

Introducción al Procesamiento de Lenguaje Natural Diciembre de 2015

Consideraciones generales

- i) La prueba es sin material escrito.
- ii) Escriba nombre y C.I. en todas las hojas.
- iii) Numere todas las hojas.
- iv) En la primera hoja, indique el total de hojas.
- v) Comience cada ejercicio en una hoja nueva.
- vi) Utilice las hojas de un solo lado.
- vii) Entregue los ejercicios en orden
- viii) El total de puntos es 40

Ejercicio 1 [6 puntos]

- i) Escriba la ecuación fundamental de la Traducción Automática Estadística (SMT), suponiendo que se quiere traducir texto del lenguaje F al lenguaje E.

$$\hat{e} = \arg \max_e P(f | e)P(e)$$

- ii) Para resolver esta ecuación se utilizan tres componentes. Describa brevemente dichos componentes, indicando qué aspecto de la ecuación resuelve cada uno.

Modelo de lenguaje: Utilizado para resolver $P(e)$, que es la componente que mide la fluidez. Se suelen utilizar modelos de n-gramas para construir este modelo.

Modelo de traducción: Utilizado para resolver $P(f|e)$, que es la componente que mide la adecuación. El modelo de traducción más simple, es una tabla que indica la probabilidad de traducir una palabra del lenguaje origen en una palabra del lenguaje destino.

Decodificador: Es un algoritmo de búsqueda que intenta devolver los candidatos e más probables de ser traducción de f. Reduce el espacio de búsqueda de todo el lenguaje E para que el problema se vuelva tratable computacionalmente.

Ejercicio 2 [12 puntos]

- i) Identifique los constituyentes de las siguientes oraciones e indique sus funciones sintácticas y roles semánticos (donde corresponda).
- a) Juan come la manzana de mañana.
 - b) La manzana es comida por Juan de mañana.

Juan come la manzana de mañana.

Constituyente	Función sintáctica	Rol semántico
Juan	Sujeto	Agente
la manzana	Objeto directo	Tema
de mañana	Complemento circunstancial	-

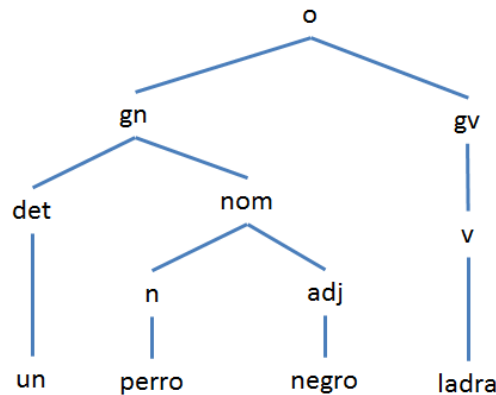
La manzana es comida por Juan de mañana.

Constituyente	Función sintáctica	Rol semántico
La manzana	Sujeto	Tema
por Juan	Complemento agente	Agente
de mañana	Complemento circunstancial	-

ii) Considere la siguiente gramática con anotaciones semánticas:

$o \rightarrow gn\ gv$	$o.sem = gn.sem(gv.sem)$
$gn \rightarrow nprop$	$gn.sem = nprop.sem$
$gn \rightarrow det\ nom$	$gn.sem = det.sem(nom.sem)$
$nom \rightarrow n$	$nom.sem = n.sem$
$nom \rightarrow n\ adj$	$nom.sem = adj.sem(n.sem)$
$gv \rightarrow v$	$gv.sem = v.sem$
$gv \rightarrow v\ gn$	$gv.sem = v.sem(gn.sem)$
$nprop \rightarrow Juan$	$nprop.sem = \lambda P . P(juan)$
$n \rightarrow perro$	$n.sem = \lambda x . perro(x)$
$n \rightarrow gato$	$n.sem = \lambda x . gato(x)$
$v \rightarrow ladra$	$v.sem = \lambda x . ladra(x)$
$v \rightarrow come$	$v.sem = \lambda P . \lambda y . P(\lambda x . come(y, x))$
$adj \rightarrow negro$	$adj.sem = \lambda P . \lambda x . P(x) \wedge negro(x)$
$det \rightarrow todo$	$det.sem = \lambda P . \lambda Q . \forall P(x) \rightarrow Q(x)$
$det \rightarrow un$	$det.sem = \lambda P . \lambda Q . \exists x P(x) \wedge Q(x)$

Dibuje el árbol sintáctico y derive la expresión lógica asociada a la oración: "Un perro negro ladra".



$$\begin{aligned}
 o.sem &= gn.sem(gv.sem) = det.sem(nom.sem)(v.sem) = (\lambda P . \lambda Q . \exists x P(x) \wedge Q(x))(adj.sem(nom.sem))(\lambda x . ladra(x)) = \\
 &= (\lambda P . \lambda Q . \exists x P(x) \wedge Q(x))((\lambda P . \lambda x . P(x) \wedge negro(x))(\lambda x . perro(x)))(\lambda x . ladra(x)) = \\
 &\text{(cambio de variable)} \\
 &= (\lambda P . \lambda Q . \exists x P(x) \wedge Q(x))((\lambda P . \lambda x . P(x) \wedge negro(x))(\lambda y . perro(y)))(\lambda x . ladra(x)) = \\
 &\text{(aplicación funcional)} \\
 &= (\lambda P . \lambda Q . \exists x P(x) \wedge Q(x))(\lambda x . (\lambda y . perro(y))(x) \wedge negro(x))(\lambda x . ladra(x)) = \\
 &\text{(aplicación funcional)} \\
 &= (\lambda P . \lambda Q . \exists x P(x) \wedge Q(x))(\lambda x . perro(x) \wedge negro(x))(\lambda x . ladra(x)) = \\
 &\text{(cambio de variable)} \\
 &= (\lambda P . \lambda Q . \exists x P(x) \wedge Q(x))(\lambda y . perro(y) \wedge negro(y))(\lambda x . ladra(x)) = \\
 &\text{(aplicación funcional)} \\
 &= (\lambda Q . \exists x (\lambda y . perro(y) \wedge negro(y))(x) \wedge Q(x))(\lambda x . ladra(x)) = \\
 &\text{(aplicación funcional)} \\
 &= (\lambda Q . \exists x perro(x) \wedge negro(x) \wedge Q(x))(\lambda x . ladra(x)) =
 \end{aligned}$$

(cambio de variable)
 $= (\lambda Q . \exists x \text{perro}(x) \wedge \text{negro}(x) \wedge Q(x))(\lambda y . \text{ladra}(y)) =$
 (aplicación funcional)
 $= \exists x \text{perro}(x) \wedge \text{negro}(x) \wedge (\lambda y . \text{ladra}(y))(x) =$
 (aplicación funcional)
 $= \exists x \text{perro}(x) \wedge \text{negro}(x) \wedge \text{ladra}(x)$

Ejercicio 3 [12 puntos]

Considere la siguiente gramática, similar a la vista en el curso:

O	→	GV GN GV
GN	→	Det Nom Nom GN GP
GV	→	V V GN V GN GP
GP	→	Prep GN
Det	→	una un el la
Nom	→	Juan sopa leche fideos tomo
V	→	tomo corro
Prep	→	con

1. Aplique el algoritmo Earley para la entrada “tomo una sopa con fideos”, usando la gramática G. ¿Qué salida devuelve el algoritmo?

Chart[0]

$\gamma \rightarrow \cdot O$	[0,0]	Dummy
$O \rightarrow \cdot GV$	[0,0]	Predict
$O \rightarrow \cdot GN GV$	[0,0]	Predict
$GV \rightarrow \cdot V$	[0,0]	Predict
$GV \rightarrow \cdot V GN$	[0,0]	Predict
$GV \rightarrow \cdot V GN GP$	[0,0]	Predict
$GN \rightarrow \cdot Det Nom$	[0,0]	Predict
$GN \rightarrow \cdot Nom$	[0,0]	Predict
$GN \rightarrow \cdot GN GP$	[0,0]	Predict

Chart[1]

$V \rightarrow \text{tomo} \cdot$	[0,1]	Scan -
$Nom \rightarrow \text{tomo} \cdot$	[0,1]	Scan -
$GV \rightarrow V \cdot$	[0,1]	Complete
$GV \rightarrow V \cdot GN$	[0,1]	Complete
$GV \rightarrow V \cdot GN GP$	[0,1]	Complete
$GN \rightarrow Nom \cdot$	[0,1]	Complete
$O \rightarrow GV \cdot$	[0,1]	Complete
$GN \rightarrow \cdot Det Nom$	[1,1]	Predict
$GN \rightarrow \cdot Nom$	[1,1]	Predict
$GN \rightarrow \cdot GN GP$	[1,1]	Predict
$O \rightarrow GN \cdot GV$	[0,1]	Complete
$GN \rightarrow GN \cdot GP$	[0,1]	Complete
$\gamma \rightarrow O \cdot$	[0,1]	Complete
$GV \rightarrow \cdot V$	[1,1]	Predict
$GV \rightarrow \cdot V GN$	[1,1]	Predict
$GV \rightarrow \cdot V GN GP$	[1,1]	Predict
$GP \rightarrow \cdot Prep GN$	[1,1]	Predict

Chart[2]
Det → una · [1,2] Scan
GN → Det · Nom [1,2] Complete

Chart[3]
Nom → sopa · [2,3] Scan
GN → Det Nom · [1,3] Complete

GV → V GN · [0,3] Complete
GV → V GN · GP [0,3] Complete
GN → GN · GP [1,3] Complete
O → GV · [0,3] Complete
γ → O · [0,3] Complete
GP → · Prep GN [3,3] Predict

chart[4]

Prep → con · [3,4] Scan
GP → Prep · GN [3,4] Complete
GN → · Det Nom [4,4] Predict
GN → · Nom [4,4] Predict
GN → · GN GP [4,4] Predict

chart[5]

Nom → fideos · [4,5] Scan -
GN → Nom · [4,5] Complete
GP → Prep GN · [3,5] Complete
GN → GN · GP [4,5] Complete
GN → GN GP · [1,5] Complete
O → GV · [0,5] Complete
GV → V GN · [0,5] Complete
GV → V GN · GP [0,5] Complete
γ → **O · [0,5] Complete**
GP → · Prep GN [5,5] Predict

El algoritmo devuelve True porque llegó a una regla con el punto a la derecha, que reconoce el símbolo inicial y abarca toda la entrada.

2. Dibuje los árboles sintácticos posibles para la oración. ¿Es posible obtener esta información del resultado del algoritmo? Justifique brevemente.

[[Tomo/V[[una/Det sopa/Nom [con/Prep fideos/Nom]_{GP}]_{GN}]_{GV}]_O
[[Tomo/V [una/Det sopa/Nom]_{GN} [con/Prep fideos/Nom]_{GP}]_{GV}]_O

No, no es posible. El algoritmo es un reconocedor, para convertirlo en un parser habría que guardar la información de qué regla se aplicó en cada paso, obteniendo los constituyentes que forman cada subárbol.

3. ¿El algoritmo utilizado expresa de alguna forma preferencia por uno de los árboles? ¿Y si se utilizaran algoritmos de *parsing* probabilísticos? ¿Y si se agregara algún tipo de información semántica a los algoritmos? Justifique brevemente las respuestas.

No, cualquier árbol consistente con la gramática es igual. Si se utilizaran algoritmos probabilísticos, entonces habría reglas que se aplicarían de forma más probable, por lo que la probabilidad de ambos árboles sería distinta. Si se incorporara información semántica (por ejemplo, que “fideos” no es una herramienta), habría forma de preferir un árbol sobre el otro.

Ejercicio 4 [10 puntos]

1. ¿Qué quiere decir que los transductores de estado finito son cerrados bajo composición? ¿Por qué es esto una ventaja para el análisis morfológico? ¿Por qué la morfología de estado finito resuelve al mismo tiempo el problema del análisis y de la generación?
2. Describa el algoritmo Naïve Bayes para clasificación. ¿Cómo se estiman las probabilidades involucradas?
3. ¿Cuál es la diferencia entre un clasificador generativo y uno discriminativo? Nombre un clasificador *secuencial* generativo y otro discriminativo.
4. Explique brevemente en qué consiste el criterio td-idf para la asignación de pesos a los términos de indización en un sistema de recuperación de información

Véase el teórico.