

Processament del Llenguatge Humà 7. Seqüències de paraules



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona

FIB

Outline

Secuencias de
palabras

Métodos

- 1 Secuencias de palabras
 - Objetivo y motivación
- 2 Métodos
 - Reglas hechas a mano
 - Modelos discriminativos
 - Conditional Random Fields

Outline

Secuencias de palabras

Objetivo y motivación

Métodos

- 1 Secuencias de palabras
 - Objetivo y motivación

- 2 Métodos
 - Reglas hechas a mano
 - Modelos discriminativos
 - Conditional Random Fields

Objetivo

Reconocer y clasificar secuencias de palabras relevantes para la interpretación del lenguaje natural (p.e., NERC y NP-chunking)

- **Entidades nombradas** (NEs - Named Entities): Persona, lugar, organización, fecha, tiempo, dinero
 - P.e.: "[John Smith]/PER was in [Picadilli Circus]/LOC at [3:00pm]/TIME"
 - P.e.: "[Heart attack]/DISEASE at [8:30am]/TIME. Admitted to the intensive care unit at [St. James]/HOSPITAL

Secuencias de palabras

Objetivo y motivación

Métodos

Objetivo

Reconocer y clasificar secuencias de palabras relevantes para la interpretación del lenguaje natural (p.e., NERC y NP-chunking)

Secuencias de palabras

Objetivo y motivación

Métodos

- **Entidades nombradas** (NEs - Named Entities): Persona, lugar, organización, fecha, tiempo, dinero
 - P.e.: "[John Smith]/PER was in [Picadilli Circus]/LOC at [3:00pm]/TIME"
 - P.e.: "[Heart attack]/DISEASE at [8:30am]/TIME. Admitted to the intensive care unit at [St. James]/HOSPITAL
- **Frasas nominales** (NPs - Noun phrases): ¿NPs simples? ¿también complejos?
 - P.e.: "[Spaniards] usually enjoy [the original dishes] cooked by [Ferràn Adrià]"
 - P.e.: "[Spaniards] usually enjoy [the original dishes cooked by Ferràn Adrià]"

Motivation

Ejemplos de aplicaciones:

- Anonimización: eliminar o enmascarar información personal de documentos privados

P.e.: Nombres de personas ,direcciones, teléfonos, etc. en informes clínicos

Secuencias de palabras

Objetivo y motivación

Métodos

Motivation

Ejemplos de aplicaciones:

- Anonimización: eliminar o enmascarar información personal de documentos privados
P.e.: Nombres de personas ,direcciones, teléfonos, etc. en informes clínicos
- Extracción de Información
P.e.: Extraer empleados de empresas, sus cargos y eventos en los que participan a partir de noticias

Motivation

Ejemplos de aplicaciones:

- Anonimización: eliminar o enmascarar información personal de documentos privados
P.e.: Nombres de personas ,direcciones, teléfonos, etc. en informes clínicos
- Extracción de Información
P.e.: Extraer empleados de empresas, sus cargos y eventos en los que participan a partir de noticias
- *Question answering*: encontrar el foco de algunos tipos de preguntas o indexar documentos por nombres propios
P.e.: Who was [Albert Einstein]?
P.e.: [Albert Einstein] was [the physicist who formulated the theory of relativity]
- Traducción automática, ...

Outline

Secuencias de palabras

Métodos

- 1 Secuencias de palabras
 - Objetivo y motivación
- 2 Métodos
 - Reglas hechas a mano
 - Modelos discriminativos
 - Conditional Random Fields

Métodos

Reglas hechas a mano:

- Usadas para casos robustos (p.e., NPs y NEs simples como teléfonos, emails, genes, proteínas)
- Pueden integrarse en métodos basados en aprendizaje automático

Métodos

Reglas hechas a mano:

- Usadas para casos robustos (p.e., NPs y NEs simples como teléfonos, emails, genes, proteínas)
- Pueden integrarse en métodos basados en aprendizaje automático

Aprendizaje automático:

- Basado en atributos: [Conditional Random Fields \(CRFs\)](#), SVMs, ...
- Deep learning:
 - word embeddings + MLP + softmax
 - word embeddings + BiLSTM + CRF
 - distribucionales (word2vec, fasttext, Glove,...) o contextuales (transformers - familias BERT, GPT, LLAMA, ...)
 - LLMs
 - ...

Métodos

Reglas hechas a mano:

- Usadas para casos robustos (p.e., NPs y NEs simples como teléfonos, emails, genes, proteínas)
- Pueden integrarse en métodos basados en aprendizaje automático

Aprendizaje automático:

- Basado en atributos: [Conditional Random Fields \(CRFs\)](#), SVMs, ...
- Deep learning:
 - word embeddings + MLP + softmax
 - word embeddings + BiLSTM + CRF
 - distribucionales (word2vec, fasttext, Glove,...) o contextuales (transformers - familias BERT, GPT, LLAMA, ...)
 - LLMs
 - ...

CRFs puede ser mejor en dominios específicos como Biomedicina.

Deep learning requiere grandes cantidades de datos d'entrenamiento

Outline

Secuencias de palabras

Métodos

Reglas hechas a mano

- 1 Secuencias de palabras
 - Objetivo y motivación
- 2 Métodos
 - Reglas hechas a mano
 - Modelos discriminativos
 - Conditional Random Fields

Reglas para casos simples de NERC

- Patrones de palabras y, quizás, etiquetas POS
- Listas de palabras clave y contextuales útiles
P.e.: Para DATE, meses, días de la semana, días señalados

Secuencias de palabras

Métodos

Reglas hechas a mano

Reglas para casos simples de NERC

- Patrones de palabras y, quizás, etiquetas POS
- Listas de palabras clave y contextuales útiles
P.e.: Para DATE, meses, días de la semana, días señalados

Ejemplos: (con expresiones regulares)

Input:

"My phone number is 934104433 . Call me on Tuesday 13 at 8:00 pm . "

Output:

"My phone number is [TEL 934104433] . Call me on [DATE Tuesday 13] at [TIME 8:00 pm] . "

Reglas para casos simples de NERC

- Patrones de palabras y, quizás, etiquetas POS
- Listas de palabras clave y contextuales útiles
P.e.: Para DATE, meses, días de la semana, días señalados

Ejemplos: (con expresiones regulares)

Input:

"My phone number is 934104433 . Call me on Tuesday 13 at 8:00 pm . "

Output:

"My phone number is [TEL 934104433] . Call me on [DATE Tuesday 13] at [TIME 8:00 pm] . "

1. ... phone number is (\d+) ... → ... phone number is [TEL *match*] ...

Reglas para casos simples de NERC

- Patrones de palabras y, quizás, etiquetas POS
- Listas de palabras clave y contextuales útiles
P.e.: Para DATE, meses, días de la semana, días señalados

Ejemplos: (con expresiones regulares)

Input:

"My phone number is 934104433 . Call me on Tuesday 13 at 8:00 pm . "

Output:

"My phone number is [TEL 934104433] . Call me on [DATE Tuesday 13] at [TIME 8:00 pm] . "

1. ... phone number is (\d+) ... → ... phone number is [TEL *match*] ...
2. DAY= '{Monday|Tuesday|Wednesday| ...}'
... on (\$DAY \d+) ... → ... on [DATE *match*]

Reglas para casos simples de NERC

- Patrones de palabras y, quizás, etiquetas POS
- Listas de palabras clave y contextuales útiles
P.e.: Para DATE, meses, días de la semana, días señalados

Secuencias de palabras

Métodos

Reglas hechas a mano

Ejemplos: (con expresiones regulares)

Input:

"My phone number is 934104433 . Call me on Tuesday 13 at 8:00 pm . "

Output:

"My phone number is [TEL 934104433] . Call me on [DATE Tuesday 13] at [TIME 8:00 pm] . "

1. ... phone number is (`\d+`) ... → ... phone number is [TEL *match*] ...
2. DAY= '{Monday|Tuesday|Wednesday| ...}'
... on (`$DAY \d+`) ... → ... on [DATE *match*]
3. SLOT= '{pm|p.m.|p.m|am|a.m.|a.m}'
... at (`\d{1:2}:\d\d $SLOT`) ... → ... at [TIME *match*] ...

Reglas para NPs simples

- Patrones de etiquetas POS
- Patrones que usan información sintáctica

Secuencias de palabras

Métodos

Reglas hechas a mano

Reglas para NPs simples

- Patrones de etiquetas POS
- Patrones que usan información sintáctica

Ejemplos: (con expresiones regulares)

Input:

"The:DT cat:NN eats:VBZ in:IN the:DT dark:JJ room:NN "

Output:

"[NP The:DT cat:NN] eats:VBZ in:IN [NP the:DT dark:JJ room:NN] "

Reglas para NPs simples

- Patrones de etiquetas POS
- Patrones que usan información sintáctica

Secuencias de palabras

Métodos

Reglas hechas a mano

Ejemplos: (con expresiones regulares)

Input:

"The:DT cat:NN eats:VBZ in:IN the:DT dark:JJ room:NN "

Output:

"[NP The:DT cat:NN] eats:VBZ in:IN [NP the:DT dark:JJ room:NN] "

1. ... (`\w+:DT \w+:NN`) ... → ... [NP *match*] ...
2. ... (`\w+:DT (\w+:JJ)+ \w+:NN`) ... → ... [NP *match*] ...

Reglas para NPs simples

- Patrones de etiquetas POS
- Patrones que usan información sintáctica

Secuencias de palabras

Métodos

Reglas hechas a mano

Ejemplos: (con expresiones regulares)

Input:

"The:DT cat:NN eats:VBZ in:IN the:DT dark:JJ room:NN "

Output:

"[NP The:DT cat:NN] eats:VBZ in:IN [NP the:DT dark:JJ room:NN] "

1. ... (`\w+:DT \w+:NN`) ... → ... [NP *match*] ...
2. ... (`\w+:DT (\w+:JJ)+ \w+:NN`) ... → ... [NP *match*] ...

Que equivalen a

1. ... (`\w+:DT (\w+:JJ)* \w+:NN`) ... → ... [NP *match*] ...

Ejercicio

Secuencias de palabras

Métodos

Reglas hechas a mano

- 1 Escribe patrones de NERC para reconocer expresiones como:
 - a) "tomorrow:NN morning::NN", "in:IN the:DT evening:NN"" after:IN this:DT Sunday:NN"
 - b) "5:CD €:NN", "one:CD million:CD dollars:NNS"
 - c) "ana.sanchez@gmail.com", "ana.sanchez at gmail dot com"
- 2 Escribe patrones para reconocer los NPs simples de estas oraciones:
 - d) "We:PRP 're:VB going:VBG to:TO the:DT best:JJ cinema:NN with:IN Gina:NNP 's:RP father:NN and:CC 24:CD friends:NNS"
 - e) "Workers:NNS of:IN car:NN parks:NNS hate:VB working:VBG after:IN 7:00:Z pm:NN "
- 3 ¿Es adecuado atacar todos estos casos usando reglas hechas a mano?

Outline

Secuencias de palabras

Métodos

Modelos
discriminativos

- 1 Secuencias de palabras
 - Objetivo y motivación
- 2 Métodos
 - Reglas hechas a mano
 - Modelos discriminativos
 - Conditional Random Fields

Anotación BIO de ejemplos

Anotación manual del corpus de entrenamiento:

$$w_1 w_2 w_3 \dots [CLASE w_i w_{i+1}] \dots w_n$$

Transformación via programa:

$$w_1:O w_2:O w_3:O \dots w_i:B-CLASE w_{i+1}:I-CLASE \dots w_n:O$$

código BIO: B: beginning; I: inside; O: outside

código BIOS: S: single token (si secuencias de 1 token frecuentes)

código BIOES [BILOU]: E: end

Anotación BIO de ejemplos

Anotación manual del corpus de entrenamiento:

$$w_1 \ w_2 \ w_3 \ \dots \ [CLASE \ w_i \ w_{i+1}] \ \dots \ w_n$$

Transformación via programa:

$$w_1:O \ w_2:O \ w_3:O \ \dots \ w_i:B-CLASE \ w_{i+1}:I-CLASE \ \dots \ w_n:O$$

código BIO: B: beginning; I: inside; O: outside

código BIOS: S: single token (si secuencias de 1 token frecuentes)

código BIOES [BILOU]: E: end

Ejemplos:

- "The president of [LOC the US] , [PER D. Trump]"
"The:O president:O of:O the:B-LOC US:I-LOC ,:O D.:B-PER
Trump:I-PER"
- "[NP The president] of [NP the US] , [NP D. Trump]"
"The:B president:I of:O the:B US:I ,:O D.:B Trump:I"

Outline

Secuencias de palabras

Métodos

Conditional Random Fields

- 1 Secuencias de palabras
 - Objetivo y motivación
- 2 Métodos
 - Reglas hechas a mano
 - Modelos discriminativos
 - Conditional Random Fields

Conditional Random Fields

- Generalización de HMMs
- **HMMs**: Naïve Bayes aplicado a una secuencia
 - Basados en probabilidad conjunta (Modelo generativo)

$$P(X|O) \approx P(X, O) = P(X_1, \dots, X_T) \cdot P(O_1, \dots, O_T | X_1, \dots, X_T)$$

Conditional Random Fields

- Generalización de HMMs
- **HMMs**: Naïve Bayes aplicado a una secuencia
 - Basados en probabilidad conjunta (Modelo generativo)

$$P(X|O) \approx P(X, O) = P(X_1, \dots, X_T) \cdot P(O_1, \dots, O_T | X_1, \dots, X_T)$$

- **CRFs**: regresión logística aplicada a una secuencia
 - Basados en probabilidad condicional (Modelo discriminativo)

$$P(X|O) = \frac{1}{Z(O)} \cdot \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

$$Z(O) = \sum_X \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

f_k : funciones de atributos binarios sobre X_{t-1} , X_t y O
(existen CRFs de orden superior)

Aprendizaje de los parámetros λ_i

$$P(X|O) = \frac{1}{Z(O)} \cdot \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

- Maximizar la probabilidad de las secuencias de etiquetas que ocurren en los datos de entrenamiento
- Métodos de optimización: métodos cuasi Newton, gradiente conjugado, escalado iterativo

Este tema queda fuera de nuestro curso

Tipos de funciones de atributos

$$P(X|O) = \frac{1}{Z(O)} \cdot \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

1 De observación:

$$\text{Ex: } f_1(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = s_3 \text{ and has_property}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

Tipos de funciones de atributos

$$P(X|O) = \frac{1}{Z(O)} \cdot \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

1 De observación:

$$\text{Ex: } f_1(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = s_3 \text{ and has_property}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

2 De transición:

$$\text{Ex: } f_2(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = s_3 \text{ and } x_{t-1} = s_6 \\ 0 & \text{otherwise} \end{cases}$$

Tipos de funciones de atributos

$$P(X|O) = \frac{1}{Z(O)} \cdot \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

1 De observación:

$$\text{Ex: } f_1(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = s_3 \text{ and has_property}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

2 De transición:

$$\text{Ex: } f_2(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = s_3 \text{ and } x_{t-1} = s_6 \\ 0 & \text{otherwise} \end{cases}$$

3 Mixta:

$$\text{Ex: } f_3(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = s_3 \text{ and } x_{t-1} = s_6 \text{ and } o_t = w_4 \\ 0 & \text{otherwise} \end{cases}$$

Diseño de plantillas de atributos

$$P(X|O) = \frac{1}{Z(O)} \cdot \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

1 De observación:

$$\text{Ex: } f_{1,a,b_1,\dots,b_k}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = a \text{ and has_property}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

2 De transición:

$$\text{Ex: } f_{2,a,c}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = a \text{ and } x_{t-1} = c \\ 0 & \text{otherwise} \end{cases}$$

3 Mixta:

$$\text{Ex: } f_{3,a,b_i,c}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = a \text{ and } x_{t-1} = c \text{ and } o_t = b_i \\ 0 & \text{otherwise} \end{cases}$$

Funciones de atributos correctas vs. útiles

$$P(X|O) = \frac{1}{Z(O)} \cdot \exp\left(\sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\right)$$

- Función correcta:
 - x_t siempre debe estar definida
 - no puede definir ningún elemento fuera de los parámetros
- Función útil:
 - tienen sentido para la tarea particular
 - λ_j no es nula

Modelado de NERC con CRFs

- Los estados s_i son las etiquetas B-CLASE_r, I-CLASE_r y O
- Plantillas de atributos

P.e.: La palabra actual esta en mayusculas y su etiqueta es a

$$f_{1,a}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = a \text{ and capitalized}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

Modelado de NERC con CRFs

Secuencias de palabras

Métodos

Conditional Random Fields

- Los estados s_i son las etiquetas B-CLASE_r, I-CLASE_r y O
- Plantillas de atributos

P.e.: La palabra actual esta en mayúsculas y su etiqueta es a

$$f_{1,a}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = a \text{ and capitalized}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

- Se derivan automáticamente las funciones de atributos.

P.e.: Dos funciones de atributos derivadas a partir de $f_{1,a}$

$$f_{1,\text{B-PER}}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = \text{B-PER and capitalized}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

$$f_{1,\text{O}}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } x_t = \text{O and capitalized}(o_t) \\ 0 & \text{otherwise} \end{cases}$$

Modelado de NP-*chunking* con CRFs

- Los estados s_i son las etiquetas B, I, O
- Plantillas de atributos

P.e.: La POS de la palabra actual es a y la etiqueta actual es b

$$f_{1,a,b}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } \text{pos}(o_t)=a \text{ and } x_t = b \\ 0 & \text{otherwise} \end{cases}$$

Modelado de NP-*chunking* con CRFs

- Los estados s_i son las etiquetas B, I, O
- Plantillas de atributos

P.e.: La POS de la palabra actual es a y la etiqueta actual es b

$$f_{1,a,b}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } \text{pos}(o_t)=a \text{ and } x_t = b \\ 0 & \text{otherwise} \end{cases}$$

- Funciones de atributos

P.e.: Tres funciones derivadas a partir de $f_{1,a,b}$:

$$f_{1,DT,B}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } \text{pos}(o_t)=DT \text{ and } x_t=B \\ 0 & \text{otherwise} \end{cases}$$

$$f_{1,NN,I}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } \text{pos}(o_t)=NN \text{ and } x_t=I \\ 0 & \text{otherwise} \end{cases}$$

$$f_{1,VB,O}(x_{t-1}, x_t, O, t) = \begin{cases} 1 & \text{if } \text{pos}(o_t)=VB \text{ and } x_t=O \\ 0 & \text{otherwise} \end{cases}$$

Ejercicio

Escribe plantillas de atributos para las siguientes descripciones. De ellas, deriva algunas funciones de atributos.

Usualmente para NERC:

- La etiqueta previa es a , la etiqueta actual es b y la palabra actual empieza con mayúscula
- La etiqueta actual es a y la siguiente palabra es w
- El nombre de una persona puede estar precedido por un título (mr., dr., . . .)

Usualmente para NP-*chunking*:

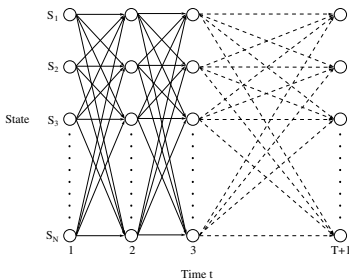
- La POS de la palabra actual es a y la etiqueta actual es b
- La POS de la palabra anterior es a , la etiqueta previa es b y la etiqueta actual es c

¿Cómo encontrar la mejor secuencia de etiquetas?

- Queremos encontrar

$$\begin{aligned}\hat{X} &= \operatorname{argmax}_X P(X|O, \lambda) = \operatorname{argmax}_X \exp \sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t) \\ &= \operatorname{argmax}_X \sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)\end{aligned}$$

- Usamos Viterbi modificado para CRFs



Trellis del CRF totalmente conectado.

Un nodo $\{s_j, X_t\}$ guarda información sobre las secuencias parciales de X_1 a $X_t = s_j$.

$$\begin{aligned}\{s_j, X_t\}: \quad \delta_t(j) &= \max_{X_1, \dots, X_{t-1}} P(X_1, \dots, X_{t-1}, X_t = s_j | O, \lambda) \\ \varphi_t(j) &= \operatorname{last}(\operatorname{argmax}_{X_1, \dots, X_{t-1}} P(X_1, \dots, X_{t-1}, X_t = s_j | O, \lambda))\end{aligned}$$

¿Cómo encontrar la mejor secuencia de etiquetas?

- Queremos encontrar

$$\hat{X} = \underset{X}{\operatorname{argmax}} \sum_t \sum_k \lambda_k \cdot f_k(x_{t-1}, x_t, O, t)$$

- Usamos Viterbi modificado para CRFs

- 1 Inicialización: $\forall j = 1 \dots N$

$$\delta_1(j) = \sum_k \lambda_k \cdot f_k(x_0 = *, x_1 = s_j, O, t)$$

- 2 Inducción: $\forall j = 1 \dots N$

$$\delta_t(j) = \max_i [\delta_{t-1}(i) + \sum_k \lambda_k \cdot f_k(x_{t-1} = s_i, x_t = s_j, O, t)]$$

$$\varphi_t(j) = \operatorname{argmax}_i [\delta_{t-1}(i) + \sum_k \lambda_k \cdot f_k(x_{i-1} = s_i, x_i = s_j, O, t)]$$

- 3 Finalización:

$$\hat{X}_T = \operatorname{argmax}_i \delta_T(i)$$

- 4 Reconstrucción del camino:

$$\hat{X}_t = \varphi_{t+1}(\hat{X}_{t+1})$$