Bases Formales de la Computación: Sesión 5. Gramáticas Estocásticas

Prof. Gloria Inés Alvarez V.

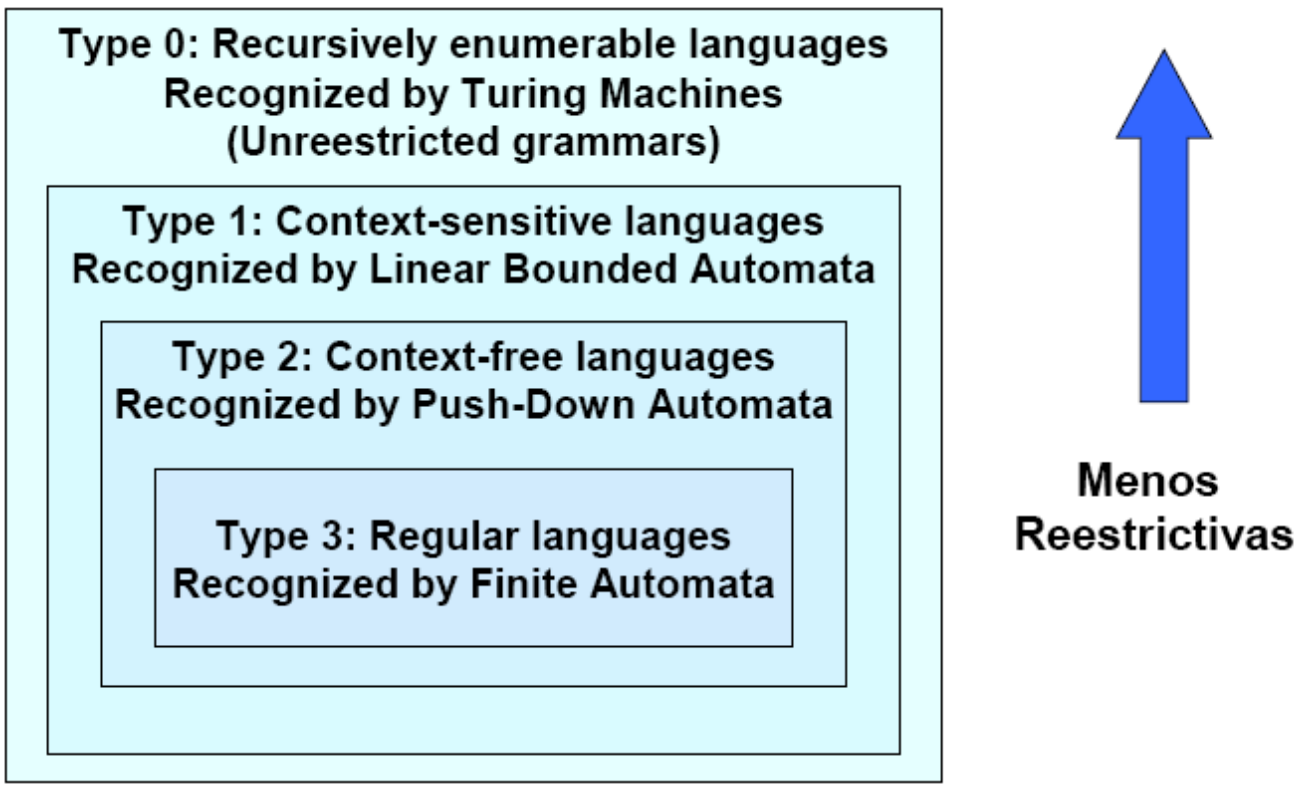
Departamento de Ciencias e Ingeniería de la Computación
Pontificia Universidad Javeriana Cali

Periodo 2008-2

Contenido

Introducción

En la teoría de los lenguajes formales también se proponen modelos que permite representar sistemas. Siguiendo la jerarquía de Chomsky los modelos se clasifican de acuerdo a su poder expresivo.



Lenguajes Regulares

Definición

Un Autómata finito es una tupla $A = \{Q, \Sigma, \delta, q_0, F\}$*, donde:*

- Q *es un conjunto finito de estados*
- Σ *es un alfabeto finito de símbolos*
- δ *es una función de transición* $\delta : Q \times \Sigma \rightarrow 2^Q$
- q_0 *es el estado inicial*
- $F \in Q$ *es un conjunto de estados finales o de aceptación*

Gramáticas Regulares

Definición

Una gramática regular es una tupla $G = \{N, T, P, S\}$ *donde:*

- 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* $\alpha \in T \times N$ *ó* $\alpha \in T$
- $S \in N$ *es el símbolo inicial de la gramática*

Gramáticas Regulares

Por ejemplo, la gramática regular que reconoce el lenguaje $L = \{a^n b^m \mid n, m \geq 0\}$ sobre $\Sigma = \{a, b\}$ está formada por las producciones:

- $S \longrightarrow aA$
- $A \longrightarrow aA$
- $A \longrightarrow bB$
- $B \longrightarrow bB$
- $B \longrightarrow \varepsilon$

Lenguajes Incontextuales

Definición

Una gramática incontextual es una tupla $G = \{N, T, P, S\}$ *donde:*

- 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*

Gramáticas Incontextuales

Por ejemplo, la gramática incontextual que reconoce el lenguaje $L = \{a^n b^n \mid n > 0\}$ sobre $\Sigma = \{a, b\}$ está formada por las producciones:

- $S \longrightarrow aSb$
- $S \longrightarrow ab$

Lenguajes Contextuales

Definición

Una gramática contextual es una tupla $G = \{N, T, P, S\}$ *donde:*

- 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* $A, \alpha \in \{N \cup T\}^*$
- $S \in N$ *es el símbolo inicial de la gramática*

Gramáticas Contextuales

Por ejemplo, la gramática contextual que reconoce el lenguaje $L = \{a^n b^n c^n \mid n > 0\}$ sobre $\Sigma = \{a, b\}$ está formada por las producciones:

- $S \longrightarrow aBSc$
- $S \longrightarrow abc$
- $Ba \longrightarrow aB$
- $Bb \longrightarrow bb$

Gramáticas Incontextuales, Derivación

Definición

Sean $u, v, w \in \{N \cup T\}^*$ *y* $(A \rightarrow w) \in P$ *se dice que* $uAv \Rightarrow uwv$.

Definición

Se dice que u *deriva* $v, (u \Rightarrow^* v)$ *si* $u = v$ *ó si existe* $u_1, u_2, \ldots, u_k$ *con* $k \geq 0$ *tal que* $u \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \cdots \Rightarrow u_k \Rightarrow v$.

Lenguaje de una gramática

Definición

El lenguaje de una gramática G es $\{w \mid w \in \Sigma^*, S \Rightarrow^* w\}$, siendo S el símbolo inicial de G.

Ejemplos:

- $G_3 = (\{S\}, \{a, b\}, \{S \rightarrow aSb \mid SS \mid \varepsilon\}, \{S\})$
- $G_4 = (\{EXP, TERM, FACT\}, \{a, +, *, (,)\}, \{EXP \rightarrow EXP + TERM \mid TERM, TERM \rightarrow TERM * FACT \mid FACT, FACT \rightarrow (EXP) \mid a\}, \{EXP\})$

Ambigüedad

Definición

Una derivación de una cadena w en una gramática incontextual G es una derivación más izquierda si a cada paso se reemplaza el símbolo no terminal que se encuentra más a la izquierda.

Definición

Una cadena w es derivada ambiguamente de una gramática incontextual G si tiene dos o más derivaciones más izquierdas. Una gramática G es ambigua si genera alguna cadena ambiguamente.

Definición

Un lenguaje inherentemente ambiguo es aquel que sólo puede reconocerse con gramáticas ambiguas.

Ejemplo 1 de Gramática ambigua

$EXP \rightarrow EXP + EXP$
$EXP \rightarrow EXP * EXP$
$EXP \rightarrow (EXP)$
$EXP \rightarrow a$
Ver árboles de análisis sintáctico para $a + a * a$

Forma Normal de Chomsky

Definición

Una gramática incontextual está en forma normal de Chomsky si todas sus reglas son de la forma: $A \to BC$ ó $A \to a$ donde A, B, C son no terminales y B, C no son el símbolo inicial de la gramática. Se permite la producción $S \to \varepsilon$ para el símbolo inicial.

Teorema

Todo lenguaje incontextual es generado por una gramática en forma normal de Chomsky.

Demostración

Demostración.

Toda gramática incontextual puede transformarse en una en forma normal de Chomsky que reconoce el mismo lenguaje. La transformación se realiza por etapas:

1. Se adiciona un nuevo símbolo inicial
2. Se consideran las reglas vacías: $A \longrightarrow \varepsilon$ se elimina y por cada vez que aparece A en el lado derecho de una regla se adiciona otra regla igual salvo que la ocurrencia de A no está.
3. Se considera las reglas unitarias: $A \longrightarrow B$ se elimina y toda regla de la forma $B \longrightarrow u$ se duplica cambiando B por A.
4. Se convierten el resto de reglas: $A \longrightarrow u_1, u_2, \ldots, u_k$ con $k \geq 3$ se reemplaza por las reglas:
   $A \longrightarrow u_1 A_1, A_1 \longrightarrow u_2 A_2, \ldots, A_{k-2} \longrightarrow u_{k-1} u_k$. Si u_i es un terminal, se reemplaza por U_i y se adiciona $U_i \longrightarrow u_i$

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 Incontextural 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.

Porqué es útil extender el modelo con probabilidades?

- Esto incrementa el poder expresivo del modelo, pueden entrenarse con muestras positivas únicamente
- Son robustas.
- A medida que una gramática crece, se hace más ambigua y sin las probabilidades es difícil establecer la plausibilidad de las diversas alternativas

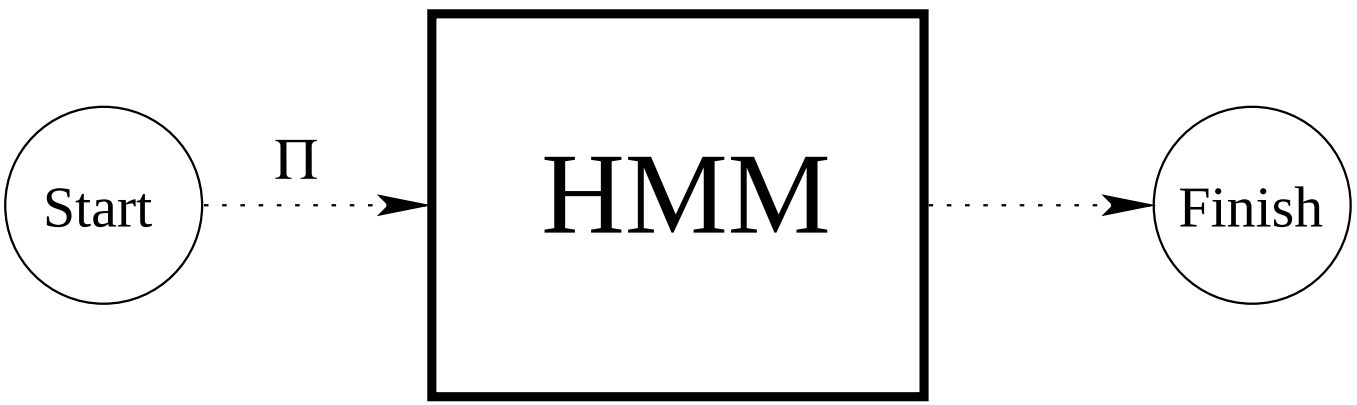
Gramáticas Incontextuales Estocásticas y HMMs

- Las gramáticas sirven tanto para analizar cadenas como para generarlas
- Hay una relación estrecha entre las gramáticas regulares estocásticas y los HMM
- El poder predictivo de las gramáticas tiende a ser más alto que el de los HMM, medido mediante el criterio de entropía. Por ejemplo, la frase *Juan decidió cocinar un* obtiene una probabilidad alta en un HMM, por ser un comienzo válido de frase, pero obtiene una probabilidad baja en una gramática debido a que no es una oración bien formada.

Gramáticas Incontextuales Estocásticas y HMMs

- Una gramática regular estocástica se puede construir a partir de un HMM agregando un estado inicial y un estado final.
- Los dos modelos se diferencian, sin embargo, en la interpretación que se da a la distribución de probabilidad sobre las cadenas del lenguaje. En los HMM la probabilidad se reparte entre las cadenas de la misma longitud, mientras en la gramática se reparte entre todas las cadenas del lenguaje

Construcción de una gramática regular estocástica a partir de un HMM



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: $argmax_t P(t \mid w_{1m}, G)$
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: $argmax_G P(w_{1m} \mid , G)$.

Problema de Evaluación

Definición

Dada una secuencia de observaciones $w_1, \ldots, w_m$ y una gramática incontextual estocástica G, cómo calcular eficientemente $P(w_{1m}|G)$, la probabilidad de la secuencia dado el modelo?

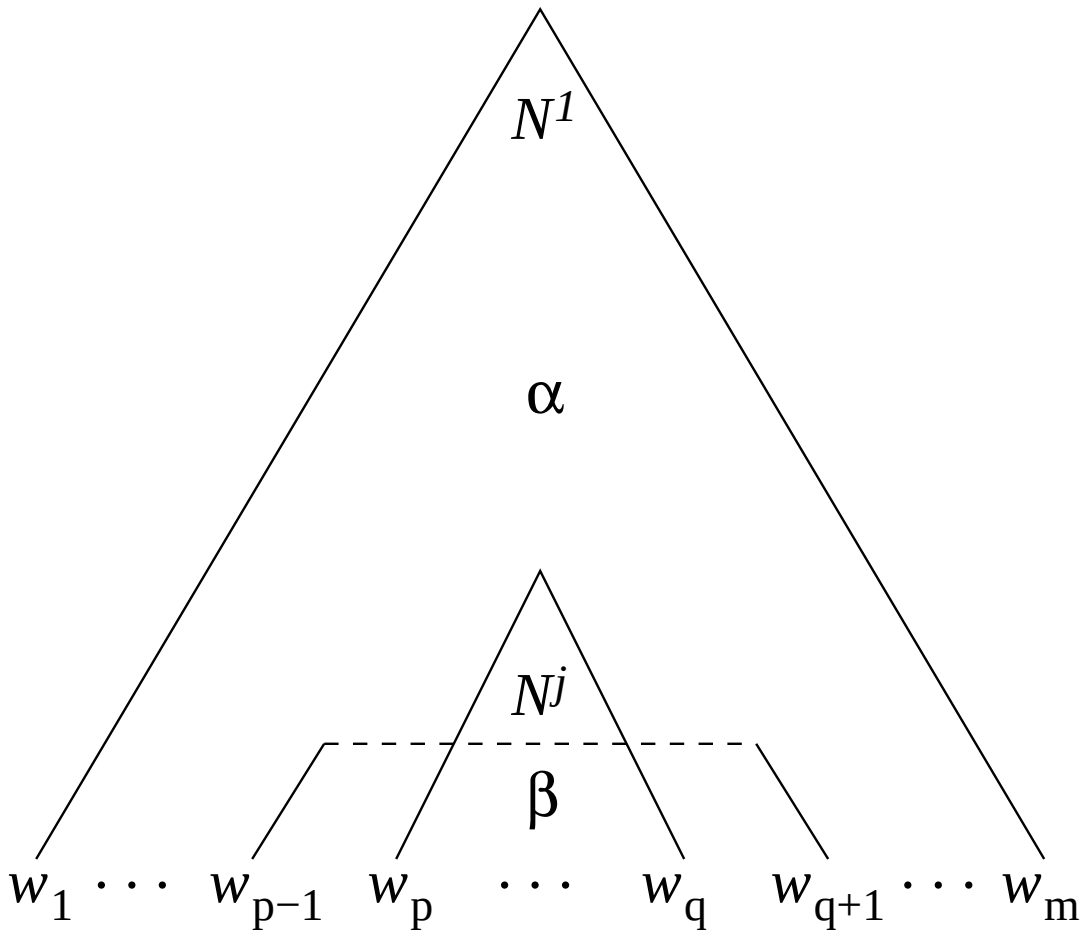
Esta probabilidad se puede calcular adaptando las variables α y β que se usaron en los HMM

Variables Inside y Outside

- Probabilidad outside: $\alpha_j(p,q) = P(w_{1(p-1)}, N^j_{pq}, w_{(q+1)m}|G)$
- Probabilidad inside: $\beta_j(p,q) = P(w_{pq}|N^j_{pq}, G)$

En este estudio se va a suponer que las gramáticas están en forma normal de Chomsky

Variables Inside y Outside



Solución al Problema de Evaluación

- A partir de la variable inside:

$$\begin{aligned} P(w_{1m}|G) &= P(N^1 \Rightarrow^* w_{1m}|G) \\ &= P(w_{1m}|N^1_{1m}, G) \\ &= \beta_1(1,m) \end{aligned}$$

- A partir de la variable outside:

$$\begin{aligned} P(w_{1m}|G) &= \sum_j P(w_{1(k-1)}, w_k, w_{(k+1)m}, N^j_{kk}|G) \\ &= \sum_j P(w_{1(k-1)}, N^j_{kk}, w_{(k+1)m}|G) \times \\ &\quad P(w_k|w_{1(k-1)}, N^j_{kk}, w_{(k+1)m}, G) \\ &= \sum_j \alpha_j(k,k) P(N^j \rightarrow w_k) \end{aligned}$$

Definición de la Variable Inside

- $\beta_j(p,q)$ es la probabilidad total de generar la secuencia $w_p, \ldots, w_q$ dado que se parte del no terminal N^j
- $\beta_j(p,q)$ se puede calcular inductivamente:
  - Caso base:

$$\begin{aligned} \beta_j(k,k) &= P(w_k | N^j_{kk}, G) \\ &= P(N^j \rightarrow w_k | G) \end{aligned}$$

Definición de la Variable Inside

- Caso general: $\beta_j(p,q)$ para $\forall j, 1 \leq p < q \leq m$

$$
\begin{aligned}
\beta_j(p,q) &= P(w_{pq} \mid N^j_{pq}, G) \\
&= \sum_{r,s} \sum_{d=p}^{q-41} P(w_{pd}, N^r_{pd}, w_{(d+1)q}, N^s_{(d+1)q}, | N^j_{pq}, G) \\
&= \sum_{r,s} \sum_{d=p}^{q-1} P(N^r_{pd}, N^s_{(d+1)q}, | N^j_{pq}, G) \times \\
&\quad P(w_{pd} | N^j_{pq}, N^r_{pd}, N^s_{(d+1)q} | G) \times \\
&\quad P(w_{(d+1)q} | N^j_{pq}, N^r_{pd}, N^s_{(d+1)q}, w_{pd}, G)
\end{aligned}
$$

Definición de la Variable Inside

- Caso general: $\beta_j(p,q)$ para $\forall j, 1 \leq p < q \leq m$

$$\begin{aligned}\beta_j(p,q) &= \sum_{r,s}\sum_{d=p}^{q-1} P(N^r_{pd}, N^s_{(d+1)q} | N^j_{pq}, G) P(w_{pd} | N^r_{pd}, G) \\ &\quad P(w_{(d+1)q} | N^s_{(d+1)q}, G) \\ &= \sum_{r,s}\sum_{d=p}^{q-1} P(N^j \rightarrow N^r N^s) \beta_r(p,d) \beta_s(d+1,q)\end{aligned}$$

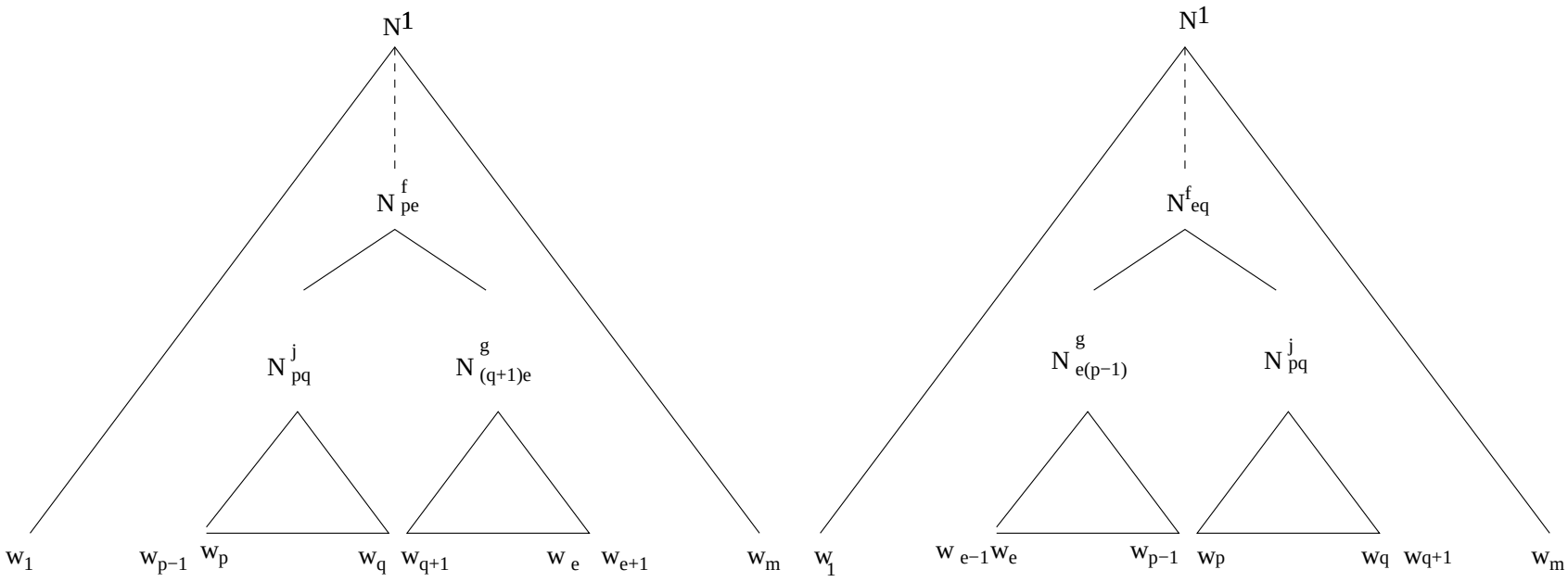
Definición de la Variable Outside

- $\alpha_j(p,q)$ es la probabilidad de empezar en el símbolo inicial y generar el no terminal N^j_{pq} y todas las palabras fuera de w_{pq}.
- $\alpha_j(p,q)$ se puede calcular inductivamente:
  - Caso base:

$$\begin{aligned} \alpha_1(1,m) &= 1 \\ \alpha_j(1,m) &= 0, j \neq 1 \end{aligned}$$

  - Caso general:

$$\alpha_j(p,q) = \left[\sum_{f,g\neq j}\sum_{e=q+1}^{m} \alpha_f(p,e)P(N^f \rightarrow N^j N^g)\beta(q+1,e)\right] + \left[\sum_{f,g}\sum_{e_1}^{p-1} \alpha_f(e,q)P(N^f \rightarrow N^g N^j)\beta_g(e,p-1)\right]$$



$$\alpha_j(p,q) = \left[\sum_{f,g\neq j}\sum_{e=q+1}^{m} \alpha_f(p,e)P(N^f \rightarrow N^j N^g)\beta(q+1,e)\right] + \left[\sum_{f,g}\sum_{e_1}^{p-1} \alpha_f(e,q)P(N^f \rightarrow N^g N^j)\beta_g(e,p-1)\right]$$

Problema 2: Determinación del árbol de derivación más probable

- Se puede adaptar el algoritmo de Viterbi para que encuentre el árbol de derivación más probable para una cadena
- Si puede ajustar el algoritmo inside para que encuentre el elemento de la suma que es máximo y almacene qué regla lo produjo
- Se define $\delta_i(p, q)$ como la más alta probabilidad inside que explica el subárbol N^i_{pq}

Definición de la Variable $\delta_i(p, q)$

- Inicialización

$$\delta_i(p, p) = P(N^i \rightarrow w_p)$$

- Inducción

$$\delta_i(p, q) = \max_{1 \leq j,k \leq n, p \leq r < q} P(N^i \rightarrow N^j N^k)\delta_j(p, r)\delta_k(r + 1, q)$$

$$\psi_i(p, q) = \arg \max_{(j,k,r)} P(N^i \rightarrow N^j N^k)\delta_j(p, r)\delta_k(r + 1, q)$$

- Finalización

$$P(\hat{t}) = \delta_1(1, m)$$

Estimación de las Probabilidades de las Reglas

En forma análoga a la estimación en HMMs, se propone el algoritmo *inside-outside* que se encarga de reestimar las probabilidades actuales con el objetivo de maximizar la probabilidad de las secuencias de entrenamiento. Este método, tiene las siguientes características:

- No se aprende todo el modelo, ya que la estructura de las reglas está dada, únicamente se estiman las probabilidades.
- Es un algoritmo EM (Expectation-maximization)
- Parte de la suposición que una buena gramática es la que asigna probabilidades altas a las secuencias de entrenamiento

Problemas del algoritmo inside-outside

- Comparado con el algoritmo Baum-Welch es lento, su costo es $O(m^3 n^3)$ donde m es la longitud de la secuencia y n es el número de no terminales de la gramática
- El algoritmo es muy sensible a los valores de inicialización de las probabilidades
- Para realizar un aprendizaje satisfactorio se debe permitir a la gramática muchos más símbolos no terminales que los que se necesitarían teóricamente para describir el lenguaje
- Los no terminales que el algoritmo va utilizando no tienen una interpretación clara a la luz del problema que se está modelando. Esto ocurre aún si se comienza con una gramática como la que construiría un experto en el tema del modelo.

Bibliografía

Foundations of Statistical Natural Language Processing. Christopher D. Manning and Hinrich Schütze. The MIT Press.2002. Capítulo 11.