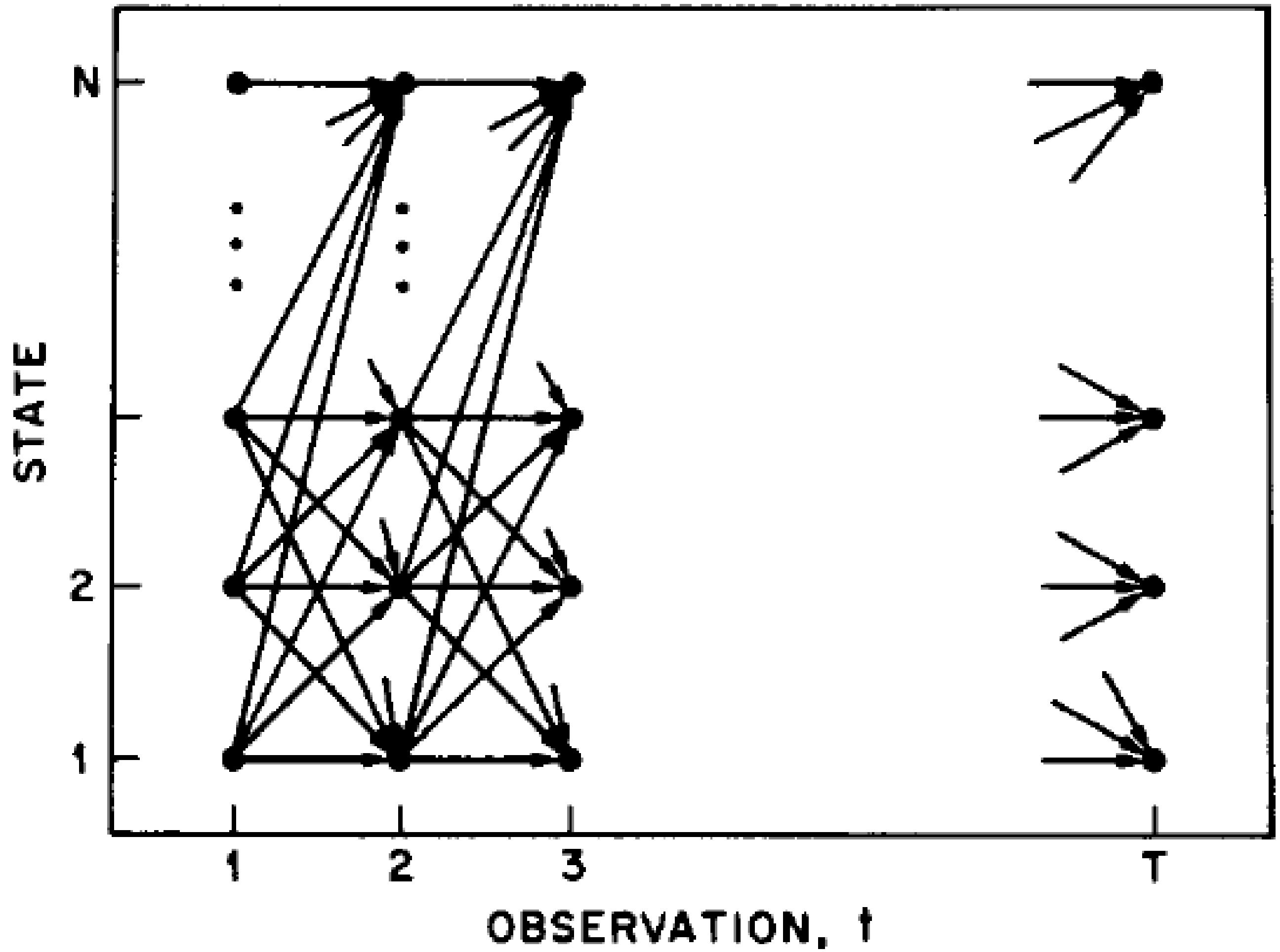
1. Terminación

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

Algoritmo *forward-backward*

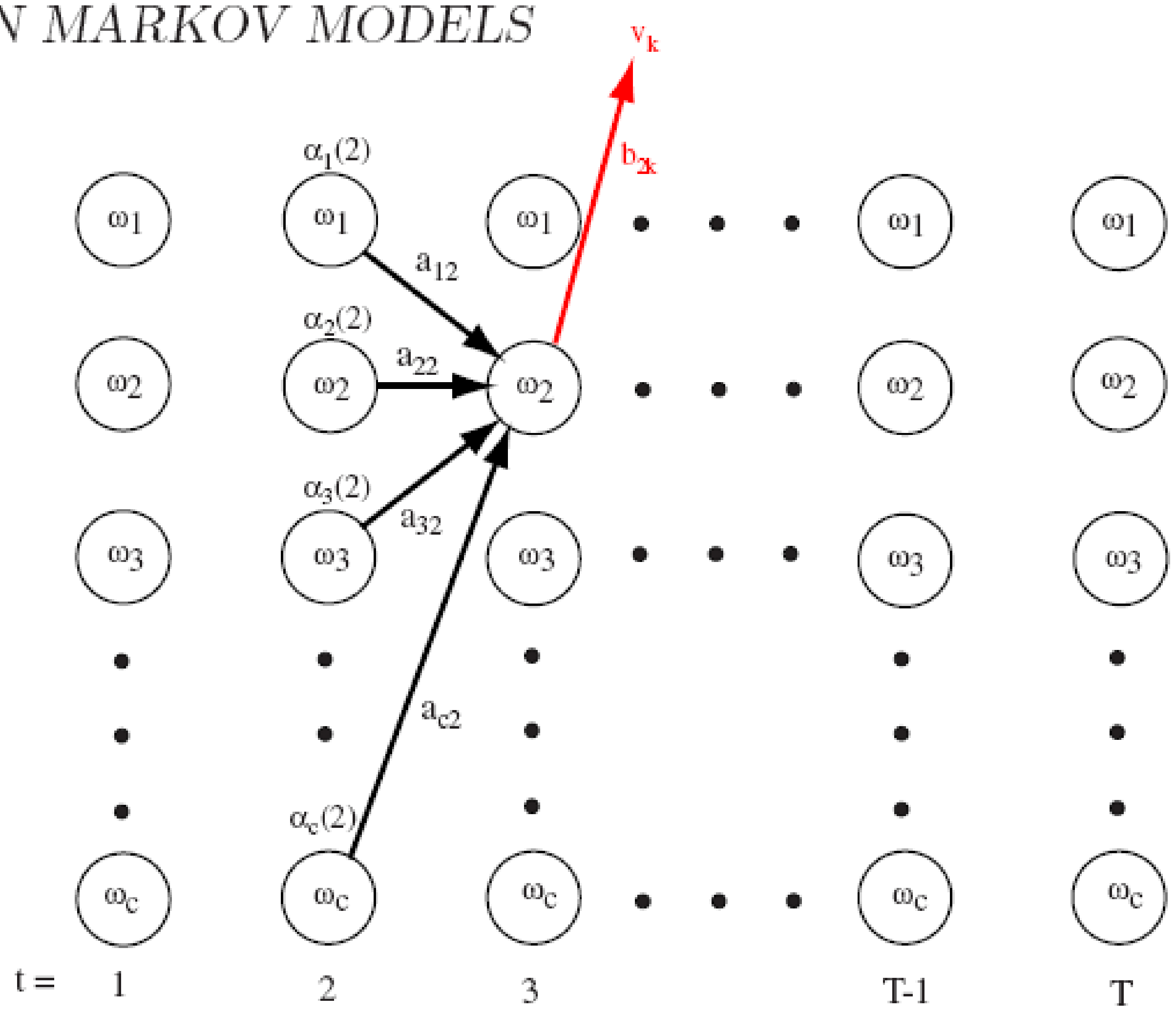


Algoritmo *forward-backward*

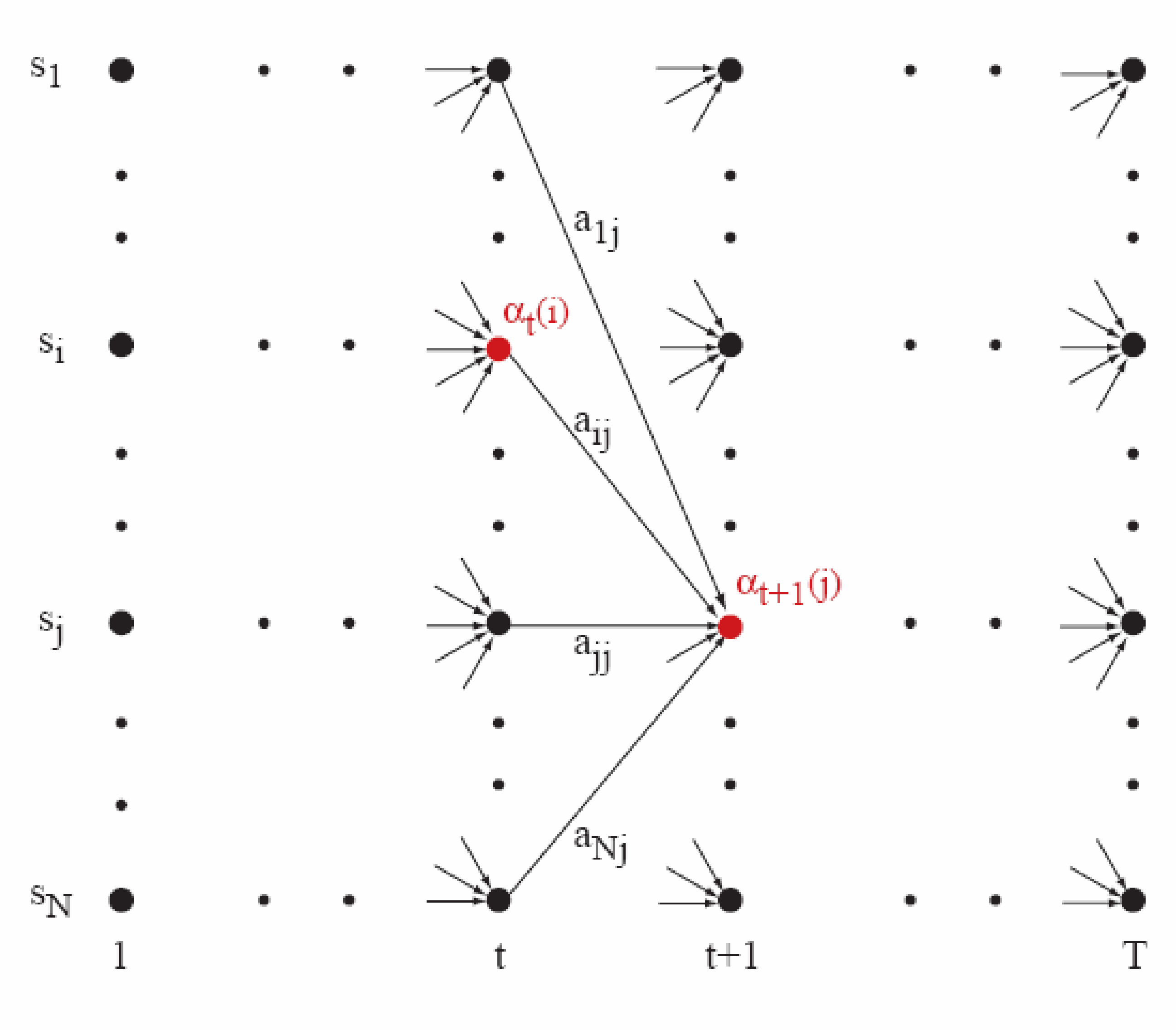


Algoritmo *forward-backward*

*HIDDEN MARKOV MODELS



Algoritmo *forward-backward*



Algoritmo *forward*-*backward*

- Complejidad $\boldsymbol{N^2T}$ operaciones de cálculo, para $\boldsymbol{N = 5}$ ***y*** $\boldsymbol{T = 100}$ se necesitarán solamente 3000 cálculos frente a los 10^{72} con un cálculo directo

Aproximadamente menos 69 ordenes de magnitud !!!

Algoritmo *forward*-*backward*

- De similar manera podemos considerar

$$\beta_t(i) = P(O_{t+1}\, O_{t+2} \cdots O_T | q_t = S_i, \lambda)$$

la probabilidad de la secuencia de observaciones parciales de ***t+1*** hasta ***T*** dado el estado $\boldsymbol{S_i}$ en el tiempo ***t*** y el modelo $\boldsymbol{\lambda}$

Algoritmo *forward-backward*

1. Inicialización

$$\beta_T(i) = 1, \qquad 1 \leq i \leq N.$$

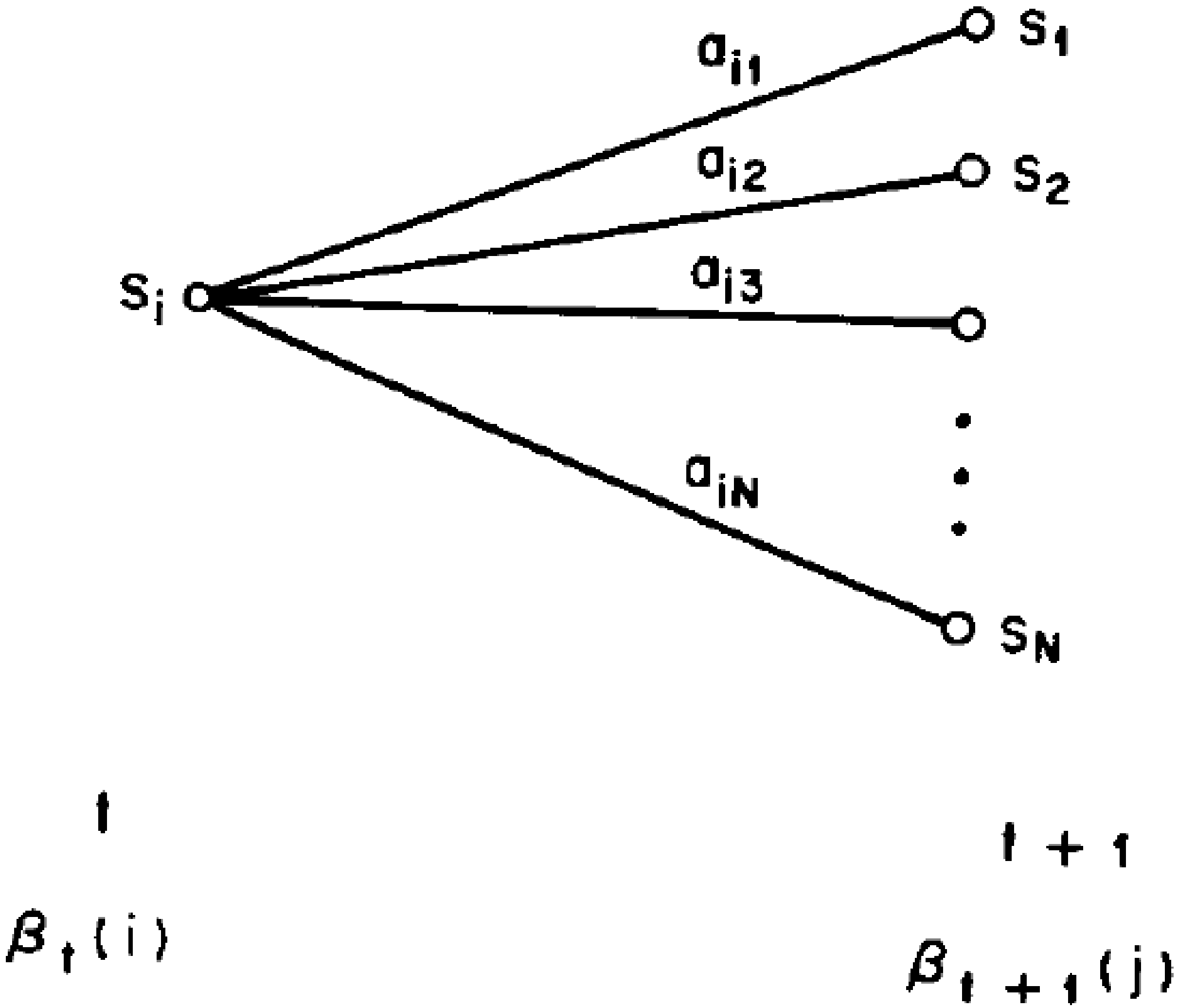
2. Inducción

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1})\, \beta_{t+1}(j),$$

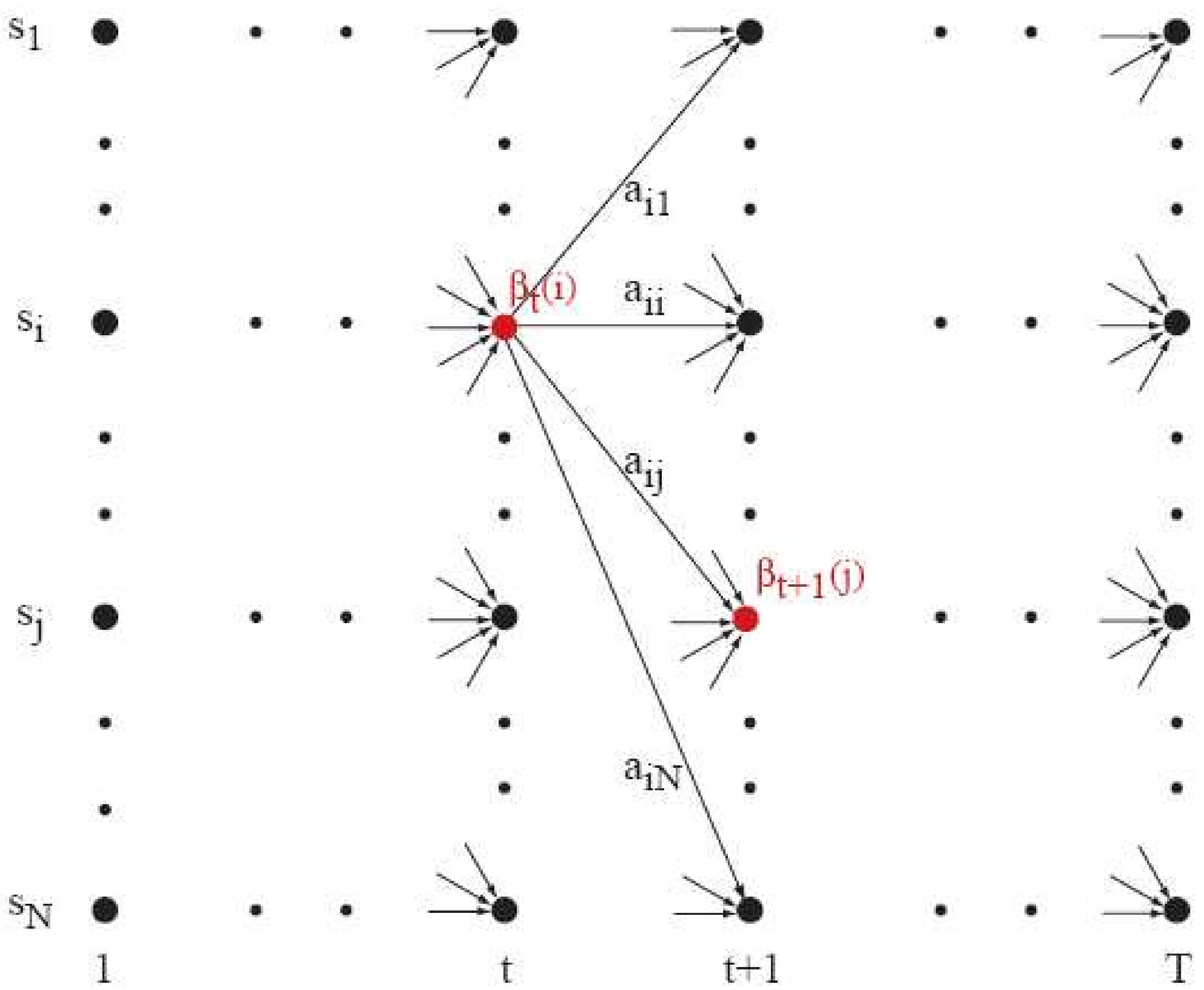
$$t = T-1, T-2, \cdots, 1, \; 1 \leq i \leq N.$$

La parte ***Backward*** del algoritmo se usará para solucionar el problema 2 más no el problema 1

Algoritmo *forward-backward*



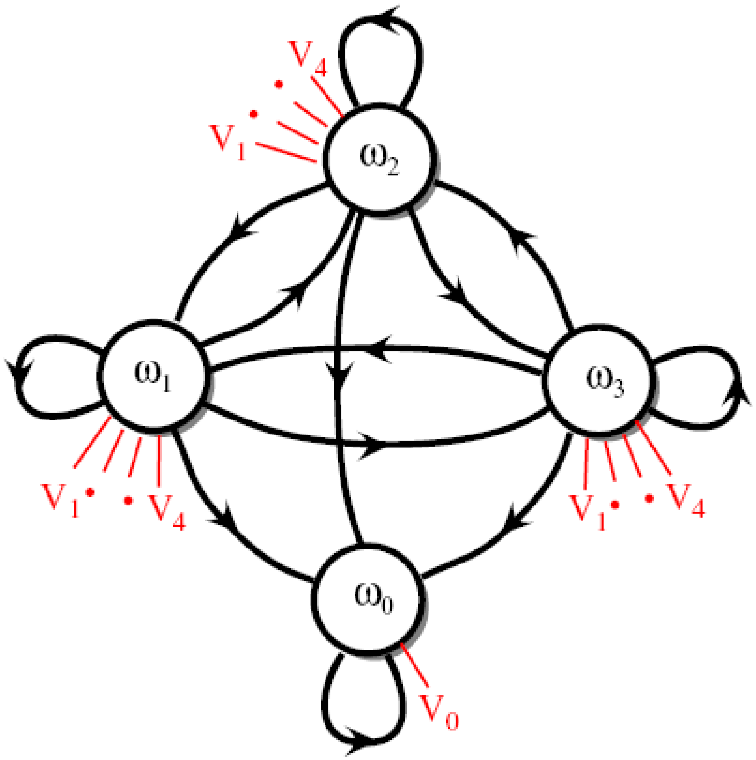
Algoritmo *forward-backward*



Ejemplo

$$a_{ij} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.1 & 0.4 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0.8 & 0.1 & 0.0 & 0.1 \end{pmatrix}$$

$$b_{jk} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0 & 0.1 & 0.1 & 0.7 & 0.1 \\ 0 & 0.5 & 0.2 & 0.1 & 0.2 \end{pmatrix}$$



Ejemplo

- ¿Cuál es la probabilidad de generar la secuencia $\boldsymbol{V^5} = \{v_3, v_1, v_3, v_2, v_0,\}$?