

Técnicas de Modelamiento Computacional con Potencial Uso en Bioinformática

Hernán Darío Benítez R. y Gloria Inés Alvarez V.

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

Noviembre de 2008

Contenido

- 1 Reconocimiento de Patrones Estadístico
- 2 Modelos Ocultos de Markov
- 3 Gramáticas Estocásticas
- 4 Inferencia Gramatical
- 5 Inteligencia Computacional

Modelos Ocultos de Markov

Definición

Un modelo oculto de Markov es una tupla $M = \{S, \Sigma, A, B, \Pi\}$ donde:

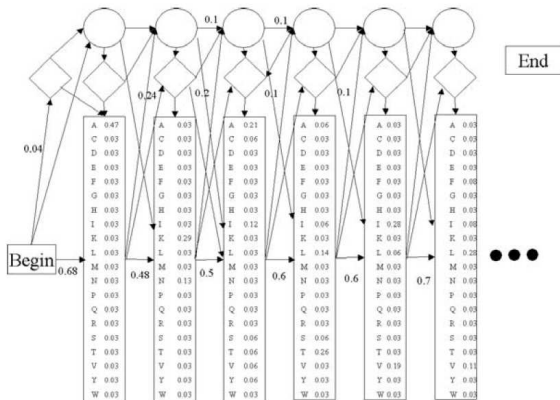
- *S es un conjunto finito de estados.*
- *Σ es un conjunto finito de símbolos.*
- *A es la matriz de probabilidades de transición entre estados de S . $A[i, j]$ es $P(X_{t+1} = S_j | X_t = s_i)$*
- *B es la matriz de probabilidad de emisión de símbolos de Σ . $B[j, k]$ es $B[j, k] = P(O_t = k | X_t = s_j)$*
- *Π es el vector de probabilidades iniciales. $\Pi[i]$ es $P(X_1 = S_i)$*

Se denota una secuencia de estados $X = (X_1, \dots, X_{T+1})$ donde $X_t : S \rightarrow \{1, \dots, N\}$ y una secuencia de observaciones $O = (o_1, \dots, o_T)$ con $o_T \in \Sigma$

Los tres problemas fundamentales en HMM

- 1 Problema de evaluación de la probabilidad (o verosimilitud) de una secuencia de observaciones dado un HMM.
- 2 Problema de determinación de la secuencia más probable de estados.
- 3 Problema de ajuste de los parámetros del modelo para que den mejor cuenta de las señales observadas.

Ejemplo de HMM



Gramáticas Incontextuales Estocásticas, Definición

Definición

Una gramática incontextual estocástica es una tupla

$$G = \{N, T, P, S, Q\}$$

- *N es un conjunto finito de símbolos no terminales*
- *T es un conjunto finito de símbolos terminales*
- *P es un conjunto finito de producciones de la forma $A \rightarrow \alpha$, con $A \in N$ y $\alpha \in \{N \cup T\}^*$*
- *$S \in N$ es el símbolo inicial de la gramática*
- *Q es un conjunto de probabilidades asociadas a las producciones tal que $\forall i, \sum_i P(N^i \rightarrow \zeta^i) = 1$, con $N^i \in N$ y $\zeta^i \in \{N \cup T\}^*$*

Ejemplo de Gramática Incontextual Estocástica

$S \rightarrow NP VP(1,0)$	$NP \rightarrow NP PP(0,4)$
$PP \rightarrow P NP(1,0)$	$NP \rightarrow astronomers(0,1)$
$VP \rightarrow V NP(0,7)$	$NP \rightarrow ears(0,18)$
$VP \rightarrow VP NP(0,3)$	$NP \rightarrow saw(0,04)$
$P \rightarrow with (1,0)$	$NP \rightarrow stars(0,18)$
$V \rightarrow saw (1,0)$	$NP \rightarrow telescopes(0,1)$

Con esta gramática incontextual estocástica es posible construir dos árboles sintácticos a partir de la frase *astronomers saw stars with ears*. La probabilidad del árbol de sintaxis se calcula multiplicando la probabilidad de las reglas que se aplican.

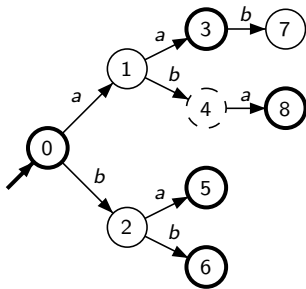
Los Tres Problemas Fundamentales en Gramáticas Incontextuales Estocásticas

- 1 Problema de evaluación de la probabilidad (o verosimilitud) de una secuencia de observaciones w_{1m} de acuerdo a una gramática G
- 2 Problema de determinación del árbol de derivación más probable para la secuencia: $\operatorname{argmax}_t P(t \mid w_{1m}, G)$
- 3 Problema de ajuste de los parámetros del modelo, es decir de la probabilidad de las reglas de la gramática para maximizar la probabilidad de una secuencia: $\operatorname{argmax}_G P(w_{1m} \mid \cdot, G)$.

Inferencia Gramatical

- El problema de la inferencia gramatical se puede plantear como el de aprender una función desconocida a partir de algunos de los valores que produce.
- La entrada del problema está formada por un conjunto no vacío de muestras positivas y un conjunto de muestras negativas, que en la definición general puede ser vacío pero que en este estudio en concreto no lo será.
- La representación resultante de la inferencia puede expresarse de diversas maneras, en este estudio dicha representación será un autómata finito.
- Se asume el modelo de aprendizaje de Gold.

Ejemplo de ejecución del algoritmo RPNI



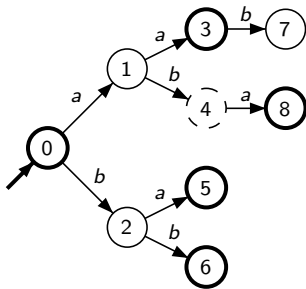
$$S = \{0\}$$

$$T = \{1, 2\}$$

Compara 0, 1: \neq

Compara 0, 2: \neq

Ejemplo de ejecución del algoritmo RPNI



$$S = \{0, 1\}$$

$$T = \{2, 3, 4\}$$

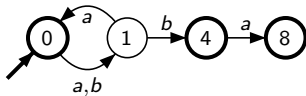
Compara 1, 2 da: =

Compara 0, 3 da: =

Compara 0, 4 da: \neq

Compara 1, 4 da: \neq

Ejemplo de ejecución del algoritmo RPNI



$$S = \{0, 1\}$$

$$T = \{2, 3, 4\}$$

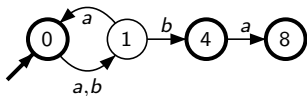
Compara 1, 2 da: =

Compara 0, 3 da: =

Compara 0, 4 da: \neq

Compara 1, 4 da: \neq

Ejemplo de ejecución del algoritmo RPNI

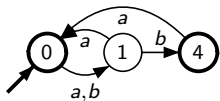


$$S = \{0, 1, 4\}$$

$$T = \{8\}$$

Compara: 0, 8 da =

Ejemplo de ejecución del algoritmo RPNI



$$S = \{0, 1, 4\}$$

$$T = \{8\}$$

Compara: 0, 8 da =

Técnicas Aplicables

- Redes Neuronales
- Algoritmos Genéticos
- Inteligencia de Enjambres
- Programación evolutiva