

Desambiguación de significados

ITALICA
Universidad de Sevilla
José A. Troyano

Índice

- **Introducción**
- Aspectos metodológicos
- Desambiguación supervisada
- Desambiguación basada en recursos lingüísticos
- Desambiguación no supervisada

Introducción

Definición de ambigüedad

Ambigüedad: varios significados distintos para una única palabra.

Desambiguación: determinar el significado que corresponde a una determinada ocurrencia de una palabra en función de su contexto.

Tipos de ambigüedad:

- Evidente
- No evidente
- En el etiquetado gramatical (POS tagging)

Introducción

Ambigüedad evidente

Significados de *bank*:

1. The rising ground bordering a lake, river or sea.
2. An establishment for the custody, loan exchange, or issue of money, for the extension of credit, and for facilitating the transmission of funds.

Este caso es muy claro pero no es lo usual, ya que no siempre es fácil establecer una línea de separación clara entre los distintos significados.

Introducción

Ambigüedad no evidente

Significados de *title*:

1. Name/heading of a book, statue, work of art or music.
2. Material at the start of a film.
3. The right of legal ownership (of land).
4. The document that is evidence of this right.
5. An appellation of respect attached to a person's name.
6. A written work (sinécdoque: la parte por el todo)

En una clasificación más fina el significado 1 se podría dividir.

En una clasificación más gruesa los significados 1-2-6 y 3-4 se podrían fusionar.

Introducción

Ambigüedad en el etiquetado gramatical

butter puede ser:

un nombre: bread and butter

un verbo: to butter a toast

En este caso nos interesa etiquetar adecuadamente la palabra, aunque implícitamente con ello averigüemos también el significado.

En la práctica son dos problemas distintos (desambiguación y etiquetado). Tanto por la naturaleza de los problemas como por sus soluciones.

Índice

- Introducción
- **Aspectos metodológicos**
- Desambiguación supervisada
- Desambiguación basada en recursos lingüísticos
- Desambiguación no supervisada

Aspectos metodológicos

Aprendizaje supervisado y no supervisado

Aprendizaje supervisado: se parte de un conocimiento previo, que en el caso de la desambiguación puede ser un corpus ya desambiguado, un diccionario electrónico, un thesaurus, ...

Aprendizaje no supervisado: no se parte de ninguna información previa, puede verse como un proceso de generación de clases (clustering). En el caso de la desambiguación se agruparían las apariciones de una palabra en distintos grupos según sus contextos.

Aspectos metodológicos

Pseudo-palabras

Es una forma fácil de obtener tests para los algoritmos de desambiguación a partir de corpus sin etiquetar (Gale92).

Pseudo-palabra: palabra ambigua artificial, por ejemplo *banana-door*. Se sustituye en un texto todas las apariciones de *banana* y de *door* por dicha pseudo-palabra, el resultado del algoritmo de desambiguación se contrasta con el texto original.

Aspectos metodológicos

Límites inferior y superior de rendimiento

La medida de rendimiento de un algoritmo (en %) debe ser interpretada en función de la dificultad del problema:

- Es fácil conseguir un 90% de éxito en el etiquetado gramatical de un texto inglés
- En situaciones claras de ambigüedad como *bank* podemos esperar un éxito del 95%.
- En situaciones menos claras como *title* un 65% puede ser un buen resultado (quizá una persona no lo haga mucho mejor).

En general:

Límite superior: rendimiento humano

Límite inferior: rendimiento del algoritmo más simple(engañoso?)

P.e. para una palabra con dos significados con probabilidades $p_1 = 0.1$ y $p_2 = 0.9$ podemos conseguir el 90% fácilmente

Aspectos metodológicos

Notación

<u>Símbolos</u>	<u>Significado</u>
w	Palabra ambigua
$s_1, \dots, s_{k'}, \dots, s_K$	Significados de la palabra ambigua w
$c_1, \dots, c_{j'}, \dots, c_I$	Contextos de w en un corpus
$v_1, \dots, v_{j'}, \dots, v_J$	Palabras de un contexto c

Índice

- Introducción
- Aspectos metodológicos
- **Desambiguación supervisada**
- Desambiguación basada en recursos lingüísticos
- Desambiguación no supervisada

Desambiguación supervisada
Clasificación Bayesiana

Regla de decisión:

Decide s' si $P(s'|c) > P(s_k|c)$ para $s' \neq s_k$

Cálculo de $P(s_k|c)$:

$$\begin{aligned} P(s_k|c) &= (P(c|s_k) * P(s_k)) / P(c) && \% P(c) \text{ es cte. para todo } s_k \\ &\Rightarrow (P(c|s_k) * P(s_k)) && \% \text{ aplicando logaritmos} \\ &\Rightarrow \log P(c|s_k) + \log P(s_k) \end{aligned}$$

En general:

$$s' = \arg_{s_k} \max [\log P(c|s_k) + \log P(s_k)]$$

Desambiguación supervisada
Clasificación Bayesiana ingenua

Para simplificar el cálculo de $P(c|s_k)$ asumimos que las palabras del contexto c son independientes, por tanto:

$$P(c|s_k) = P(\{v_j | v_j \text{ in } c\} | s_k) = \prod_{v_j \text{ in } c} P(v_j | s_k)$$

donde c es un multiconjunto de palabras v_j .

Tanto $P(v_j | s_k)$ como $P(s_k)$ son calculadas a través de la estimación de la máxima probabilidad:

$$P(v_j | s_k) = C(v_j, s_k) / C(s_k)$$

$$P(s_k) = C(s_k) / C(w)$$

Desambiguación supervisada

Algoritmo para la desambiguación Bayesiana

Comentario: Entrenamiento

Para todos los significados s_k de w **hacer**

Para todas las palabras v_j del vocabulario **hacer**

 calcular $P(v_j | s_k)$

Fin para

Fin para

Para todos los significados s_k de w **hacer**

 calcular $P(s_k)$

Fin para

Comentario: Desambiguación

Para todos los significados s_k de w **hacer**

 cuenta(s_k) = $\log P(s_k)$

Para todas las palabras v_j de la ventana del contexto c **hacer**

 cuenta(s_k) = cuenta(s_k) + $\log P(v_j | s_k)$

Fin para

Fin para

$s' = \arg \max_{s_k} \text{cuenta}(s_k)$

Desambiguación supervisada

Método basado en la información mutua

En este caso en lugar de utilizar un contexto amplio (que incluya varias palabras), se realiza la desambiguación mediante una sola palabra escogida cuidadosamente.

Por ejemplo para el verbo francés prendre, el objeto da suficiente información para desambiguar:

prendre une mesure => **take** a measure

prendre une décision => **make** a decision

Desambiguación supervisada Información mutua: flip-flop

Traducciones de w : t_1, \dots, t_m

Valores del indicador: x_1, \dots, x_n

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Comentario: Obtiene dos particiones para los significados e indicadores de w

encontrar una partición aleatoria $P = \{P_1, P_2\}$ de $\{t_1, \dots, t_m\}$

Mientras haya mejora **hacer**

encontrar una partición $Q = \{Q_1, Q_2\}$ de $\{x_1, \dots, x_n\}$

que maximice $I(P;Q)$

encontrar una partición de $P = \{P_1, P_2\}$ de $\{t_1, \dots, t_m\}$

que maximice $I(P;Q)$

Fin mientras

Desambiguación supervisada Información mutua: desambiguación

Por ejemplo, para *prendre*:

$\{t_1, \dots, t_m\} = \{\text{take, make, rise, speak}\}$

$\{x_1, \dots, x_n\} = \{\text{mesure, note, exemple, décision, parole}\}$

Comentario: Desambigua w dada la partición $Q = \{Q_1, Q_2\}$

determina el valor x_i del indicador de w

Si $x_i \in Q_1$ **entonces**

s' = significado 1

Si $x_i \in Q_2$ **entonces**

s' = significado 2

Fin si

Resultado del Flip-Flop

$P_1 = \{\text{take, rise}\}, P_2 = \{\text{make, speak}\}$

$Q_1 = \{\text{mesure, note, exemple}\}, Q_2 = \{\text{décision, parole}\}$

$P_1 = \{\text{take}\}, P_2 = \{\text{make, rise, speak}\}$

...

Índice

- Introducción
- Aspectos metodológicos
- Desambiguación supervisada
- **Desambiguación basada en recursos lingüísticos**
- Desambiguación no supervisada

Desambiguación basada en recursos lingüísticos

Desambiguación basada en diccionarios

Método: Aprovechar las palabras de la definición de un diccionario como indicadores para desambiguar la palabra definida.

Definiciones de *cone*:

1. a mass of ovule-bearing or pollen-bearing scales or bracts in trees of the pine family or in cycads that are arranged usually on a somewhat elongated axis.
2. something that resembles a cone in shape: as a crisp cone-shaped wafer for holding ice cream.

tree y *ice* son indicadores de los significados 1 y 2, respectivamente

Desambiguación basada en recursos lingüísticos
Algoritmo basado en diccionario

D_k es la definición de diccionario de la palabra w para cada significado s_k

E_{v_j} es la definición de la palabra v_j que aparece en el contexto c de la palabra w (v_j no se desambigua, por tanto

$E_{v_j} = \cup D_{j,i}$).

Comentario: Desambiguación dada w , el contexto c y las definiciones del diccionario

Para todos los significados s_k de w **hacer**

$\text{cuenta}(s_k) = \text{coincidencias}(D_k, \cup_{v_j \text{ in } c} E_{v_j})$

Fin para

$s' = \arg \max_{s_k} \text{cuenta}(s_k)$

Se pueden utilizar funciones más finas que *coincidencias* para realizar el contraste.

Desambiguación basada en recursos lingüísticos
Desambiguación basada en thesaurus

En un thesaurus, el significado de una palabra w se establece asociándole distintas categorías semánticas t_i .

$t(s_k)$ = Categoría del significado k -ésimo de la palabra w .

$t(v_j)$ = Conjunto de categorías t para la palabra v_j .

Comentario: Desambiguación dada w , el contexto c y las clasificaciones de un thesaurus

Para todos los significados s_k de w **hacer**

$\text{cuenta}(s_k) = \sum_{v_j \text{ in } c} \text{aparece}(t(s_k), t(v_j))$

Fin para

$s' = \arg \max_{s_k} \text{cuenta}(s_k)$

Desambiguación basada en recursos lingüísticos

Thesaurus: categorización de contextos

Puede que palabras que den mucha información no estén bien catalogadas en el thesaurus. Por ejemplo:

Navratilova da mucha información en un texto ya que nos puede decir que versa sobre tenis.

A no ser que tengamos un thesaurus específico sobre deportes es difícil que esa palabra esté contemplada.

Posible mejora: Que el algoritmo sea adaptativo, de manera que calcule nuevas categorías de todas las palabras del corpus en función de las categorías de sus contextos.

Desambiguación basada en recursos lingüísticos

Thesaurus: algoritmo adaptativo (I)

Comentario: Entrenamiento

Para todos los contextos c_i del corpus **hacer**

calcular las categorías del contexto $t(c_i)$ con un corte α

Fin para

Comentario: Se aprovechan las categorías de los contextos para calcular las categorías de las palabras

Para todas las palabras v_j del vocabulario **hacer**

calcular V_j : el conjunto de contextos en los que aparece v_j

Fin para

Para todas las categorías t_i del thesaurus **hacer**

calcular T_i : el conjunto de contextos con categoría t_i

Fin para

Para todas las palabras v_j y todas las categorías t_i **hacer**

$$P(v_j|t_i) = |V_j \cap T_i| / |\Sigma_j V_j \cap T_i|$$

Fin para

Para todas las categorías t_i **hacer**

$$P(t_i) = (\Sigma_j |V_j \cap T_i|) / (\Sigma_i \Sigma_j |V_j \cap T_i|)$$

Fin para

Desambiguación basada en recursos lingüísticos

Thesaurus: algoritmo adaptativo (II)

Una vez que tenemos $P(t(s_k))$ y $P(v_j|t(s_k))$, se aplica desambiguación Bayesiana

Comentario: Desambiguación

Para todos los significados s_k de w en el contexto c **hacer**

$$\text{cuenta}(s_k) = \log P(t(s_k)) + \sum_{v_j \text{ in } c} \log P(v_j|t(s_k))$$

Fin para

$$s' = \arg \max_{s_k} \text{cuenta}(s_k)$$

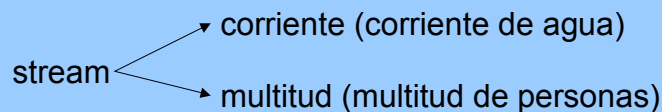
La diferencia está, por tanto, en el cálculo de esas probabilidades:

- El método basado en corpus etiquetado las calcula en función de las palabras que rodean a un significado.
- En el método basado en thesaurus se incorpora un factor más que es la categoría del contexto (visión más global)

Desambiguación basada en recursos lingüísticos

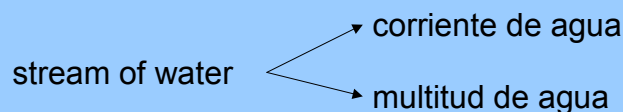
Desambiguación basada en traducción

Se utilizan las traducciones de una palabra a un segundo lenguaje como etiquetas semánticas:



Se necesita una relación R , en este caso la relación se establece con “de” y otra palabra (agua, personas)

El método se basa en traducir la palabra ambigua (stream) con sus dos significados y comprobar si la frase traducida aparece en un corpus del segundo lenguaje:



Desambiguación basada en recursos lingüísticos

Desambiguación basada en traducción: algoritmo

S es el corpus en un segundo lenguaje.

$T(s_k)$ es el conjunto de posibles traducciones del significado s_k

$T(v)$ es el conjunto de posibles traducciones de v

Comentario: Desambiguación de la palabra w que aparece relacionada con v a través de $R(w,v)$

Para todos los significados s_k de w **hacer**

$\text{cuenta}(s_k) = |\{c \in S \mid \exists w' \in T(s_k), v' \in T(v) : R(w',v') \in c\}|$

Fin para

$s' = \arg \max_{s_k} \text{cuenta}(s_k)$

Para que sea efectivo no debe reproducirse la ambigüedad en el segundo lenguaje. Por ejemplo, *interest* e *interés* significan atención y rédito en español e inglés.

Desambiguación basada en recursos lingüísticos

Un significado por discurso y colocación

Un significado por discurso: Habitualmente el significado de las distintas apariciones de una palabra ambigua en un texto es siempre el mismo.

Un significado por colocación: Las palabras cercanas determinan en gran medida el significado de una palabra ambigua. Ya lo vimos en el método de la información mutua.

La importancia de una colocación f frente a dos significados s_{k_i} y s_{k_j} se puede medir con la siguiente fórmula:

$$\frac{P(s_{k_i}|f)}{P(s_{k_j}|f)}$$

Desambiguación basada en recursos lingüísticos

Un significado por discurso y colocación: algoritmo(I)

F_k es el conjunto de colocaciones características del significado s_k

E_k es el conjunto de contextos donde se asigna el significado s_k a w

Comentario: Inicialización

Para todos los significados s_k de w **hacer**

$F_k =$ colocaciones de s_k en un diccionario

$E_k = \emptyset$

Fin para

Desambiguación basada en recursos lingüísticos

Un significado por discurso y colocación: algoritmo(II)

Comentario: Un sentido por colocación (idea similar al Flip-Flop)

Mientras cambie algún E_k en la última iteración **hacer**

Para todos los significados s_k de w **hacer**

$E_k = \{c_i \mid \exists f_m : f_m \in c_i \wedge f_m \in F_k\}$

Fin para

Para todos los significados s_k de w **hacer**

$F_k = \{f_m \mid \forall n \neq k P(s_k|f_m)/P(s_n|f_m) > \alpha\}$ % se calcula P sobre E_k

Fin para

Fin mientras

Comentario: Un sentido por discurso

Para todos los documentos d_m **hacer**

determinar el significado mayoritario s_k de w en d_m

asigna s_k a todas las apariciones de w en d_m

Fin para

Índice

- Introducción
- Aspectos metodológicos
- Desambiguación supervisada
- Desambiguación basada en recursos lingüísticos
- **Desambiguación no supervisada**

Desambiguación no supervisada

Clustering

Estrictamente hablando la desambiguación no supervisada no es posible ya que se necesita un conocimiento mínimo de los significados.

Es más correcto hablar de etiquetado de significados.

Se pueden generar K clusters para diferenciar K significados:

- Es apropiado para la ambigüedad no evidente ya que puede sacar a la luz diferencias sutiles.
- No es apropiado para significados poco frecuentes ya que estos quedan aislados

Desambiguación no supervisada
Algoritmo EM(I)

Modelo μ : está compuesto por las probabilidades $P(v_j|s_k)$ y $P(s_k)$ para toda palabra del vocabulario v_j y todo significado s_k de la palabra ambigua w .

Probabilidad del modelo $I(C|\mu)$: Mide la bondad del modelo y se calcula con la siguiente fórmula:

$$I(C|\mu) = \sum_{i=1..I} \log \left(\sum_{k=1..K} P(c_i|s_k)P(s_k) \right)$$

Coefficiente h_{ik} : Probabilidad posterior de s_k y c_i . Se calcula con la siguiente fórmula:

$$h_{ik} = \frac{P(c_i|s_k)}{\sum_{k=1..K} P(c_i|s_k)}$$

Desambiguación no supervisada
Algoritmo EM(II)

Comentario: Inicialización

inicializar aleatoriamente $P(s_k)$ y $P(v_j|s_k)$
calcular la probabilidad del modelo $I(C|\mu)$

Mientras mejore sensiblemente $I(C|\mu)$ en la última iteración **hacer**

Comentario: Paso E (Estimación)

calcular los coeficientes h_{ik}

Comentario: Paso M (estimación de Máxima probabilidad)

recalcular las probabilidades $P(v_j|s_k)$ con la fórmula

$$P(v_j|s_k) = \frac{\sum_{\{c_i:v_j \in c_i\}} h_{ik}}{\sum_{k=1..K} \sum_{\{c_i:v_j \in c_i\}} h_{ik}}$$

recalcular las probabilidades $P(s_k)$ con la fórmula

$$P(s_k) = \frac{\sum_{i=1..I} h_{ik}}{\sum_{k=1..K} \sum_{i=1..I} h_{ik}}$$

Fin mientras

Desambiguación no supervisada
Algoritmo EM(III)

Cuanto mayor sea K (el número de *clusters*), mayor será la probabilidad del modelo $I(C|\mu)$. Merecerá la pena añadir más significados mientras la mejora de $I(C|\mu)$ sea significativa.

Una vez calculadas las probabilidades $P(v_j|s_k)$ y $P(s_k)$, se aplica desambiguación bayesiana:

Comentario: Desambiguación

Para todos los significados s_k de w en el contexto c **hacer**

$$\text{cuenta}(s_k) = \log P(s_k) + \sum_{v_j \text{ in } c} \log P(v_j|s_k)$$

Fin para

$$s' = \arg \max_{s_k} \text{cuenta}(s_k)$$