Bases Formales de la Computación: Sesión 4. Modelos Ocultos de Markov

Prof. Gloria Inés Alvarez V.

Departamento de Ciencias e Ingeniería de la Computación
Pontificia Universidad Javeriana Cali

Periodo 2008-2

Contenido

Evaluación de la segunda parte del curso

- Cada persona debe elegir un artículo científico de las bases de datos IEEE Explorer o Science Direct the Elsevier, que reporte la aplicación de Modelos Ocultos de Markov a un tema específico. Deben ponerse de acuerdo para que cada persona presente una aplicación diferente.
- Deben preparar una presentación de 15 minutos donde expliquen el mencionado artículo.
- El dia 31 de octubre tendremos una evaluación escrita de una hora y las presentaciones de los artículos.

Problema 3

El Problema 3 consiste en estimar los parámetros del modelo de manera que se maximice la probabilidad de una secuencia de observaciones dado el modelo. es decir: ajustar M para maximizar $P(O|M)$

Problema de Estimación del modelo

- Este es sin duda el más difícil de resolver de los tres problemas
- Dada una secuencia de observaciones de entrenamiento, no hay una manera óptima de estimar los parámetros del modelo: A, B, Π
- Se puede elegir $M = (A, B, \Pi)$ tal que $P(O|M)$ se maximiza localmente.

Definición de $\gamma_t(i)$

Definición

La variable $\gamma_t(i)$ representa la probabilidad de estar en el estado i en el tiempo t dadas la secuencia de observaciones O y el modelo M.

$$\gamma_t(i) = P(X_t = s_i | O, M)$$

$$\begin{aligned}
\gamma_t(i) &= P(X_t = s_i | O, M) \\
&= \frac{\alpha_t(i)\beta_t(i)}{P(O|M)} \\
&= \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}
\end{aligned}$$

Procedimiento de Reestimación de Parámetros

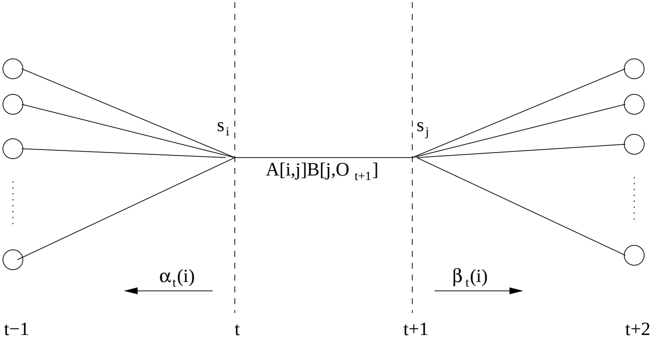
- Es un proceso iterativo que aplica la estrategia EM (expectation-maximization)
- La estrategia se conoce como el algoritmo de Baum-Welch

Definición

Sea $\xi_t(i,j)$ la probabilidad de estar en el estado s_i en el tiempo t y en el estado s_j en el tiempo $t+1$ dada una secuencia de observaciones O y el modelo M. Es decir:

$$\xi_t(i,j) = P(X_t = s_i, X_{t+1} = s_j | O, M)$$

Explicación Intuitiva



Definición precisa de $\xi_t(i,j)$

Definición

$$\begin{aligned}\xi_t(i,j) &= \frac{\alpha_t(i)A[i,j]B[j,O_{t+1}]\beta_{t+1}(j)}{P(O|M)} \\ &= \frac{\alpha_t(i)A[i,j]B[j,O_{t+1}]\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)A[i,j]B[j,O_{t+1}]\beta_{t+1}(j)}\end{aligned}$$

Se puede calcular $\gamma_t(i)$ a partir de $\xi_t(i,j)$:

Definición

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$

Valores esperados de paso por un estado y por una transición

Definición

El valor esperado (en el tiempo) del número de veces que se visita el estado s_i es: $\sum_{t_1}^{T-1} \gamma_t(i)$

Definición

El valor esperado (en el tiempo) del número de veces que se transita desde el estado s_i hasta el estado s_j es: $\sum_{t_1}^{T-1} \xi_t(i,j)$

Reestimación de Valores

Utilizando las variables previamente definidas y el concepto de conteo de ocurrencias de eventos, se propone una forma de ajustar los valores de los parámetros del modelo:

- $\bar{\pi}[i]$ es el número de veces en el estado s_i en el tiempo $t = 1$
- $\bar{A}[i, j]$ número esperado de transiciones desde s_i hasta s_j / número esperado de transiciones desde s_i
- $\bar{B}[j, k]$ es el número esperado de veces que en el estado s_j se observa el evento v_k / el número esperado de veces en el estado s_j

Reestimación de Valores

Definición

$$
\begin{aligned}
\bar{\Pi}[i] &= \gamma_1(i) \\
\bar{A}[i,j] &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\
\bar{B}[j,k] &= \frac{\sum_{t=1}^{T-1} \gamma_t(j), O_t = v_k}{\sum_{t=1}^{T-1} \gamma_t(j)}
\end{aligned}
$$

Reestimación de Valores

- Baum y sus colegas demostraron que si se parte de un modelo $M = (A, B, \Pi)$, se calculan las reestimaciones anteriores y se obtiene un nuevo modelo $\bar{M} = (\bar{A}, \bar{B}, \bar{\Pi})$. Pueden ocurrir dos cosas:
  - $M = \bar{M}$
  - $\bar{M}$ es más probable que el modelo M, en el sentido que $P(O|\bar{M}) > P(O|M)$
- El proceso de reestimación opera iterativamente hasta un punto
- Cuando se detiene se ha calculado el estimado de máxima verosimilitud del modelo oculto de Markov

Algoritmo BaumWelch$(O^1, O^2, \ldots)$

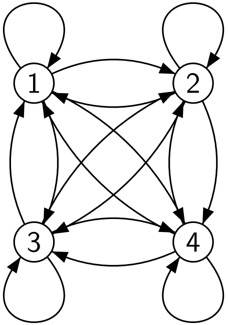
Select arbitrary model parameters $M' = (A', B', \Pi')$
score $= \sum_d P(O^d|M')$
repeat
{
    M = M'
    for each sequence O^d do
    {
        Calculate $\alpha_t(i)$
        Calculate $\beta_t(i)$
        Calculate *newA*, contribution of O^d to A
        Calculate *newB*, contribution of O^d to B
    }
    $A[i,j] = \frac{newA[i,j]}{\sum_l A[i,l]}$
    $B[j,k] = \frac{newB[j,k]}{\sum_l B[j,l]}$
    $score = \sum_d P(O^d|A, B)$
} until (the change in score is less than some predefined threshold)

Variaciones sobre la Estructura del Modelo

- Modelo ergódico
- Modelo de izquierda a derecha

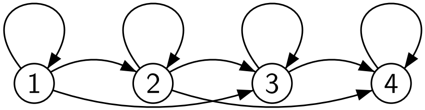
Modelo Ergódico

Cada estado se comunica con cualquier otro en un número finito de pasos. La matriz de probabilidades es de la forma:



$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

Modelo Izquierda a derecha



$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{pmatrix}$$

Modelo Izquierda a derecha

- El modelo empieza siempre por el primer estado ($\Pi[1] = 1$ y el resto valen cero)
- Todas las transiciones apuntan a un estado que está más adelante en la secuencia ($A[i,j] = 0$ si $i > j$)

Notar que las modificaciones en la arquitectura del modelo no tienen impacto sobre el proceso de reestimación, pues si una probabilidad está en cero desde el principio seguirá siendo cero luego de la reestimación.

Variaciones sobre la Función de probabilidad de Emisión

- Función de densidad de probabilidad discreta
- Función de densidad de probabilidad continua

Densidades de Observación Continuas

- Si la naturaleza de las emisiones es continua, forzar su discretización causa degradación en la información que representa
- Se deben establecer algunas restricciones sobre la función continua para garantizar que la reestimación sea consistente
- La representación más general de una función de distribución de probabilidad continua que se puede reestimar es una mixtura finita de la forma:

$$B[j, O] = \sum_{m=1}^{M} C_{jm} \mathfrak{N}[O, \mu_{jm}, U_{jm}]$$

- En el caso continuo, se deben reestimar los parámetros de la función que comúnmente es una Gaussiana

Densidades de Observación Continuas

Definición

$$
\begin{aligned}
\bar{C}_{jk} &= \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_t(j,k)} \\
\bar{\mu}_{jk} &= \frac{\sum_{t=1}^{T} \gamma_t(j,k) O_t}{\sum_{t=1}^{T} \gamma_t(j,k)} \\
\bar{U}_{jk} &= \frac{\sum_{t=1}^{T} \gamma_t(j,k)(O_t - \mu_{jk})(O_t - \mu_{jk})'}{\sum_{t=1}^{T} \gamma_t(j,k)}
\end{aligned}
$$

Variaciones sobre el Lugar de Emisión

- Emisión en el estado
- Emisión en la transición

Otras Variaciones

- Uso de estados nulos
- Tiempo de permanencia en un estado

Campos de Aplicación de los HMM

- Para detectar la familia estructural a la que pertenece una proteina desconocida
- En reconocimiento de texto manuscrito
- En reconocimiento automático de habla
- En detección y filtrado de spam
- En recuperación de información no estructurada
- En etiquetamiento sintáctico
- En reconocimiento de imágenes, aplicado a reconocimiento de rostros, gestos u objetos
- En robótica