Extracción de Features en aplicaciones específicas

Taller de Procesamiento de Señales

Agenda

1. Procesamiento de Lenguaje Natural

2. Procesamiento de Sonido

¿Como convertir un texto en un vector?

One-hot Encoding

Dado un vocabulario $V = \{\omega_1, \cdots, \omega_{|V|}\}$, se puede convertir cada palabra en un vector *one-hot*.

$$\omega_i \longrightarrow \boxed{\begin{array}{c}\text{One-Hot}\\ \text{Encoding}\end{array}} \longrightarrow x_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{|V|}$$

i-ésima posición

¿Como convertir un texto en un vector?

One-hot Encoding

Dado un vocabulario $V = \{\omega_1, \cdots, \omega_{|V|}\}$, se puede convertir cada palabra en un vector *one-hot*.

$$\omega_i \longrightarrow \boxed{\text{One-Hot Encoding}} \longrightarrow x_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{|V|}$$

i-ésima posición

Bolsa de palabras

Para vectorizar un documento $f(x_1, \cdots, x_n)$, la manera más simple es *bolsa de palabras*: $f(x_1, \cdots, x_n) = x_1 + \cdots + x_n$.

Procesamiento del Lenguaje Natural

Vectorizaciones Sofisticadas

En la práctica suelen utilizarse representaciones pre-entrenadas (ej. FastText).

Procesamiento del Lenguaje Natural

Vectorizaciones Sofisticadas

En la práctica suelen utilizarse representaciones pre-entrenadas (ej. FastText).

Normalizaciones de NLP

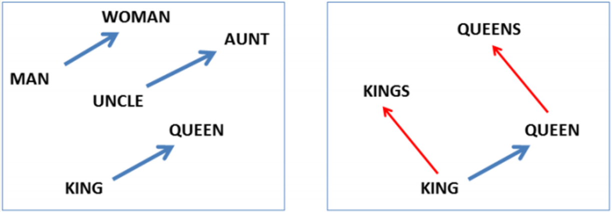
- Eliminar caracteres raros e inusuales
- Convertir todo a minúsculas
- Eliminar palabras no informativas (stop words)
- Descartar las palabras poco observadas
- Descartar las palabras más comunes
- Lemmatization (significado)
- Stemming (quedarse con la raíz)

Term Frequency - Inverse Document Frequency

Transformación tf-idf

Medida numérica que expresa cuán relevante es una palabra para un documento dentro de un dataset. El tf-idf para un término t de un documento d perteneciente a una colección de n documentos es $\text{tf-idf}(t,d) = \text{tf}(t,d) \cdot \text{idf}(t)$. El primer factor $\text{tf}(t,d) = \frac{\#(t \in d)}{\#(d)}$ es la cantidad de veces que aparece el término t en el documento d dividido la cantidad de términos que aparecen en el documento d. El segundo factor $\text{idf}(t) = 1 - \log\left(\frac{\text{df}(t)}{n}\right)$, donde $\text{df}(t)$ es la cantidad de documentos que poseen el término t en su interior.

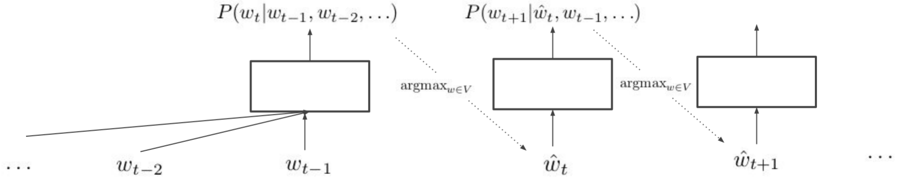
Word Vectors + PCA



(Mikolov et al., NAACL HLT, 2013)

vector(KINGS) - vector(KING) + vector(QUEEN) = vector(QUEENS)

Síntesis de texto



Outline

1 Procesamiento de Lenguaje Natural

2 Procesamiento de Sonido

Coeficientes de Frecuencia Cepstrum en escala Mel

MFCC

Los MFCC son características muy utilizadas en procesamiento de sonido, las cuáles combinan tres conceptos:

- **Ventaneo:** Se suele trabajar con ventanas del orden de los 25mseg, ya que en estos intervalos de tiempo las señales de habla son cuasi-estacionarias.
- **Escala de Mel:** Respuesta en frecuencia del oído. Básicamente se filtra una señal de forma afín a como lo hace nuestro oído (mediante un banco de filtros).
- **Coeficientes de Frecuencia Cepstrum:** Los Cepstrum se definen como la anti transformada del logaritmo del módulo de la transformada de Fourier.

Transformada Cepstrum

Tiene la virtud de transformar convoluciones en sumas:

$$\begin{aligned} \mathcal{C}\left(a[n] * b[n]\right) &= \mathcal{F}^{-1}\left[\log|A(e^{j\Omega})B(e^{j\Omega})|\right] \\ &= \mathcal{F}^{-1}\left[\log|A(e^{j\Omega})|\right] + \mathcal{F}^{-1}\left[\log|B(e^{j\Omega})|\right] \\ &= \mathcal{C}\left(a[n]\right) + \mathcal{C}\left(b[n]\right) \end{aligned}$$

Transformada Cepstrum

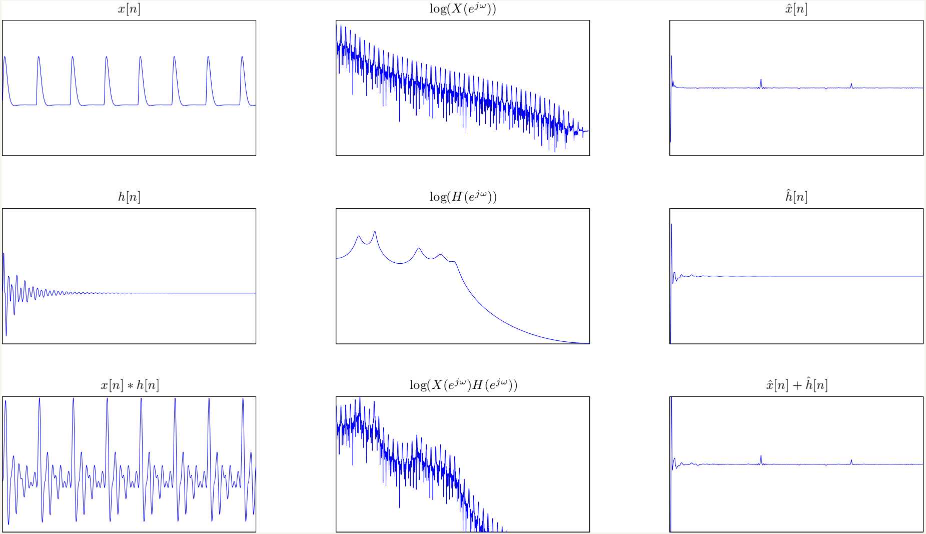
Tiene la virtud de transformar convoluciones en sumas:

$$\begin{aligned}
\mathcal{C}\left(a[n] * b[n]\right) &= \mathcal{F}^{-1}\left[\log\left|A(e^{j\Omega})B(e^{j\Omega})\right|\right] \\
&= \mathcal{F}^{-1}\left[\log\left|A(e^{j\Omega})\right|\right] + \mathcal{F}^{-1}\left[\log\left|B(e^{j\Omega})\right|\right] \\
&= \mathcal{C}\left(a[n]\right) + \mathcal{C}\left(b[n]\right)
\end{aligned}$$

Robusto a ruido

Normalizar (en media y varianza) elimina ruidos convolucionales, ya que equivale a sumar un sesgo (constante). Se ha demostrado experimentalmente que normalizar reduce significativamente también el ruido aditivo.

Transformada Cepstrum



Coeficientes MFCC dinámicos

Δ-MFCC y $\Delta\Delta$-MFCC

Los MFCC son conocidos como coeficientes estáticos ya que poseen información de la señal de habla sólo en la ventana actual. Para incorporar información acerca de la evolución temporal de los MFCC se incluyen los coeficientes dinámicos, es decir, las primeras y segundas diferencias entre coeficientes de ventanas consecutivas (velocidades y aceleraciones).

$$\Delta y_i[j] = \frac{y_i[j+1] - y_i[j-1]}{2}, \qquad \Delta\Delta y_i[j] = \frac{\Delta y_i[j+1] - \Delta y_i[j-1]}{2}$$

Coeficientes MFCC dinámicos

Δ-MFCC y $\Delta\Delta$-MFCC

Los MFCC son conocidos como coeficientes estáticos ya que poseen información de la señal de habla sólo en la ventana actual. Para incorporar información acerca de la evolución temporal de los MFCC se incluyen los coeficientes dinámicos, es decir, las primeras y segundas diferencias entre coeficientes de ventanas consecutivas (velocidades y aceleraciones).

$$\Delta y_i[j] = \frac{y_i[j+1] - y_i[j-1]}{2}, \qquad \Delta\Delta y_i[j] = \frac{\Delta y_i[j+1] - \Delta y_i[j-1]}{2}$$

¿Predictor o Muestra?

Si mi señal se procesa con n ventanas, y para cada una extraigo d coeficcientes MFCC, tendré finalmente n muestras de dimensión d cada una. Si incluyo los Δ-MFCC y $\Delta\Delta$-MFCC tendré n muestras de dimensión $3d$ cada una (d predictores estáticos, d velocidades y d aceleraciones).