Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas.
Departamento Ingeniería Eléctrica.

Laboratorio de Procesamiento y Transmisión de Voz

Capítulo 5.
Reconocimiento de Voz.

5.1 Introducción

- Es el proceso de convertir por medio de hardware (computadora o hardware dedicado) una señal acústica a una secuencia palabras representadas en forma de texto.
- Las palabras reconocidas pueden servir como entrada para otro sistema que emplee estos datos para realizar una acción en específico.

5.1 Introducción

- En la actualidad cada vez más empresas se enfocan en el desarrollo de tecnología que facilite y la comunicación entre las computadoras y los usuarios. El reconocimiento de voz es un área con un gran futuro y con un auge en las organizaciones, con los usuarios finales y en las casas.
- Los investigadores han centrado sus esfuerzos con la finalidad de obtener una máquina que sea capaz de reconocer voz, de manera continua, espontánea, independiente del locutor y sin restricciones de vocabulario. Por esta razón esta área de investigación posee cinco década de trabajos y avances.

5.1 Introducción

- El reconocimiento de voz en los últimos años se basa principalmente en tres técnicas:

1. DTW("Dynamic Time Warping").
2. HMM("Hidden Markov Models").
3. Redes Neuronales.

5.2 DTW(Alineamiento temporal)

- Dynamic Time Warping DTW
- Método el cual tiene en cuenta la variación en la escala del tiempo de dos palabras a comparar.
- El problema que se presenta cuando se pronuncia una palabra es que esta no siempre se realiza a la misma velocidad, lo que produce importantes distorsiones temporales.

5.2 DTW(Alineamiento temporal)

- Estas distorsiones afectan no sola a la palabra considerada sino también a sus componentes acústicos.
- Las variaciones temporales no son generalmente proporcionales a la velocidad de locución y podrán variar de locutor a locutor. Es por esto que se hace necesario un procedimiento que permita comparar dos palabras, sin considerar las distorsiones temporales.

5.2 DTW(Alineamiento temporal)

- Los métodos que se usan para realizar lo expuesto se basan en algoritmos de programación dinámica.

5.2 DTW(Alineamiento temporal)

- Supongamos que tenemos dos secuencias R y T vectores de parámetros observados, cada secuencia corresponde a una palabra:

$$X = X_1, X_2, X_3, \ldots\ldots, X_{Tx}$$
$$Y = Y_1, Y_2, Y_3, \ldots\ldots\ldots, Y_{Ty}$$

Donde X_i y Y_j corresponden a un vector de parámetros (coeficientes LPC, Mel-ceptral, etc) calculados dentro de un cuadro.

5.2 DTW(Alineamiento temporal)

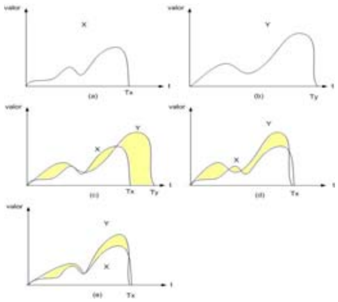
Comparación de patrones.

Patrones acústicos X e Y (a) y (b).

Comparación sin alineamiento (c).

Comparación con alineamiento (d).

Alineamiento no lineal (e).



5.2 DTW(Alineamiento temporal)

Comparación sin alineamiento.

Comparación de dos secuencias calculando la distancia entre cuadros .

$$\begin{matrix} d(X_i, Y_i) & 1 \le i \le Tx \\ 0 & Ty \le i \le Tx \end{matrix}$$

Comparación con alineamiento lineal o compresión lineal.

Para este caso se tiene que hacer Tx=Ty y luego se calcula la distancia D(X,Y) como en el caso anterior.

5.2 DTW(Alineamiento temporal)

Comparación con alineamiento no lineal.

Comprime y dilata de manera no lineal las secuencias X e Y. Comparando ambas formas en el mismo eje temporal resultante.

5.2 DTW(Alineamiento temporal)

- Consideremos una comparación entre dos patrones acústicos X e Y con alineamiento temporal en un plano de dos dimensiones.
- Donde cada secuencia X e Y se localiza en dos ejes ortogonales al plano. (i índice de secuencia X y j son los índices Y, luego c(k)=(i(k),j(k)).

$$F = c(1), c(2), \ldots\ldots, c(K)$$

Corresponde al camino de la función de alineamiento temporal.

5.2 DTW(Alineamiento temporal)

- La distancia temporal es entonces la suma ponderada de todas las distancias locales entre cuadros:

$$d(c(k)) = d(X_{i(k)}, Y_{j(k)})$$

Luego el camino de alineamiento ponderado se expresa como:

$$D(X,Y,F) = \frac{\sum_{k=1}^{K} d(c(k)) \bullet w(k)}{\sum_{k=1}^{K} w(k)}$$

5.2 DTW(Alineamiento temporal)

- (REVISAR) Formalizando el problema a resolver, este consiste en encontrar el camino F que minimice la función de distancia total, dadas las restricciones de coincidencia en dos puntos terminales, continuidades y monotonicidades, en el camino F que se encuentra dentro de la ventana de búsqueda definida por 2r

5.2 DTW(Alineamiento temporal)

$$D(X,Y) = \min_F \frac{\sum_{k=1}^{K} d(c(k)) \bullet w(k)}{\sum_{k=1}^{K} w(k)}$$

Dadas las siguientes restricciones:

1.Coincidencia entre dos puntos terminales:

No punto inicial: $i(1) = j(1) = 1$

No punto final: $i(K) = Tx \qquad j(k) = Ty$

5.2 DTW(Alineamiento temporal)

2. Coincidencias y Monotonicidades:

Para garantizar la monotonicidad entre pares de cuadros consecutivos c(k) y c(k-1), establecemos que:

$$0 \le j(k) - j(k-1) \qquad 0 \le i(k) - i(k-1)$$

Y para las continuidades establecemos que:

$$j(k) - j(k-1) \le 1 \qquad i(k) - i(k-1) \le 1$$

5.2 DTW(Alineamiento temporal)

Luego c(k-1) se puede expresar como:

$$c(k-1)=\begin{cases}(i(k),\ j(k)-1)\\(i(k)-1,\ j(k)-1)\\(i(k)-1,\ j(k))\end{cases}$$

La relación de arriba puede ser cambiada de restricción local que establece que dado un punto c(k) en un camino de alimiento solo existen tres posibles c(k-1) dados por las restricciones de monotonicidades y continuidad.

5.2 DTW(Alineamiento temporal)

3.Ventana de búsqueda (Janela de busca).

Con el fin de evitar un camino de alineamiento F poco razonable y de limitar la carga computacional se escoje un camino F que minimice D(X,Y) se establece una región de onda permitida en búsqueda de F óptimo. Esta región puede ser limitada por dos lineas paralelas espaciadas por 2r.

5.2 DTW(Alineamiento temporal)

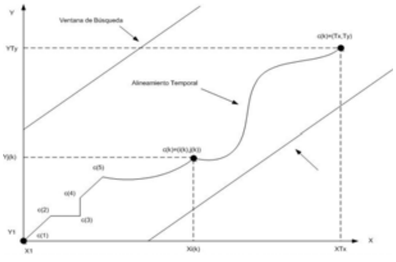


Figura. Alineamiento no lineal entre palabra en DTW. Cada palabra representada por una secuencia de vectores de parámetros.

5.2 DTW(Alineamiento temporal)

- (REVISAR)El denominador del factor de normalización da la distancia total, como veremos a continuación en la relación con la compresión de X e Y, da como tentativa a medir la diferencia entre palabras independiente de sus duraciones. Esto es coherente con la idea de eliminar las diferencias temporales relacionadas con los periodos estacionarios de la voz. Podemos definir w(k) de diferentes maneras:

1. Simétrica

$$w(k)=(i(k)-i(k-1))+(j(k)-j(k-1))$$

$$\sum_{k=1}^{K} w(k)=Tx+Ty$$

5.2 DTW(Alineamiento temporal)

Introduciendo la definición simétrica:

$$D(X,Y) = \min_F \frac{\sum_{k=1}^{K} d(c(k)) \bullet w(k)}{Tx+Ty} \qquad \textbf{(*)}$$

2. Asimétrica:

$$w(k) = i(k) - i(k-1)$$

$$\sum_{k=1}^{K} w(k) = Tx$$

5.2 DTW(Alineamiento temporal)

- Una solución del problema de minimizar la función (*) puede ser visto en primera instancia como un proceso de decisión en forma de árbol con K periodos de entrenamiento.
- Se utilizan herramientas de programación dinámica y es posible tratar el problema como una secuencia de K procesos de decisión con un proceso de entrenamiento.
- Gracias a lo anterior se puede reducir el tiempo de procesamiento para encontrar el camino óptimo.

5.2 DTW(Alineamiento temporal)

- Sea G(c(k)) una mínima distancia de D(X,Y) con denominador Tx+Ty. Como G(c(k)) representa una mínima distancia acumulada entre los cuadros k=1 a k=K, tenemos que:

$$G(c(K)) = G(Tx,Ty) = \min_{c(1),\ldots,c(K-1)} \sum_{k=1}^{K} d(c(k)) \bullet w(K)$$

Donde c(K) es fijo. La expresión puede ser expandida nuevamente:

$$G(c(K)) = G(Tx,Ty) = \min_{c(1),\ldots,c(K-1)} \left[\sum_{k}^{K-1} d(c(k) \bullet w(k) + d(c(K)) \bullet w(K) \right] =$$

5.2 DTW(Alineamiento temporal)

$$= \min_{c(K-1)} \left[\min_{c(1),\ldots,c(K-2)} \left[\sum_{k=1}^{K-1} d(c(k) \bullet w(k) \right] + d(c(K)) \bullet w(K) \right]$$

El primer termino dentro del paréntesis externo puede ser sustituido por G(c(K-1)). Luego la expresión puede ser reescrita como:

$$G(c(K)) = \min_{c(K-1)} \left[G(c(K-1) + d(c(K)) \bullet w(K) \right]$$

Generalizando, sustituido K por $k\,(1 \le k \le K)$

tenemos:

$$G(c(k)) = \min_{c(k-1)} \left[G(c(k-1) + d(c(k)) \bullet w(k) \right]$$

5.2 DTW(Alineamiento temporal)

Las expresiones anteriores indican que la secuencia de K procesos de decisión de un proceso de entrenamiento puede ser sustituido por un proceso de K procesos de entrenamiento. Esta es precisamente la expresión matemática para el principio de optimización, en la cual se basa la programación dinámica:

"Un criterio optimo tiene una propiedad de que cualquiera que sea un estado inicial de una decisión inicial, las decisiones restantes deben construir una estrategia óptima como relación de un estado resultante de la decisión anterior"

5.2 DTW(Alineamiento temporal)

$$G(i,j) = \min \begin{bmatrix} G(i, j-1) + d(X_i, Y_j) \\ G(i-1, j-1) + 2 \bullet d(X_i, Y_j) \\ G(i-1, j) + d(X_i, Y_j) \end{bmatrix}$$

Con:

$$g(1,1) = 2 \bullet d(1,1)$$

Considerando la restricción de la ventana de búsqueda

$$j - r \leq i \leq j + r$$

5.2 DTW(Alineamiento temporal)

- DTW es una método no paramétrico
- Para cada patrón de referencia corresponde una palabra pronunciada.
- Lo anterior conduce a tener varios patrones de la misma palabra. Con la inetción de cubrir todas las variaciones posibles del locutor.
- Lo anterior significa un aumento de procesamiento y uso de memoria.
- Como veremos a continuación DTW se puede definir como un caso particular de HMM.

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

- Esta ha sido la mejor técnica aplicada al problema de reconocimiento de palabras habladas.
- Es una herramienta basada en modelamiento estocástico.
- Se adapta muy bien al fenómeno del habla una vez que este presenta una variabilidad intra e inter-locutor muy grande.
- Son muy flexibles y permiten modelar una palabra entera o un fonema.
- HMM consigue asimilar información en grandes cantidades de datos de entrenamiento en los parámetros de los modelos, permitiendo en operación de reconocimiento, lidiar con muchas locuciones de referencia la mismo tiempo.

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

Procesos de Markov.
Un evento actual determinado es determinado o influenciado por los eventos anteriores.

"*Una de las propiedades de Markov dice que para cualquier secuencia temporal de eventos, la densidad de probabilidades condicional de un evento actual depende solamente de dos j eventos mas recientes*".

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

Procesos de Markov.

Nos restringiremos a estudiar procesos Markovianos de primer orden, pues son estas las que se utilizan en los procesos del reconocimiento de voz.

Veamos ejemplo de un proceso de Markov de primer orden.

Supongamos tres símbolos (a,b,c).

- La probabilidad del símbolo "a" puede ser seguido por cualquiera de los otros tres símbolos "a", "b" y "c" es 1/3 .
- La probabilidad del símbolo "b" ser seguido por otro "b" es de ½ y por los restantes corresponde a ¼.
- La probabilidad del símbolo "c" ser seguido por otro "c" es de ½ y por los restantes corresponde a ¼.

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

Resumiendo lo anterior:

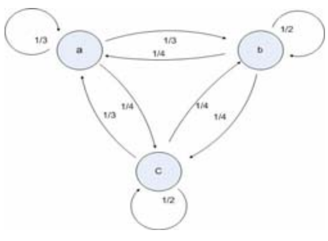
$$\Pr(a/a)=1/3 \quad , \quad \Pr(b/a)=1/3 \quad , \quad \Pr(c/a)=1/3$$
$$\Pr(a/b)=1/4 \quad , \quad \Pr(b/b)=1/2 \quad , \quad \Pr(c/b)=1/4$$
$$\Pr(a/c)=1/4 \quad , \quad \Pr(b/c)=1/4 \quad , \quad \Pr(c/c)=1/2$$

Este tipo de modelo puede ser utilizado para describir estados meteorológicos. Donde "a" corresponde soleado, "b" nublado y "c" chubascos. Ahora escribimos lo anterior en una matriz de transiciones.

$$A=\begin{bmatrix} \Pr(a/a) & \Pr(b/a) & \Pr(c/a) \\ \Pr(a/b) & \Pr(b/b) & \Pr(c/b) \\ \Pr(a/c) & \Pr(b/c) & \Pr(c/c) \end{bmatrix}=\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

- Diagrama de Estado del Modelo de Markov.



5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

Definición de Modelos ocultos de Markov (HMM).

Veamos un ejemplo:

1.Una persona tras una cortina.

2.Hay tres cestas con varias bolas coloridas N=3.

3.Existen bolas de L=4 colores diferentes.

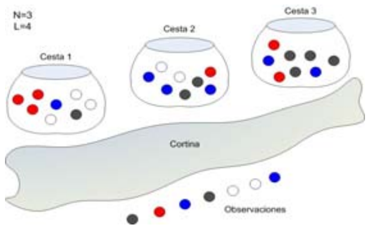
4. Una cesta es escogida aleatoriamente y de ella una bola es escogida aletoriamente.

5.La persona dice en voz alta el color de la bola y la coloca en la cesta nuevamente.

6.Luego se repite el proceso anterior de la elección de la cesta y la bola.

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

HMM es una técnica que tiene como objetivo modelar u establecer las probabilidades de una secuencia de cestas (información oculta) dada la secuencia de bolas de colores (Información disponible)



5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

- Para formalizar, vamos a usar la siguiente notación:

D=largo de la secuencia de observación O_1, O_2,…….,O_D. (número de bolas coloridas observadas).

N=número de estados en el modelo. (número de cestas).

O_t=vector de observación resultante de medidas físicas en cuadro t.

L=número de símbolos observables.(número de cestas o canastas).

S={s} un conjunto de estados. Por simplificación el estado i en instante t puede ser denominado $s_t=1$.

$V=\{v_1,v_2,\ldots,v_L\}$ conjunto finito de posibles símbolos observables. O_t esta relacionado con un elemento del conjunto V por medio de la cuantización vectorial.

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

$A=\{a_{ij} \mid a_{ij}=Pr(s_{t+1}=j|s_t=i)\}$ Distribución de probabilidades de transición de estados.

$B=\{b_j(O_t) \mid b_j(O_t)=Pr(O_t|s_t=j)\}$
Para cada estado hay una correspondiente función de probabilidades de salida (función discreta o continua).

$\Pi=\{\Pi_i \mid \Pi_i= Pr(s_1=i)\}$ Función de distribución de probabilidad para estado inicial.

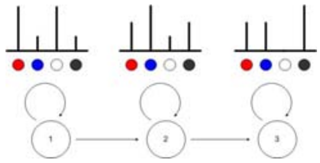
5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

El experimento de las bolas coloridas puede ser modelado con HMM utilizando las siguientes definiciones:

- Para cada estado i corresponde una cesta.
- Como las probabilidades de salida $b_i(O)$ son definidas por las bolas de colores observadas en cada estado. Esto es una distribución de probabilidad en cada urna.
- El símbolo observado v_k es un color de la bola seleccionada de una de las urnas.
- A cada una de las cesta se le da la prob de ser la 1a (distribución de estado inicial). Se define la distribución de probabilidad de transición entre estados.

5.3 HMM (Hidden Markov Models o Modelos Ocultos de Markov)

- Este modelo puede ser considerado como un generador de una secuencia de experimentos, una vez que la cadena de Markov genera una secuencia de estados, a partir de un estado inicial, la función de probabilidad de salida se encarga de generar un bola de color para cada estado. La secuencia de observaciones de bolas de colores da una idea del estado y de los parámetros de modelo.



5.3.1 Algoritmo de Viterbi HMM (Modelos Ocultos de Markov)

- La información oculta de HMM, una secuencia de estados (e.g. secuencia de cestas), no puede ser recuperada , pero puede ser estimada por el criterio de probabilidades máximas. El algoritmo de Viterbi encuentra la secuencia de estados que tiene la mayor probabilidad de ser generada para una secuencia de estados observada. Maximiza Pr(O,S/λ).
- El algoritmo de Viterbi es muy semejante al DTW, en el sentido que se intenta encontrar una secuencia de alineamiento que minimice la distancia global al compar dos secuencia de vectores observados.

5.3.1 Algoritmo de Viterbi HMM (Modelos Ocultos de Markov)

- **Paso 1.** Inicialización. Para todos los estados i.

$$\delta_1(i) = \pi_i b_i(O_1)$$
$$\Psi_1(i) = 0$$

- **Paso2.** Recursión. En el instante t=2 en T. Para todos los estados j:

$$\delta_t(j) = Max\left[\delta_{t-1}(i) \cdot a_{ij}\right] \cdot b_j(O_t)$$
$$\Psi_t(j) = \arg\max\left[\delta_{t-1}(i) \cdot a_{ij}\right]$$

5.3.1 Algoritmo de Viterbi HMM (Modelos Ocultos de Markov)

- **Paso 3.** Finalización. El asterisco* indica los resultados óptimos.

$$P^* = \underset{s \in S_F}{Max} \left[\delta_T(s) \right]$$

$$s_T^* = \arg \underset{s \in S_F}{\max} \left[\delta_T(s) \right]$$

- **Paso4.** Determinación del camino óptimo, para atrás. De t=T-1 a t=1.

$$s_t^* = \psi_{t+1}\left(s_{t+1}^*\right)$$

La expresión $\arg\max\left[\delta_{t-1}(i) \cdot a_{ij}\right]$ indica el estado i que maximiza $\delta_{t-1}(i) \cdot a_{ij}$

5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

- Una manera más directa de calcular las probabilidades de un secuencia de observaciones es enumerando toda posible secuencia de estados de largo T (o numero de observaciones). Para una secuencia de estados:

$$\Pr(O/S,\lambda) = b_{s_1}(O_1) \cdot b_{s_{21}}(O_2) \ldots\ldots\ldots b_{s_T}(O_T)$$

Por otro lado la secuencia de probabilidades esta dada **:**

$$\Pr(S/\lambda) = \pi_{s_1} \cdot a_{s_1 s_2} \cdot a_{s_2 s_3} \cdot \ldots\ldots \cdot a_{s_{T-1} s_T}$$

Por simplicidad de notación llamaremos $\pi_{s_1} = a_{s_0 s_1}$

5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

- La probabilidad conjunta:

$$Pr(O,S/\lambda) = Pr(O/S,\lambda) Pr(S/\lambda)$$

- La probabilidad es la sumatoria de ecuaciones sobre todas las posibles secuencia de estados

$$Pr(O/\lambda) = \sum_{Todas\ secuencia\ S} Pr(O/S,\lambda) Pr(S/\lambda)$$

$$Pr(O/\lambda) = \sum_{Toda\ secuencia\ S} \prod_{t=1}^{T} a_{s_{t-1}s_t} \cdot b_{s_t}(O_t)$$

5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

- Definamos la siguiente variable:

$$\alpha_t(i) = \Pr(O_1, O_2, \ldots\ldots, O_t, s_t = i/\lambda)$$

- Esta es la probabilidad de observar una parte de la secuencia de observaciones de O_1 a O_T.. Las probabilidades $\alpha_t(i)$ pueden ser calculadas por inducción por medio del algoritmo de Forward, mostrado a continuación.

5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

Algoritmo de Forward.

Paso1:

$$\alpha_1(i) = \pi_i b_i(O_1)$$

Para todos los estados i (si i ∈ S_I), entonces: $\pi_i = \frac{1}{N_I}$,caso contrario $\pi_i = 0$

Paso2:

Calculo de α() a través de la siguiente ecuación inductiva:

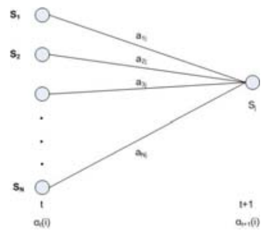
$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i) \cdot a_{ij}\right] \cdot b_j(O_t)$$

Paso3:

La probabilidad final esta dada por: $\Pr(O/\lambda) = \sum_{i \in S_F} \alpha_T(i)$

5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

- Ilustración del algoritmo de Forward



5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

- De manera semejante vamos a definir la la siguiente variable de Backward:

$$\beta_t(i) = \Pr(O_{t+1}, O_{t+2}, \ldots\ldots, O_T / s_t = i, \lambda)$$

- Esto es observar una secuencia de Ot+1 a O_T dado que en instante t el estado era i , y dado el modelo λ. Esta variable $\alpha_t(i)$ puede ser determinada inductivamente para el algoritmo de Backward para todos los estados e instantes de manera análoga al algoritmo de Forward.

5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

Algoritmo de Backward.

Paso1:

$$\beta_T(i) = 1 / N_F$$

Para todos los estados i ∈ S_F, caso contrario $\beta_T(i) = 0$

Paso2:

Calculo de β() a través de la siguiente ecuación inductiva:

$$\beta_t(j) = \left[\sum_{i=1}^{N} a_{ji} \cdot b_j(O_{t+1}) \cdot \beta_{t+1}(i)\right]$$

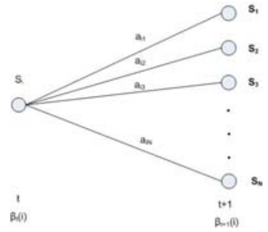
Paso3:

La probabilidad final esta dada por:

$$\Pr(O/\lambda) = \sum_{i \in S_I} \pi_i \cdot b_i(O_1) \cdot \beta_1(i)$$

5.3.2 Algoritmo de Forward y Backward HMM (Modelos Ocultos de Markov)

- Ilustración algoritmo de Backward



5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- El problema más difícil en HMM tiene relación a cómo ajustar los parametros de modelo (A,B,∏) de manera de maximizar las probabilidades P(O/λ).
- No existe una solución analítica para el problema, pero existe un algoritmo de iteración conocido como Baum- Welch.
- El algoritmo de Baum- Welch esta basado en la técnica de maximización de esperanza de máxima verosimilitud.

5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

Considere una HMM cualquiera con parámetros λ= (A,B,∏) todos dintintos de cero. La probabilidad de transición para un estado i a un estado j después de una secuencia de observación dada es:

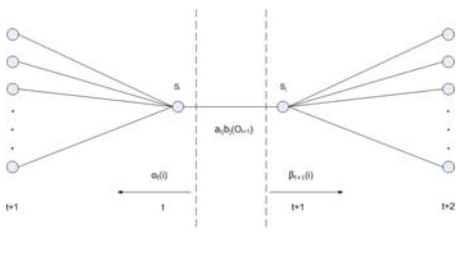
$$\gamma_t(i,j) = \Pr(s_t = i, s_{t+1} = j / O, \lambda)$$

La expresión puede ser colocada en función de las variables de Fordware y Backward como:

$$\gamma_t(i,j) = \frac{\alpha_t(i) \cdot a_{ij} b_j(O_{t+1}) \cdot \beta_{t+1}(j)}{\Pr(O/\lambda)} = \frac{\alpha_t(i) \cdot a_{ij} b_j(O_{t+1}) \cdot \beta_{t+1}(j)}{\sum_{k \in S_F} \alpha_T}$$

5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- Ilustración del cálculo de las variables gamma.



5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- De manera semejante definimos las probabilidades a posteriori de estas en estado i en instante t, dad la secuencia de observaciones de el modelo:

$$\gamma_t(i) = \Pr(s_t = i / O, \lambda) = \frac{\alpha_t(i) \cdot \beta_t(i)}{\sum_{k \in S_F} \alpha_T}$$

Donde :

$$\gamma_t(i) = \sum_{j=1}^{N} \gamma_t(i, j) \qquad t < T$$

(REVISAR)Seja entao lamda parametros de un HMM y Lamda gorro parametros de una misma HMM apos uma reestimacao.

5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- Se define :

$$Q(\lambda, \bar{\lambda}) = \frac{1}{\Pr(O/\lambda)} \cdot \sum_{todos/estados} \Pr(O, S/\lambda) \cdot \log[\Pr(O, S/\bar{\lambda})]$$

- La función anterior es función de lamda gorro, pues lamda son los parámetros actuales de HMM antes de la siguiente re-estimación de parámetros.
- Se puede observar la concavidad de la función logaritmo:

$$\log\left[\frac{\Pr(O/\bar{\lambda})}{\Pr(O/\lambda)}\right] \geq Q(\lambda, \bar{\lambda}) - Q(\lambda, \lambda)$$

5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- A partir de la expresión anterior se observa un nuevo modelo de re estimación de lamda gorro maximiza la probabilidad Pr(), maximiza tambien la expresión Q().
- Sea $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$, aplicadas a una secuencia de estados S=s_1,s_2,…..,s_T, tenemos:

$$\log[\Pr(O, S/\bar{\lambda})] = \log[\bar{\pi}_1] + \sum_{t=1}^{T-1} \log[\bar{a}_{s_t s_{t+1}}] + \sum_{t=1}^{T} \log[\bar{b}_{s_t}(O_t)]$$

5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- Sustituyendo la expresión anterior en la expresión para Q() y reagrupando los términos correspondientes a las probabilidades de transiciones y las probabilidades de transición de símbolos, llegamos a:

$$Q(\lambda, \bar{\lambda}) = \sum_i \sum_j c_{ij} \cdot \log[\bar{a}_{ij}] + \sum_j \sum_{k=1}^{L} d_{jk} \cdot \log\left[\bar{b}_j(k)\right] + \sum_i e_i \cdot \log[\bar{\pi}_i]$$

Donde:

$$c_{ij} = \sum_{t=1}^{T-1} \gamma_t(i, j) \qquad d_{jk} = \sum_{t \in O_t = v_k} \gamma_t(j) \qquad e_i = \gamma_1(i)$$

5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- (revisar) Q se maximiza si solamente si:

$$\bar{a}_{ij} = \frac{c_{ij}}{\sum_j c_{ij}} = \frac{\sum_{t=1}^{T-1} \gamma_t(i, j)}{\sum_{t=1}^{T-1} \sum_j \gamma_t(i, j)} \qquad (*)$$

$$\bar{b}_j(k) = \frac{d_{jk}}{\sum_k d_{jk}} = \frac{\sum_{t \in O_t = v_k} \gamma_t(j)}{\sum_j \gamma_t(j)} \qquad (**)$$

$$\bar{\pi}_i = \frac{e_i}{\sum_i e_i} = \gamma_1(i) \qquad (***)$$

5.3.2 Algoritmo de Baum-Welch HMM (Modelos Ocultos de Markov)

- Las expresiones (*),(**) y (***) se conocen como las ecuaciones de re-estimación de Baum-Welch, para el caso en que la distribución de probabilidad de salida es discreta.