

A Priori

Minería de Datos (CC3074) - 2026

A Priori

Semestre 01, 2026

Contexto

- No siempre interesa predecir
- A veces interesa entender relaciones
- Especialmente en datos transaccionales

Resultado

- Reglas de asociación
- Co-ocurrencia de elementos en una colección
- Del tipo: $A \rightarrow B$
- Muchas reglas, no todas útiles

Probabilidades

Calcula la probabilidad de un elemento de estar presente en un conjunto dado que otro u otros elementos estén presentes.

Reto

- Apriori genera candidatos
- El analista decide calidad
- La dificultad no es calcular, es interpretar

Escenario de pruebas

| Transacción | Ítems | |-----|-----| | Compra 1 | Laptop, Monitor, Teclado | | Compra 2 | Office, Laptop, Mouse, Teclado | | Compra 3 | Laptop, Mouse, Monitor, Teclado |

Conjuntos frecuentes

Identificar los conjuntos de elementos más frecuentes.

Compra 1

- {Laptop, Teclado}
- {Laptop, Monitor}
- {Monitor, Teclado}
- {Laptop, Monitor, Teclado}

Compra 2

- {Office, Laptop}
- {Office, Mouse}
- {Office, Teclado}
- {Laptop, Mouse}

- {Laptop, Teclado}
- {Mouse, Teclado}
- {Office, Laptop, Mouse}
- {Office, Laptop, Teclado}
- {Office, Mouse, Teclado}
- {Laptop, Mouse, Teclado}

Compra 3

- {Laptop, Mouse}
- {Laptop, Monitor}
- {Laptop, Teclado}
- {Mouse, Monitor}
- {Mouse, Teclado}
- {Monitor, Teclado}
- {Laptop, Mouse, Monitor}
- {Laptop, Mouse, Teclado}
- {Laptop, Monitor, Teclado}
- {Mouse, Monitor, Teclado}

Criterio de selección

Para determinar qué conjuntos de elementos repetidos son significativos usamos una métrica llamada soporte.

$$\text{Soporte}(X) = \frac{\text{Número de transacciones con } X}{n}$$

$$\{Laptop, Teclado\} = 3/3 = 1.00$$

$$\{Laptop, Monitor\} = 2/3 = 0.66$$

$$\{\text{Monitor, Teclado}\} = 2/3 = 0.66$$

$$\{\text{Laptop, Monitor, Teclado}\} = 2/3 = 0.66$$

$$\{\text{Laptop, Mouse}\} = 2/3 = 0.66$$

$$\{\text{Mouse, Teclado}\} = 2/3 = 0.66$$

$$\{\text{Laptop, Mouse, Teclado}\} = 2/3 = 0.66$$

$$\text{Todos los demás} = 1/3 = 0.33$$

La tolerancia depende del dominio de los datos! En ciertos casos vamos a necesitar una tolerancia más relajada o más estricta.

Reglas de asociación

Definir las reglas de co-ocurrencia de los elementos.

Antecedente -> Consecuente

La cardinalidad es importante al trabajar con nuestros conjuntos. Queremos evitar hacer análisis de más por lo que vamos a eliminar algunos conjuntos.

Criterio de selección

No existe la mejor regla absoluta

Para determinar qué reglas si son relevantes podemos usar 3 métricas para evaluar.

- Soporte (otro)
- Confianza
- Lift

Soporte(A->C)

Qué tal frecuente es la regla

Por lo general se acepta un nivel de tolerancia ≥ 0.6

$$\text{Soporte}(A \rightarrow C) = \frac{\text{Conteo}(A \cup C)}{n}$$

Confianza(A->C)

Qué tanto aparece el consecuente en las trasacciones que contienen al antecedente

Por lo general se acepta un nivel de tolerancia ≥ 0.6

$$\text{Confianza}(A \rightarrow C) =$$

$$\frac{\text{Conteo}(\text{Antecedente} \cup \text{Consecuente})}{\text{Conteo}(\text{Antecedente})}$$

Lift(A->C)

Fuerza de la regla sobre su ocurrencia aleatoria.

Un valor menor a 1 indica una mala regla. Entre más grande sea el número, mejor.

$$\text{Lift}(A \rightarrow C) =$$

$$\frac{\text{Soporte}(\text{Antecedente} \rightarrow \text{Consecuente})}{\text{Soporte}(\text{Antecedente}) \ast \text{Soporte}(\text{Consecuente})}$$