



Procesamiento de Lenguaje Natural

Vectores Semánticos

Mauricio Toledo-Acosta
mauricio.toledo@unison.mx

Departamento de Matemáticas
Universidad de Sonora



Section 1

Introducción



Section 2

Ejemplos ilustrativos





Modelo BOW

Procesamiento
de Lenguaje
Natural

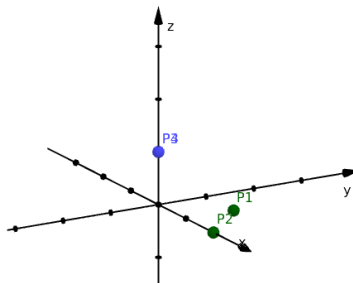
Introducción

Ejemplos
ilustrativos

LSA

Texto	Europa	cambios	arroz	pescado
1	1	1	0	0
2	2	0	0	0
3	0	0	1	0
4	0	0	1	1

El modelo BOW asigna a cada documento el vector correspondiente a la fila. El vector de cada palabra es su columna.





Ejercicio BOW

Documentos:

- El gato come ratones y juega con el perro. El perro duerme al lado y come.
- El gato come pescado.
- El perro ladra fuerte y come.
- El código tiene un error.
- El programa ejecuta código.

Palabras consideradas:

- | | | |
|-----------|-----------|------------|
| • gato | • duerme | • código |
| • come | • lado | • error |
| • ratones | • pescado | • programa |
| • juega | • ladra | • ejecuta |
| • perro | • fuerte | |



Section 3

LSA



- **Definición:** LSA (Latent Semantic Analysis) es una técnica de procesamiento de lenguaje natural.
- **Contexto:** Usado en Topic Modelling para descubrir temas en textos.
- **Objetivo:** Identificar temas ocultos en un conjunto de documentos.



- **Matriz Término-Documento:** Representación numérica de textos.
- **SVD:** Reducción de dimensionalidad para capturar relaciones semánticas.
- **Espacio semántico latente:** Representación compacta de palabras y documentos.



Matriz Término-Documento

Procesamiento
de Lenguaje
Natural

Introducción

Ejemplos
ilustrativos

LSA

- **Estructura:** Filas = palabras, columnas = documentos.
- **Valores:** Frecuencia de términos (TF) o TF-IDF.
- **Ejemplo:**

$$\begin{bmatrix} 2 & 0 & 1 \\ 1 & 3 & 0 \\ 0 & 2 & 1 \end{bmatrix}$$



SVD (Descomposición en Valores Singulares)

Procesamiento
de Lenguaje
Natural

Introducción

Ejemplos
ilustrativos

LSA

- **Definición:** Descompone una matriz en U , Σ , V^T .
- **Propósito:** Reducir la dimensionalidad conservando la estructura semántica.
- **Ecuación:**

$$A = U\Sigma V^T$$



- **Concepto:** Representación de palabras y documentos en un espacio de menor dimensión.
- **Ventaja:** Captura relaciones semánticas entre términos y documentos.
- **Ejemplo:** Palabras como "coche" y "automóvil" estarán cerca.



Proceso de LSA

Procesamiento
de Lenguaje
Natural

Introducción

Ejemplos
ilustrativos

LSA

- **Preprocesamiento:** Tokenización, eliminación de stopwords, etc.
- **Matriz Término-Documento:** Creación y ponderación (TF-IDF).
- **SVD:** Aplicación y reducción de dimensionalidad.
- **Interpretación:** Identificación de temas latentes.



Preprocesamiento

Procesamiento de Lenguaje Natural

Introducción

Ejemplos ilustrativos

LSA

- **Tokenización:** Dividir textos en palabras o frases.
- **Eliminación de stopwords:** Quitar palabras comunes (ej: "el", "de").
- **Lematización:** Reducir palabras a su forma base (ej: "corriendo" → "correr").



Creación de la matriz término-documento

- **Paso 1:** Contar la frecuencia de cada término en cada documento.
- **Paso 2:** Aplicar TF-IDF para ponderar términos importantes.
- **Resultado:** Matriz numérica que representa los textos.



Aplicación de SVD

Procesamiento
de Lenguaje
Natural

Introducción

Ejemplos
ilustrativos

LSA

- **Entrada:** Matriz término-documento.
- **Proceso:** Descomposición en U, Σ, V^T .
- **Salida:** Matrices reducidas que capturan la estructura semántica.



- **Componentes:** Cada columna de U y V^T representa un tema.
- **Relaciones:** Palabras y documentos se agrupan según su similitud semántica.
- **Ejemplo:** Tema 1 = deportes, Tema 2 = tecnología.



Ventajas de LSA

Procesamiento de Lenguaje Natural

Introducción

Ejemplos ilustrativos

LSA

- Captura relaciones semánticas entre palabras.
- Reduce el ruido en grandes conjuntos de datos.
- Simple y fácil de implementar.



Limitaciones de LSA

Procesamiento de Lenguaje Natural

Introducción

Ejemplos ilustrativos

LSA

- Dificultad para interpretar temas explícitamente.
- No captura relaciones no lineales entre palabras.
- Depende del preprocesamiento y parámetros.



Aplicaciones de LSA

Procesamiento
de Lenguaje
Natural

Introducción

Ejemplos
ilustrativos

LSA

- Recuperación de información.
- Clasificación de textos.
- Análisis de sentimientos.
- Recomendación de contenido.



Comparación con otras técnicas

Procesamiento
de Lenguaje
Natural

Introducción

Ejemplos
ilustrativos

LSA

- **LSA:** Basado en álgebra lineal (SVD).
- **LDA:** Basado en probabilidad (distribuciones de temas).
- **Uso:** LSA para simplicidad, LDA para interpretabilidad.



- **Dataset:** Textos de noticias.
- **Proceso:** Aplicación de LSA y visualización de temas.
- **Resultados:** Grupos de palabras y documentos relacionados.



Conclusión

Procesamiento
de Lenguaje
Natural

Introducción

Ejemplos
ilustrativos

LSA

- LSA es una herramienta poderosa para Topic Modelling.
- Captura relaciones semánticas en textos.
- Invitación a explorar más sobre el tema.