

Aprendizaje de etiquetas basado en transformaciones

El etiquetador de Brill

Miguel A. Alonso Jorge Graña Jesús Vilares

Departamento de Computación, Facultad de Informática, Universidade da Coruña

- 1 Etiquetador léxico
- 2 Etiquetador de palabras desconocidas
- 3 Etiquetador contextual
- 4 Aprendizaje basado en transformaciones y dirigido por el error
- 5 Consideraciones de eficiencia

El etiquetador de Brill se fundamenta en el **aprendizaje basado en transformaciones y dirigido por el error**

Consta de:

- Etiquetador léxico
- Etiquetador de palabras desconocidas
- Etiquetador contextual

Etiquetador léxico

- Cada palabra se etiqueta inicialmente con su etiqueta más probable
- No se tiene en cuenta el contexto en el que aparece
- Palabras desconocidas:
 - sustantivo propio si la primera letra es mayúscula
 - sustantivo común en otro caso

Etiquetador de palabras desconocidas

Intenta *adivinar* una etiqueta para una palabra desconocida en función de su sufijo, de su prefijo, y de otras propiedades relevantes similares

Cada transformación consta de dos partes

- una descripción del contexto de aplicación
- una regla de reescritura que reemplaza una etiqueta por otra

Plantilla genérica de transformaciones léxicas

- A y B son variables sobre el conjunto de todas las etiquetas
- x es cualquier cadena de caracteres de longitud 1, 2, 3 o 4
- l es la longitud de dicha cadena
- w es cualquier palabra
- z es cualquier caracter.

Plantilla genérica de transformaciones léxicas

- $x \text{ haspref } l A$: si los primeros l caracteres de la palabra son x , se asigna a la palabra desconocida la etiqueta A
- $A x \text{ fhaspref } l B$: si la etiqueta actual de la palabra es A y sus primeros l caracteres son x , se cambia dicha etiqueta por B
- $x \text{ deletepref } l A$: si borrando el prefijo x de longitud l obtenemos una palabra conocida, se asigna a la palabra desconocida la etiqueta A
- $A x \text{ fdeletepref } l B$: si la etiqueta actual de la palabra es A y borrando el prefijo x de longitud l obtenemos una palabra conocida, se cambia dicha etiqueta por B
- $x \text{ addpref } l A$: si añadiendo el prefijo x de longitud l obtenemos una palabra conocida, se asigna a la palabra desconocida la etiqueta A
- $A x \text{ faddpref } l B$: si la etiqueta actual de la palabra es A y añadiendo el prefijo x de longitud l obtenemos una palabra conocida, se cambia dicha etiqueta por B
- $x \text{ hassuf } l A$: si los últimos l caracteres de la palabra son x , se asigna a la palabra desconocida la etiqueta A
- $A x \text{ fhassuf } l B$: si la etiqueta actual de la palabra es A y sus últimos l caracteres son x , se cambia dicha etiqueta por B
- $x \text{ deletesuf } l A$: si borrando el sufijo x de longitud l obtenemos una palabra conocida, se asigna a la palabra desconocida la etiqueta A
- $A x \text{ fdeletesuf } l B$: si la etiqueta actual de la palabra es A y borrando el sufijo x de longitud l obtenemos una palabra conocida, se cambia dicha etiqueta por B
- $x \text{ addsuf } l A$: si añadiendo el sufijo x de longitud l obtenemos una palabra conocida, se asigna a la palabra desconocida la etiqueta A
- $A x \text{ faddsuf } l B$: si la etiqueta actual de la palabra es A y añadiendo el sufijo x de longitud l obtenemos una palabra conocida, se cambia dicha etiqueta por B

Plantilla genérica de transformaciones léxicas

- w `goodright` A : si la palabra aparece inmediatamente a la derecha de la palabra w , se asigna a la palabra desconocida la etiqueta A
- A w `fgoodright` B : si la etiqueta actual de la palabra es A y aparece inmediatamente a la derecha de la palabra w , se cambia dicha etiqueta por B
- w `goodleft` A : si la palabra aparece inmediatamente a la izquierda de la palabra w , se asigna a la palabra desconocida la etiqueta A
- A w `fgoodleft` B : si la etiqueta actual de la palabra es A y aparece inmediatamente a la izquierda de la palabra w , se cambia dicha etiqueta por B
- z `char` A : si el carácter z aparece en la palabra, se asigna a la palabra desconocida la etiqueta A
- A z `fchar` B : si la etiqueta actual de la palabra es A y el carácter z aparece en la palabra, se cambia dicha etiqueta por B

Ejemplos en español

- **rse hassuf 3 V000f0PE1**: si los últimos 3 caracteres de la palabra son rse, se asigna a la palabra desconocida la etiqueta V000f0PE1, es decir, verbo infinitivo con un pronombre enclítico
- **r hassuf 1 V000f0**: si el último carácter de la palabra es r, se asigna a la palabra desconocida la etiqueta V000f0, es decir, verbo infinitivo
- **V000f0 or fhassuf 2 Scms**: si la etiqueta actual de la palabra es V000f0, es decir, verbo infinitivo, y sus últimos 2 caracteres son or, se cambia dicha etiqueta por la etiqueta Scms, es decir, sustantivo común, masculino, singular
- **ría deletesuf 3 Vysci0**: si borrando el sufijo ría de longitud 3 obtenemos una palabra conocida, se asigna a la palabra desconocida la etiqueta Vysci0, es decir, verbo, primera y tercera personas del singular, postpretérito de indicativo
- **textttScfs r faddsuf 1 V3spi0**: si la etiqueta actual de la palabra es Scfs, es decir, sustantivo, común, femenino, singular, y añadiendo el sufijo r de longitud 1 obtenemos una palabra conocida, se cambia dicha etiqueta por la etiqueta V3spi0, es decir, verbo, tercera persona del singular, presente de indicativo

Ejemplos en español

- `el goodright Scms`: si la palabra aparece inmediatamente a la derecha de la palabra `e1`, se asigna a la palabra desconocida la etiqueta `Scms`, es decir, sustantivo común, masculino, singular
- `Scmp las fgoodright Scfp`: si la etiqueta actual de la palabra es `Scmp`, es decir, sustantivo común, masculino, plural, y aparece inmediatamente a la derecha de la palabra `las`, se cambia dicha etiqueta por la etiqueta `Scfp`, es decir, sustantivo común femenino, plural
- `% goodleft Ncyyp`: si la palabra aparece inmediatamente a la izquierda de la palabra `%`, se asigna a la palabra desconocida la etiqueta `Ncyyp`, es decir, numeral cardinal, determinante y no determinante, masculino y femenino, plural
- `w char Ze00`: si el carácter `w` aparece en la palabra, se asigna a la palabra desconocida la etiqueta `Ze00`, es decir, palabra extranjera

Problema: limitación de 4 caracteres de longitud en los prefijos y sufijos de las reglas extraídas automáticamente

Etiquetador contextual

- El etiquetador contextual actúa justo después del etiquetador de palabras desconocidas
- Aplica en orden una secuencia de reglas contextuales que han sido aprendidas de manera automática a partir del corpus de entrenamiento
 - A , B , C y D son variables sobre el conjunto de todas las etiquetas
 - w y x son cualquier palabra

Plantilla genérica de transformaciones contextuales

- **A B prevtag C**: cambiar la etiqueta A por B si la palabra anterior aparece etiquetada con la etiqueta C
- **A B prev1or2tag C**: cambiar la etiqueta A por B si una de las dos palabras anteriores aparece etiquetada con la etiqueta C
- **A B prev1or2or3tag C**: cambiar la etiqueta A por B si una de las tres palabras anteriores aparece etiquetada con la etiqueta C
- **A B prev2tag C**: cambiar la etiqueta A por B si la segunda palabra anterior aparece etiquetada con la etiqueta C
- **A B nexttag C**: cambiar la etiqueta A por B si la palabra siguiente aparece etiquetada con la etiqueta C
- **A B next1or2tag C**: cambiar la etiqueta A por B si una de las dos palabras siguientes aparece etiquetada con la etiqueta C
- **A B next1or2or3tag C**: cambiar la etiqueta A por B si una de las tres palabras siguientes aparece etiquetada con la etiqueta C
- **A B next2tag C**: cambiar la etiqueta A por B si la segunda palabra siguiente aparece etiquetada con la etiqueta C
- **A B prevbigram C D**: cambiar la etiqueta A por B si la palabra anterior aparece etiquetada con la etiqueta C y la segunda palabra anterior con D
- **A B nextbigram C D**: cambiar la etiqueta A por B si la palabra siguiente aparece etiquetada con la etiqueta C y la segunda palabra siguiente con D
- **A B surroundtag C D**: cambiar la etiqueta A por B si la palabra anterior aparece etiquetada con la etiqueta C y la siguiente con D

Plantilla genérica de transformaciones contextuales

- $A B \text{ curwd } w$: cambiar la etiqueta A por B si la palabra actual es w
- $A B \text{ prevwd } w$: cambiar la etiqueta A por B si la palabra anterior es w
- $A B \text{ prev1or2wd } w$: cambiar la etiqueta A por B si una de las dos palabra anteriores es w
- $A B \text{ prev2wd } w$: cambiar la etiqueta A por B si la segunda palabra anterior es w
- $A B \text{ nextwd } w$: cambiar la etiqueta A por B si la palabra siguiente es w
- $A B \text{ next1or2wd } w$: cambiar la etiqueta A por la etiqueta B si una de las dos palabras siguientes es w
- $A B \text{ next2wd } w$: cambiar la etiqueta A por B si la segunda palabra siguiente es w
- $A B \text{ lbigram } w x$: cambiar la etiqueta A por B si las dos palabras anteriores son w y x
- $A B \text{ rbigram } w x$: cambiar la etiqueta A por B si las dos palabras siguientes son w y x
- $A B \text{ wdand2bfr } x w$: cambiar la etiqueta A por B si la palabra actual es w y la segunda palabra anterior es x
- $A B \text{ wdand2aft } w x$: cambiar la etiqueta A por B si la palabra actual es w y la segunda palabra siguiente es x
- $A B \text{ wdprevttag } C w$: cambiar la etiqueta A por B si la palabra actual es w y la anterior aparece etiquetada con la etiqueta C
- $A B \text{ wdnexntag } w C$: cambiar la etiqueta A por B si la palabra actual es w y la siguiente aparece etiquetada con la etiqueta C
- $A B \text{ wdand2tagbfr } C w$: cambiar la etiqueta A por B si la palabra actual es w y la segunda palabra anterior aparece etiquetada con la etiqueta C
- $A B \text{ wdand2tagaft } w C$: cambiar la etiqueta A por B si la palabra actual es w y la segunda palabra siguiente aparece etiquetada con la etiqueta C

Contextos de las reglas

	$i-3$	$i-2$	$i-1$	i	$i+1$	$i+2$	$i+3$
1				*			
2				*			
3				*			
4				*			
5				*			
6				*			
7				*			
8				*			
9				*			
10				*			
11				*			
12				*			
13				*			
14				*			
15				*			
16				*			
17				*			
18				*			
19				*			
20				*			
21				*			
22				*			

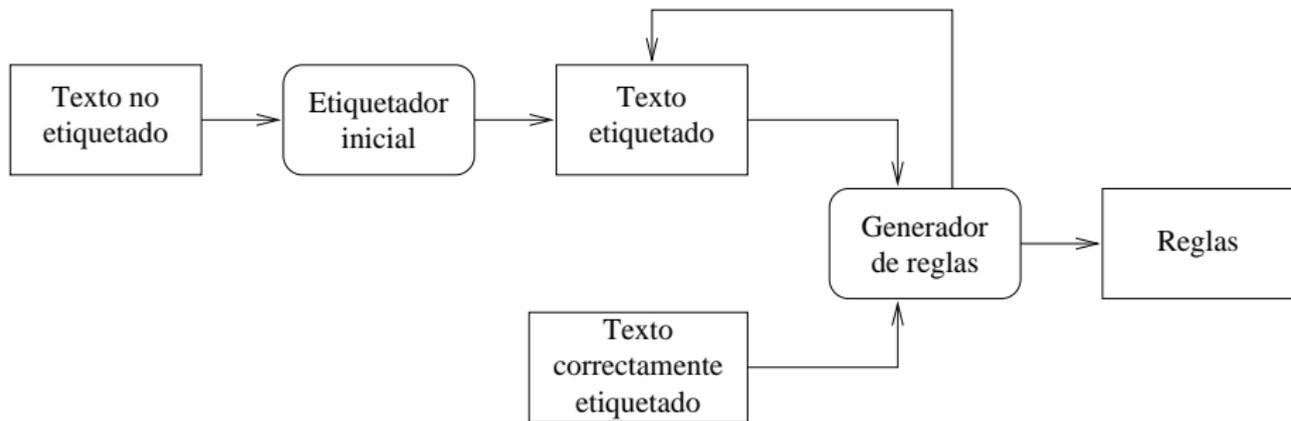
* = palabra a etiquetar; recuadros = etiqueta; sombreados = palabra

Ejemplos en español

- **Afp0 Amp0 prevtag Scmp**: cambiar la etiqueta Afp0, es decir, adjetivo, femenino, plural, sin grado, por Amp0, es decir, adjetivo, masculino, plural, sin grado, si la palabra anterior aparece etiquetada con la etiqueta Scmp, es decir, sustantivo común, masculino, plural
- **Scms Ams0 wdprevtag Scms receptor**: cambiar la etiqueta Scms, es decir, sustantivo común, masculino, singular, por Ams0, es decir, adjetivo, masculino, singular, sin grado, si la palabra actual es receptor y la anterior aparece etiquetada con la etiqueta Scms
- **Scms Ams0 wdand2bfr el transmisor**: cambiar la etiqueta Scms, es decir, sustantivo común, masculino, singular, por Ams0, es decir, adjetivo, masculino, singular, sin grado, si la palabra actual es transmisor y la segunda palabra anterior es el
- **P Scms nexttag P**: cambiar la etiqueta P, es decir, preposición, por Scms, es decir, sustantivo común, masculino, singular, si la palabra siguiente aparece etiquetada con la etiqueta P

Aprendizaje basado en transformaciones y dirigido por el error

- 1 Se toma una porción de texto no etiquetado
- 2 Se pasa a través de la fase o fases de etiquetación
- 3 Se compara la salida con el texto correctamente etiquetado, y se genera una lista de errores de etiquetación con sus correspondientes contadores.
- 4 Para cada error, se determina qué instancia concreta de la plantilla genérica de reglas produce la mayor reducción de errores.
- 5 Se aplica la regla, se calcula el nuevo conjunto de errores producidos, y se repite el proceso hasta que la reducción de errores cae por debajo de un umbral dado.



Consideraciones de eficiencia

- El etiquetador de Brill resulta considerablemente más lento que los basados en modelos probabilísticos, tanto en la fase de entrenamiento como de etiquetación
- Razón: la interacción entre las reglas, de manera que el algoritmo puede producir cálculos innecesarios
- El algoritmo de Brill podría necesitar RKn pasos elementales para etiquetar una cadena de entrada de n palabras, con R reglas aplicables en un contexto de hasta K palabras.

Ejemplo de interacción de reglas

- Etiquetas (corpus Brown):
 - VBN: verbo en participio pasado
 - VBD: verbo en tiempo pasado
 - NP: sustantivo propio
 - BEDZ: palabra `was`
 - BY: palabra `by`
 - PPS: pronombre nominativo singular en tercera persona
- VBN y VBD son las etiquetas más probables para las palabras `killed` y `shot`, respectivamente

Ejemplo de interacción de reglas

- Después de la etiquetación léxica:

(1) Chapman/NP **killed/VBN** John/NP Lennon/NP

(2) John/NP Lennon/NP was/BEDZ **shot/VBD** by/BY Chapman/NP

(3) He/PPS witnessed/VBD Lennon/NP **killed/VBN** by/BY Chapman/NP

- Errores:

en (1) **killed** etiquetada incorrectamente como verbo en participio pasado

en (2) **shot** incorrectamente etiquetada como verbo en tiempo pasado

Ejemplo de interacción de reglas

- El etiquetador contextual aplica las reglas:

VBN VBD prevtag NP (cambiar la etiqueta VBN por VBD si la etiqueta previa es NP)

VBD VBN nexttag BY (cambiar VBD por VBN si la siguiente etiqueta es BY)

- Al aplicar la primera regla la palabra `killed` que aparece en las frases (1) y (3) cambia su etiqueta VBN por VBD:

(4) Chapman/NP **killed/VBD** John/NP Lennon/NP

(5) John/NP Lennon/NP was/BEDZ **shot/VBD** by/BY Chapman/NP

(6) He/PPS witnessed/VBD Lennon/NP **killed/VBD** by/BY Chapman/NP

Ejemplo de interacción de reglas

- Al aplicar la segunda regla:

la palabra `shot` de la frase (5) cambia su etiqueta VBD por VBN

la palabra `killed` de la frase (6) vuelve a cambiar su etiqueta VBD otra vez por VBN

(7) Chapman/NP `killed/VBD` John/NP Lennon/NP

(8) John/NP Lennon/NP was/BEDZ `shot/VBN` by/BY Chapman/NP

(9) He/PPS witnessed/VBD Lennon/NP `killed/VBN` by/BY Chapman/NP