1. MODELOS OCULTOS DE MARKOV

Un Modelo Oculto de Markov (Hidden Markov Model HMM) es un proceso estocástico que consta de un proceso de Markov no observado (oculto) $\boldsymbol{q} = \{q_t\}_{t \in N}$ y un proceso observado $\boldsymbol{O} = \{o_t\}_{t \in N}$ cuyos estados son dependientes estocásticamente de los estados ocultos, es decir, es un proceso bivariado $\boldsymbol{(q,O)}$. Los HMMs se pueden considerar también como sistemas generativos estocásticos, los cuales se emplean en la modelación de series de tiempo.

1.1 CADENAS DE MARKOV

Una cadena de Markov $\boldsymbol{q} = \{q_t\}_{t \in N}$ es un proceso estocástico de Markov discreto. Un proceso estocástico se llama de Markov si conocido el presente, el futuro no depende del pasado, esto quiere decir, que dada una variable estocástica q_{t-1} que denota el estado del proceso en el tiempo $t-1$, entonces la probabilidad de transición en el momento t se define como $P[q_t = \sigma_t \mid q_{t-1} = \sigma_{t-1}]$. Formalmente, una cadena de Markov se define como (Q,A), donde $Q = \{1,2,,...,N\}$ son los posibles estados de la cadena y $A = (a_{ij})_{nxn}$ es una matriz de transición de estados en el modelo. Si $A(t) = a_{ij}(t)_{nxn}$ es independiente del tiempo entonces el proceso se llama homogéneo y las probabilidades de transición de estados son de la forma $a_{ij}(t) = P[q_t = j \mid q_{t-1} = i]$ con las siguientes propiedades:

i) $0 \leq a_{ij} \leq 1, \quad 1 \leq i,j \leq \boldsymbol{N};$

ii) $\sum_{j=1}^{N} a_{ij} = 1, \quad 1 \leq i \leq \boldsymbol{N}.$

La condición fundamental de que sea una cadena de Markov establece que las probabilidades de transición y emisión dependen sólamente del estado actual y no del pasado, esto es, $P[q_t = j \mid q_{t-1} = i, q_{t-2} = k, \ldots] = P[q_t = j \mid q_{t-1} = i] = a_{ij}(t)$.

En este trabajo se considera el conjunto de estados finitos.

1.2 DEFINICION DE MODELOS OCULTOS DE MARKOV

Un modelo oculto de Markov es una cadena de $\boldsymbol{q}$ junto con un proceso estocástico que toma valores en un alfabeto Σ y el cual depende de $\boldsymbol{q}$.

Estos sistemas evolucionan en el tiempo pasando aleatoriamente de estado a estado y emitiendo en cada momento al azar algún símbolo del alfabeto Σ. Cuando se encuentra en el estado $q_{t-1} = i$, tiene la probabilidad a_{ij} de moverse al estado $q_t = j$ en el siguiente instante y la probabilidad $b_j(k)$ de emitir el símbolo $o_t = v_k$ en el tiempo t.

Sólamente los símbolos emitidos por el proceso $\boldsymbol{q}$ son observables, pero no la ruta o secuencia de estados $\boldsymbol{q}$, de ahí el calificativo de "oculto" de Markov, ya que el proceso de Markov $\boldsymbol{q}$ es no observado.

El siguiente ejemplo ilustra un proceso $\boldsymbol{q}$ independiente del tiempo. Supóngase que en un salón se encuentra un número $\boldsymbol{N}$ muy grande de urnas de vidrio. Dentro de cada urna se tiene una cantidad $\boldsymbol{M}$ de bolas de colores. Un mago está en el salón y de acuerdo con algún procedimiento aleatorio elige una urna inicial. De ésta saca al azar una bola y registra su color como una observación. La bola es retornada a la urna de la cual fué seleccionada. A continuación selecciona una nueva urna de acuerdo con un procedimiento aleatorio que depende de la urna actual y la elección de alguna bola es repetida. Este proceso completo se realiza en un tiempo $\boldsymbol{T}$ y genera una secuencia de observación finita de colores $\boldsymbol{O}$ de longitud $\boldsymbol{T}$, la cual puede modelarse como la salida observable de un HMM. Se asume que las urnas son seleccionadas independientemente.

Urna 1 urna 2 . . . urna N

Figura. 1.1 Modelo de urnas y bolas de $\boldsymbol{N}$ estados que ilustra el caso general de un HMM con símbolos discretos.

Los siguientes son ejemplos de posibles secuencias de observación del modelo de las urnas y las bolas:

$\boldsymbol{O^1}$ = (amarillo, verde, azul, verde, rojo, amarillo, naranja, rojo, verde, azul, amarillo),

$\boldsymbol{O^2}$ = (amarillo, rojo, verde, rojo, azul, naranja, verde, rojo, azul, amarillo, rojo, verde),

$\boldsymbol{O^3}$ = (rojo, azul, amarillo, rojo, azul, vede, rojo, amarillo, naranja, naranja, verde, rojo),

$\boldsymbol{O^4}$ = (rojo, verde, naranja, rojo, rojo, azul, verde, amarillo, azul, rojo, verde, rojo).

El alfabeto es:

$\boldsymbol{\Sigma}$ = verde, azul, rojo, amarillo, naranja

Los estados ocultos son:

$\boldsymbol{Q} = \{1,2,...,N\}$

Las probabilidades de obtener un color en cada urna son:

urna 1	urna 2	. . .	urna N
P(rojo) = $b_1(1)$	P(rojo) = $b_2(1)$	. . .	P(rojo) = $b_N(1)$
P(azul) = $b_1(2)$	P(azul) = $b_2(2)$	. . .	P(azul) = $b_N(2)$
P(verde) = $b_1(3)$	P(verde) = $b_2(3)$	. . .	P(verde) = $b_N(3)$
P(amarillo) = $b_1(4)$	P(amarillo) = $b_2(4)$	. . .	P(amarillo) = $b_N(4)$
. . .	. . .	. . .	. . .
P(naranja) = $b_1(\boldsymbol{M})$	P(naranja) = $b_2(\boldsymbol{M})$	. . .	P(naranja) = $b_N(\boldsymbol{M})$

Las probabilidades de pasar de una urna a otra son:

P(1,1) = a_{11}	P(2,1) = a_{21}	P(3,1) = a_{31}	. . .	P(N,1) = a_{N1}
P(1,2) = a_{12}	P(2,2) = a_{22}	P(3,2) = a_{32}	. . .	P(N,2) = a_{N2}
. . .	. . .	. . .	. . .	. . .
P(1,N) = a_{1N}	P(2,N) = a_{2N}	P(3,N) = a_{3N}	. . .	P(N,N) = a_{NN}

El primer problema consiste en decidir cual proceso es representado por los estados y después decidir cuantos estados pueden estar en el modelo.

Como se ilustró antes, el HMM más simple que corresponda al comportamiento de este proceso es aquel en el cual cada estado representa una urna específica y cada color representa un posible símbolo de observación. Por cada estado se define una probabilidad de extraer una bola (color) y una probabilidad de pasar a la siguiente urna. Los colores de las bolas dentro de cada urna pueden o no ser los mismos y pueden existir números diferentes de bolas de cada color en cada urna. Por lo tanto, una observación aislada de un color en particular no dice inmediatamente de cuál urna procede.

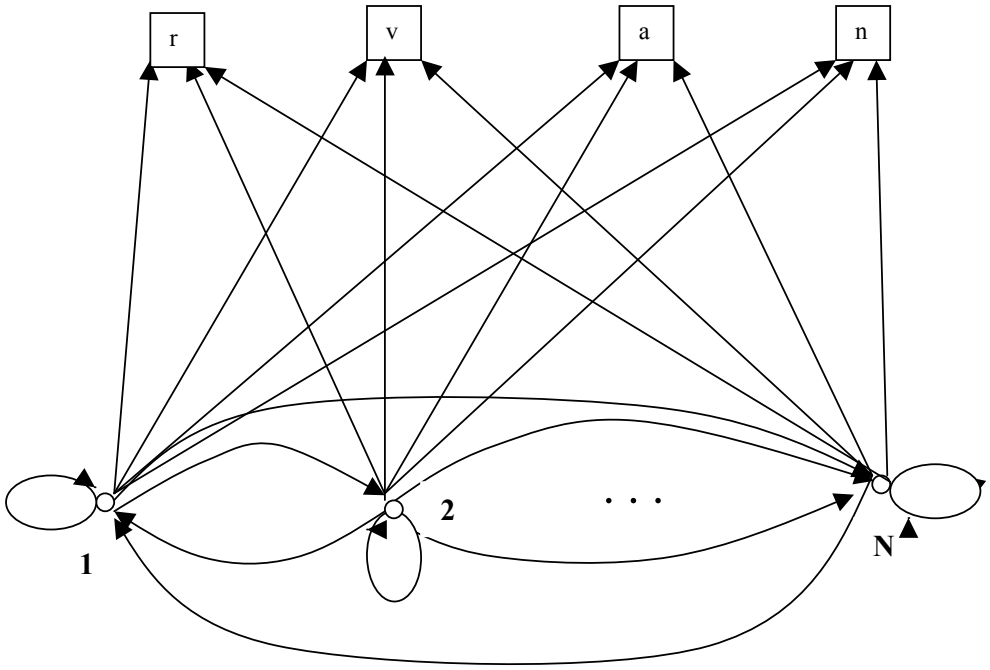

Figura 1.2 Arquitectura del grafo del modelo de urnas y bolas

1.3 ARQUITECTURAS DE HMMS

Un HMM puede ser representado como un grafo dirigido de transiciones/emisiones como se ilustra en la figura 1.2. La arquitectura específica que permita modelar de la mejor forma posible las propiedades observadas depende en gran medida de las características del problema. Las arquitecturas mas usadas son:

1) *Ergódicas o completamente conectadas* en las cuales cada estado del modelo puede ser alcanzado desde cualquier otro estado en un número finito de pasos (figura 1.2).
2) *Izquierda-derecha*, *hacia adelante* o *Bakis* las cuales tienen la propiedad de que en la medida que el tiempo crece se avanza en la secuencia de observación asociada $\boldsymbol{O}$, y en esa misma medida el índice que señala el estado del modelo permanece o crece, es decir, los estados del sistema van de izquierda a derecha (figura 1.3). En secuencias biológicas y en reconocimiento de la voz estas arquitecturas modelan bien los aspectos lineales de las secuencias.
3) *Izquierda-derecha paralelas*, son dos arquitecturas *izquierda-derecha* conectadas entre sí (figura 1.4).

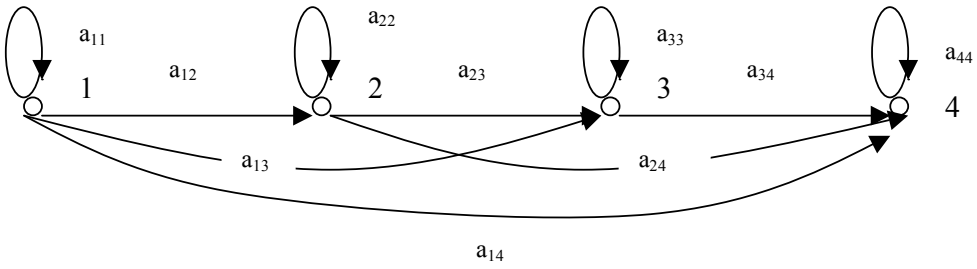

Figura 1.3 Modelo izquierda-derecha con 4 estados

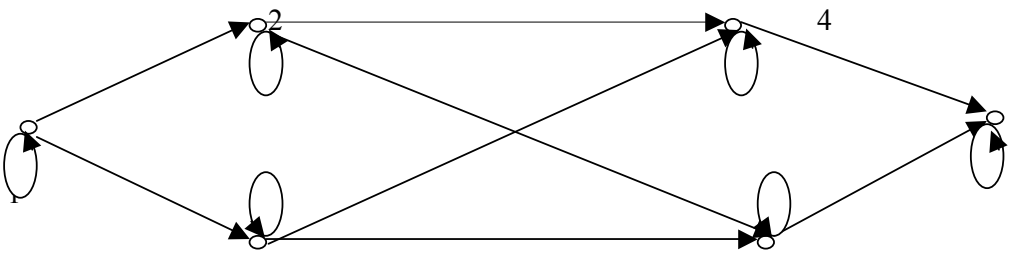

Figura 1.4 Modelo izquierda derecha paralelo con 6 estados

1.4 DEFINICION DE LOS ELEMENTOS DE UN HMM

Formalmente, un HMM discreto de primer orden se define como una cinco-tupla

$$\boldsymbol{\lambda = (\Sigma, Q, A, B, \pi)}$$

donde

i) $\boldsymbol{\Sigma} = \{v_1, v_2, ..., v_{\boldsymbol{M}}\}$ es un alfabeto o conjunto discreto finito de $\boldsymbol{M}$ símbolos.

ii) $\boldsymbol{Q} = \{1,2,...,\boldsymbol{N}\}$ es un conjunto finito de $\boldsymbol{N}$ estados.

iii) $\boldsymbol{A} = (a_{ij})_{\boldsymbol{N}x\boldsymbol{N}}$ es una matriz de probabilidades de transición donde a_{ij} es la probabilidad de transición desde el estado i al estado j, para todo $i,j \in \boldsymbol{N}$.

iv) $\boldsymbol{B} = (b_j(o_t))_{\boldsymbol{N}x\boldsymbol{M}}$ es un vector de probabilidades de emisión de símbolos, uno por cada estado, donde $b_j = (b_{j1}, b_{j2}, \ldots, b_{j\boldsymbol{M}})$ es la probabilidad de emisión del símbolo v_k del alfabeto en el estado j.

v) $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_{\boldsymbol{N}})$ es un vector de probabilidades del estado inicial q_o en $\boldsymbol{Q}$.

Las probabilidades de iniciación, transición y emisión son los parámetros del modelo. Un HMM puede ser usado como un generador de secuencias de observaciones $\boldsymbol{O} = (o_1, o_2, ..., o_{\boldsymbol{T}})$ donde:

- o_t es uno de los símbolos de Σ para $t \in \{1,2,...,\boldsymbol{T}\}$
- $\boldsymbol{T}$ es la longitud de la secuencia de observación $\boldsymbol{O}$, es decir, el número de observaciones en la secuencia
- $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ son los parámetros del modelo.

Un HMM define una medida de probabilidad μ, sobre el espacio de secuencias $\boldsymbol{\Sigma}^*$.

1.5 PROBLEMAS BASICOS DE LAS HMMs

Existen tres problemas básicos relacionados con los HMMs:

1. Calcular eficientemente $P(\boldsymbol{O}|\lambda)$ la probabilidad de la secuencia de observación $\boldsymbol{O}$ dado el modelo $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ y la secuencia de observación $\boldsymbol{O} = (o_1 o_2 ... o_{\boldsymbol{T}})$.
2. Encontrar la trayectoria mas probable $\boldsymbol{q} = (q_1 q_2 ... q_{\boldsymbol{T}})$ dado el modelo λ y la secuencia de observación $\boldsymbol{O} = (o_1 o_2 ... o_{\boldsymbol{T}})$, es decir, $\boldsymbol{q} = arg_{r \in \boldsymbol{Q}^*}\{ max_{r \in Q^*} P(r)\}$.
3. Ajustar los parámetros $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}$ para maximizar $P(\boldsymbol{O}|\lambda)$

1.5.1 Estimativos iniciales de los parámetros del HMM

Existe diferentes formas de incorporar la información inicial para el diseño del HMM y estimar sus parámetros. La experiencia ha mostrado [9] que los estimativos iniciales uniformes o aleatorios de los parámetros $\boldsymbol{\pi}$ y $\boldsymbol{A}$ son adecuados en la mayoría de los casos. Y que los buenos estimativos iniciales para el parámetro $\boldsymbol{B}$ son útiles en el caso de secuencias con símbolos discretos y esenciales en el caso de distribuciones continuas. Por ejemplo, en biología computacional [1] se utiliza la distribución de Dirichlet para las probabilidades de transición y emisión y las distribuciones uniformes o aleatorias para las probabilidades del estado inicial.

1.5.2 Evaluación de la probabilidad de la secuencia

El problema 1 de los HMMs consiste en calcular la probabilidad de la secuencia de observación $\boldsymbol{O} = (o_1 o_2 \ldots o_T)$, dado el modelo λ, es decir, $\boldsymbol{P(O|\lambda)}$. La forma mas simple de resolver el problema 1 consiste en enumerar todas las posibles secuencias de estado de longitud $\boldsymbol{T}$. Una de dichas secuencias es de la forma

$$q = (q_1 q_2 \ldots q_T)$$

donde q_1 es el estado inicial.

La probabilidad de la secuencia de observación $\boldsymbol{O}$ dada la anterior secuencia $\boldsymbol{q}$ es:

$$P(\mathbf{O}|\mathbf{q},\lambda) = \prod_{t=1}^{T} P(o_t | q_t, \lambda) = b_{q1}(o_1) b_{q2}(o_2) \ldots b_{qT}(o_T)$$

La probabilidad de la secuencia de estados $\boldsymbol{O}$ es:

$$P(\mathbf{q}|\lambda) = \pi_{q1} a_{q1q2} a_{q2q3} \ldots a_{qT-1qT}$$

La probabilidad de que $\boldsymbol{O}$ y q ocurran simultáneamente está dada por:

$$P(\mathbf{Q},\mathbf{q}|\lambda) = P(\mathbf{O}|\mathbf{q},\lambda) P(\mathbf{q}|\lambda)$$

La probabilidad de $\boldsymbol{O}$ sobre todas las posibles secuencias de estados Q, es el cálculo de:

$$P(O \mid \lambda) = \quad P(O \mid q, \lambda)P(q \mid \lambda) = \pi_{q1} b_{q1}(o_1) a_{q1q2} \ldots a_{qT-1qT} b_{qT}(o_T)$$

Esta expresión no permite un cómputo eficiente de la probabilidad debido a que el número de rutas en una arquitectura es exponencial. En cada tiempo t = 1,2, ...,$\boldsymbol{T}$ se tienen $\boldsymbol{N}$ posibles estados alcanzables, por lo tanto $\boldsymbol{N^T}$ operaciones (figura 1.5). Existe un procedimiento mas eficiente para calcular dicha probabilidad, denominado *algoritmo de avance (forward algorithm)*.

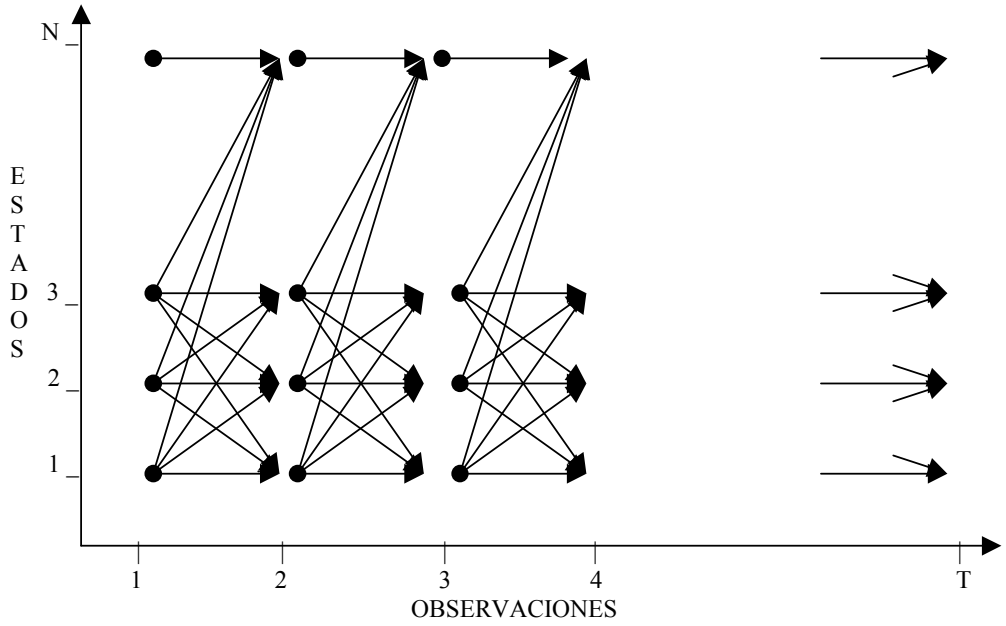

Figura 1.5 Rutas de un modelo ergódico con $\boldsymbol{N}$ estados y $\boldsymbol{T}$ observaciones

Algoritmo de avance (forward algorithm)

Sea $\alpha_t(i) = P(o_1 o_2 \ldots o_t, q_t = i \mid \lambda)$

la probabilidad de la secuencia de observación parcial $o_1 o_2 \ldots o_t$ en el estado i hasta el tiempo t, dado el modelo λ. Puede calcularse $\alpha_t(i)$, así:

1. Inicializa $: \alpha_1(i) = \pi_i b_i(o_1) \quad 1 \leq i \leq N$

$$2.\ \text{Inducción}: \alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad 1 \le t \le T-1,\ 1 \le j \le N$$

$$3.\ \textit{Terminación}: P(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

El número de operaciones requeridas para calcular $\alpha_t(j)$, $1 \le t \le \boldsymbol{T}$, $1 \le j \le \boldsymbol{N}$, ilustradas en la figura 1.6, es exactamente de $\boldsymbol{N}(\boldsymbol{N}+1)(\boldsymbol{T}-1)+\boldsymbol{N}$ multiplicaciones y $\boldsymbol{N}(\boldsymbol{N}-1)(\boldsymbol{T}-1)$ sumas, entonces es del orden de N^2T.

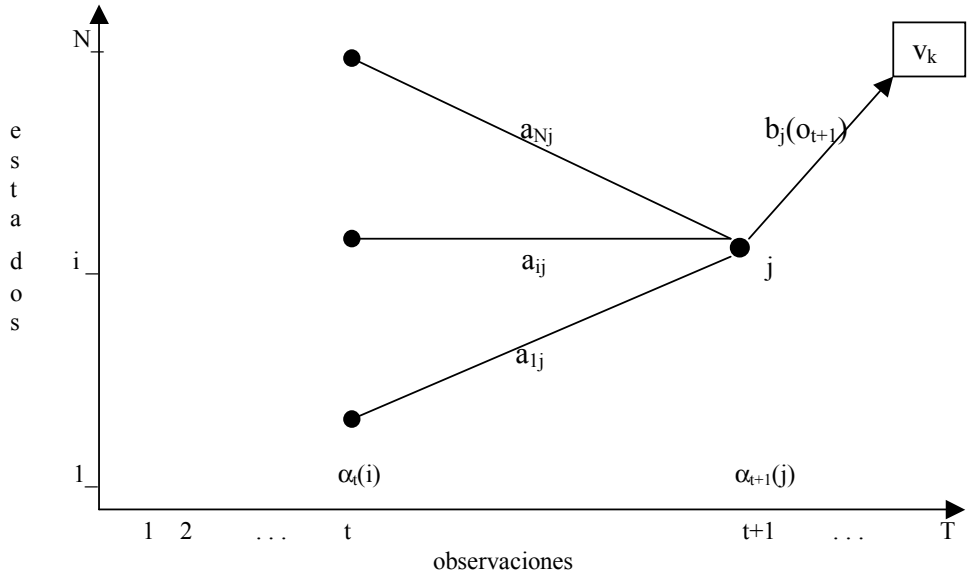

Figura 1.6 Secuencia de operaciones necesarias para el cómputo de la variable $\alpha_{t+1}(j)$

Para el aprendizaje del HMM se deben propagar las probabilidades en forma inversa. El *algoritmo de retroceso (backward algorithm)* es la versión inversa del *algoritmo de forward* se define como sigue:

Algoritmo de retroceso (backward algorithm)

Sea $\beta_t(i) = P(o_{t+1}o_{t+2}...o_T \mid q_t = i, \lambda)$

La probabilidad de la secuencia de observación parcial desde t+1 hasta el final, dado el estado i en el tiempo t y el modelo λ. La variable $\beta_t(i)$ puede resolverse como sigue:

1. Inicializa $\beta_T(i) = 1, \quad 1 \le i \le N$

2. Inducción: $\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, ..., 1, \; 1 \le i \le N$

Los cómputos necesarios de $\beta_t(i)$, $1 \le t \le \boldsymbol{T}$, $1 \le i \le \boldsymbol{N}$, es del orden de N^2T operaciones, según se ilustra en la figura 1.7.

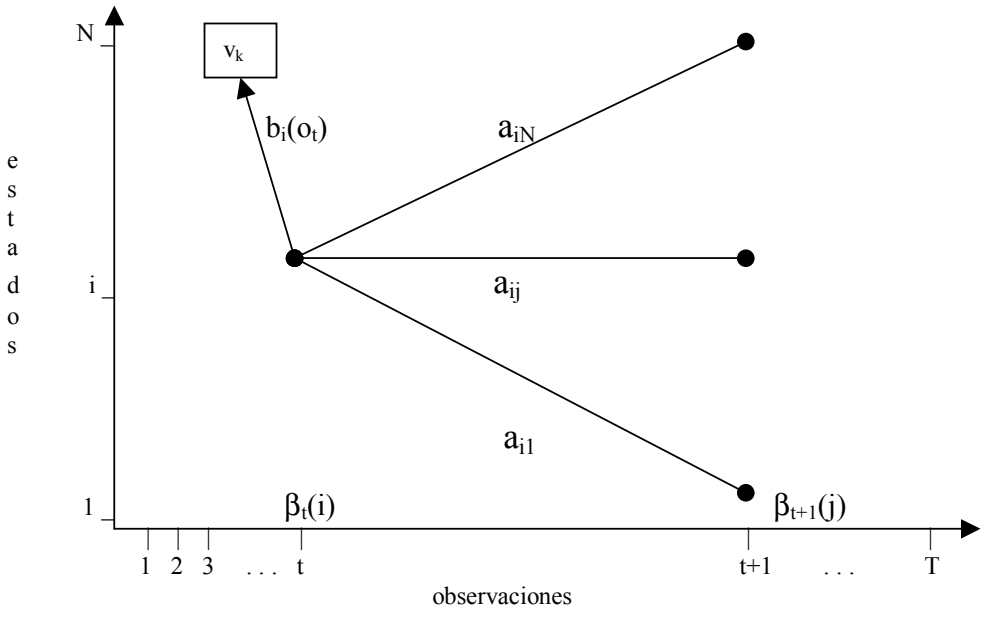

Figura 1.7 Secuencia de operaciones requeridas para el cómputo de la variable de retroceso $\beta_t(i)$

1.5.3 Secuencia optimal de estados

La ruta mas probable en un HMM es útil para el aprendizaje y para el alineamiento de secuencias con el modelo. La ruta mas probable $q = (q_1 q_2 \ldots q_T)$ para la secuencia de observación $\boldsymbol{O} = (o_1 o_2 \ldots o_T)$ puede ser calculada utilizando el *algoritmo de Viterbi*.

$$Sea \quad \delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P[q_1 q_2 \ldots q_{t-1}, q_t = i, o_1 o_2 \ldots o_t \mid \lambda]$$

la probabilidad asociada con la ruta mas probable a lo largo de un camino simple que toma en cuenta las primeras t observaciones y finaliza en el estado i. Por inducción se tiene que

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij}\right] b_j(o_{t+1})$$

Para conservar la secuencia de estados se define la variable $\psi_t(j)$. El procedimiento para encontrar la mejor secuencia de estados se establece así:

1. Inicialización :
$$\delta_1(i) = \pi_i b_i(o_1) \quad 1 \le i \le N$$
$$\Psi_1(i) = 0, \qquad 1 \le i \le N$$

2. Inducción :
$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(o_t), \quad 2 \le t \le T, \; 1 \le j \le N$$
$$\Psi_t(j) = \arg \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], \quad 2 \le t \le T, \; 1 \le j \le N$$

3. Terminación :
$$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$
$$Q_T^* = \arg \max_{1 \le i \le N} [\delta_T(i)]$$

4.. Ruta inversa :
$$q_t^* = \Psi_{t+1}(q^*_{t+1}), \qquad t = T-1, T-2, \ldots, 1.$$

El algoritmo de Viterbi está basado en el método de la programación dinámica. La estructura de retículo implementa eficientemente los cómputos del algoritmo (figura 1.8).

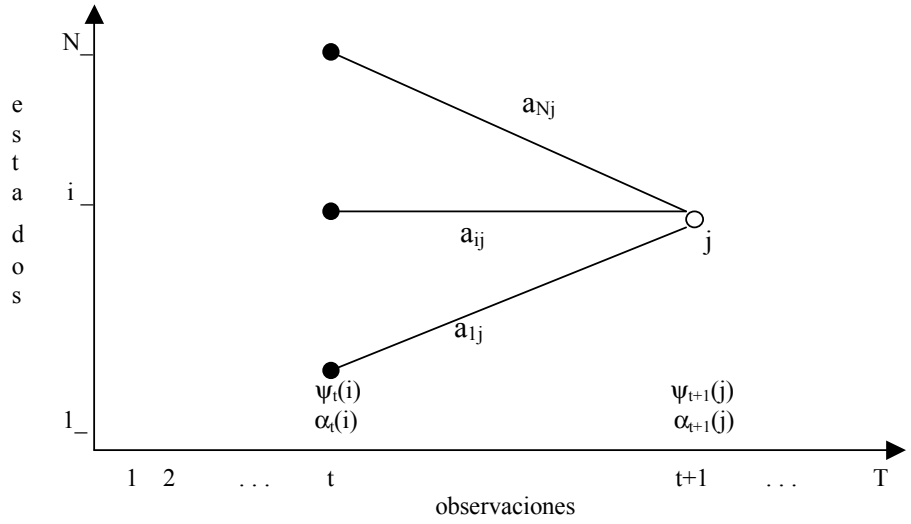

Figura 1.8 Secuencia de operaciones necesarias para calcular las variables $\alpha_t(j)$ y $\psi_t(j)$.

1.5.4 Aprendizaje del modelo

El problema mas difícil de los HMMs es determinar un método para ajustar los parámetros $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi})$ del modelo para satisfacer los criterios de optimización. No se conoce una forma analítica para fijar los parámetros que maximice la probabilidad de la secuencia de observación. Varios algoritmos están disponibles para el entrenamiento de un HMM, entre ellos, Baum-Welch o EM (expectation maximization), GEM (EM generalizado) y diferentes formas de descenso por gradiente. El procedimiento de entrenamiento con EM [9] para reestimar los parámetros $\boldsymbol{A}$,$\boldsymbol{B}$ y $\boldsymbol{\pi}$, utiliza la variable $\xi_t(i,j)$, que es la probabilidad de encontrarse en el estado i en el tiempo t y en el estado j en el tiempo $t+1$, dado el model λ y la secuencia de observación $\boldsymbol{O}$, esto es

$$\xi_t(i,j) = P[q_t = i,\ q_{t+1} = j \mid \boldsymbol{O},\ \lambda] = \frac{P[q_t = i,\ q_{t+1} = j,\ \boldsymbol{O} \mid \lambda]}{P(O|\lambda)}$$

La variable $\xi_t(i,j)$ puede reescribirse a partir de las variables $\alpha_t(j)$ y $\beta_t(i)$, así:

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

La sumatoria de $\xi_t\ (i,j)$ sobre t puede interpretarse como el número esperado de transiciones desde del estado i hasta el estado j, es decir,

$$\sum_{i=1}^{T-1}\sum_{j=1}^{N}\xi_t(i,j) = \text{número esperado de transiciones desde el estado i en } \mathbf{O}$$

$$\sum_{t=1}^{T-1}\xi_t(i,j) = \text{número esperado de transiciones desde el estado i hasta el estado j en } \mathbf{O}$$

Con las anteriores expresiones se pueden enunciar las siguientes fórmulas de reestimación para $\boldsymbol{A}$,$\boldsymbol{B}$ y $\boldsymbol{\pi}$.

π'_i = frecuencia esperada en el estado i en el tiempo $t = 1$, entonces es igual a

$$\pi'_i = \sum_{i=1}^{N}\xi_1(i,j)$$

$$a_{ij}' = \frac{\text{número esperado de transiciones desde el estado i hasta el estado j}}{\text{número esperado de transiciones desde el estado i}}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

$$b_j(k)' = \frac{\text{número esperado de veces en el estado j y observando el símbolo } v_k}{\text{número esperado de veces en el estado j}}$$

$$= \frac{\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{\substack{t=1\\ o_t=v_k}}^{N}\alpha_t(j)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{t=1}^{N}\alpha_t(j)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

Las trayectorias que satisfacen las condiciones requeridas por la ecuación se ilustran en la figura 1.9.

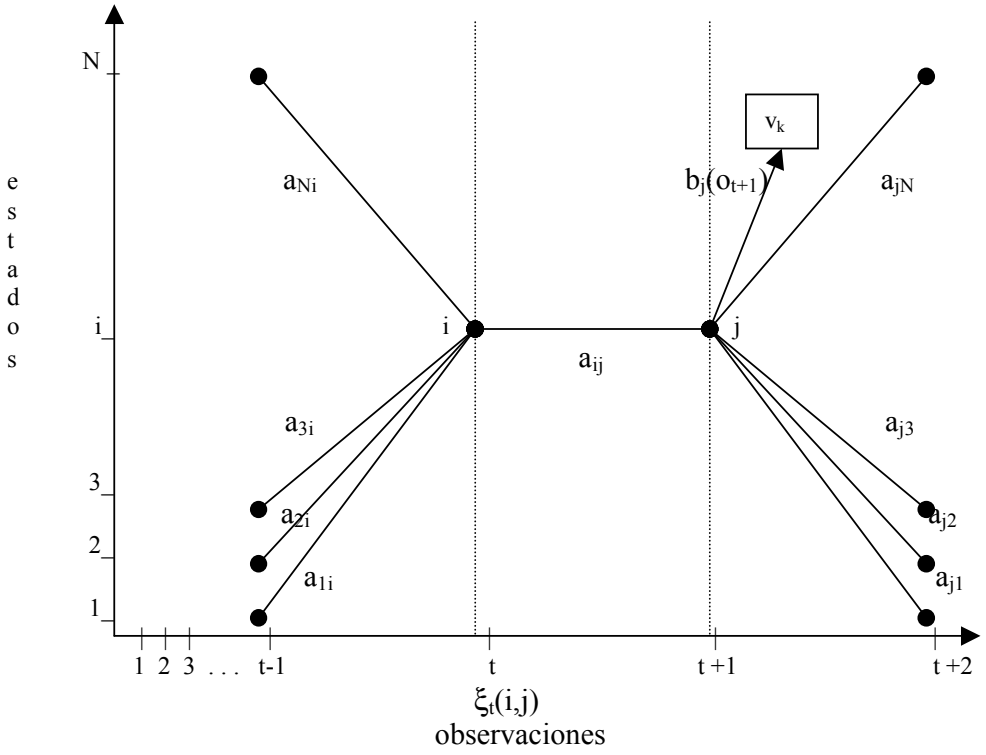

Figura 1.9 Secuencia de operaciones requeridas para el cálculo de $\xi_t(i,j)$.

Si designamos al modelo reestimado como $\lambda' = (A', B', \pi')$ pueden presentarse dos posibilidades (1) el modelo inicial $\lambda = (A, B, \pi)$ define un punto crítico de la función de probabilidad, en cuyo caso, $\lambda' = \lambda$; o (2) modelo λ' es mas probable que el modelo λ, es decir, $P(O \mid \lambda') > P(O \mid \lambda)$ y se ha conseguido un nuevo modelo λ' a partir del cual la secuencia de observación es mas probable.

1.6 SECUENCIAS DE OBSERVACIÓN CONTINUAS Y SECUENCIAS DE OBSERVACIÓN MULTIPLES

Hasta ahora se han considerado secuencias de observación caracterizadas por símbolos discretos que pertenecen a un alfabeto finito y que usan probabilidades discretas en cada estado del modelo, no obstante, en algunos problemas las secuencias de observación son señales continuas y por lo tanto es conveniente usar HMMs con densidades de observación continuas y funciones de densidad de probabilidades que aseguren la reestimación consistente de los parámetros del modelo. Por otra parte, se ha revisado el entrenamiento de los parámetros del modelo con una sóla secuencia de observación, pero en la práctica existen muchas aplicaciones, tales como reconocimiento de la voz y alineamiento de secuencias biológicas en las que se debe trabajar con múltiples secuencias de observación para hacer mas confiable el modelo, esto es,

$$\boldsymbol{O} = [\boldsymbol{O}^{(1)}, \boldsymbol{O}^{(2)}, ..., \boldsymbol{O}^{(k)}]$$

donde

$\boldsymbol{O}^{(k)} = (\boldsymbol{O}_1^{(k)} \boldsymbol{O}_2^{(k)} ... \boldsymbol{O}_T^{(k)})$ es la k-ésima secuencia de observación