Aprendizaje no supervisado

VC10: Análisis de transacciones y reglas de asociación

Félix José Fuentes Hurtado
felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia

Reglas de asociación

Reglas de asociación

$$\text{leche} \rightarrow \text{galletas}$$

$$\text{leche} \wedge \text{cereales} \rightarrow \text{galletas}$$

Métricas

Soporte (support) de un (sub)conjunto de ítems $X \subseteq I$

Proporción de transacciones del conjunto S en las que aparece el subconjunto de ítems X:

$$soporte(X; S) = \frac{|\{T \in S : X \subseteq T\}}{|S|}$$

Se puede entender como la probabilidad conjunta (marginal) de X.

El **soporte de una regla** de asociación $X \rightarrow Y$: soporte del conjunto resultante de la unión de antecedente y consecuente,

$$soporte(X \rightarrow Y; S) = soporte(X \cup Y; S)$$

Métricas

Confianza (confidence) de una regla de asociación $X \rightarrow Y$

Frecuencia con la que se cumple la regla de asociación:

$$confianza(X \rightarrow Y; S) = \frac{soporte(X \cup Y; S)}{soporte(X; S)}$$

La confianza tiende a 1 a medida que se observa con mayor frecuencia Y cada vez que aparece X.
Se puede entender como la precisión de la regla o la probabilidad condicionada de Y dado X.

Métricas

Mejora (lift) de una regla de asociación $X \rightarrow Y$

Ratio del soporte como indicador de la creencia de que la regla pueda ser producto del azar:

$$mejora(X \rightarrow Y; S) = \frac{soporte(X \cup Y; S)}{soporte(X; S) \cdot soporte(Y; S)}$$

donde según el valor de $mejora(X \rightarrow Y; S) =$

- 1 indica que ambos subconjuntos de ítems se relacionan al azar
- > 1 indica cierto grado de co-ocurrencia
- < 1 indica complementariedad (cuando se observa un subconjunto, no se observa el otro)

Búsqueda de subconjuntos frecuentes

Algoritmo APRIORI

Método de búsqueda en anchura de subconjuntos de elementos frecuentes

Intuición: un conjunto sólo puede ser frecuente si todos sus subconjuntos también lo son

Se reduce el espacio de búsqueda y se aborda el problema de manera iterativa-aglomerativa

Definiciones:

- **Itemset**: conjunto de elementos

- **Itemset frecuente**: conjunto de elementos que aparece en al menos ϵ transacciones del conjunto de referencia S

- k-**itemset**: conjunto de k elementos

Búsqueda de subconjuntos frecuentes

Algoritmo APRIORI

Dado un umbral de soporte ϵ, se seleccionan los 1-itemsets (items individuales) que superan el ϵ

Se buscan iterativamente los k-itemsets con $k = \{2, 3, \dots\}$

- Para que un k-itemset I sea frecuente, los k diferentes $(k-1)$-itemsets subconjuntos de I deben ser frecuentes

El algoritmo Apriori devuelve un conjuntos de itemsets frecuentes. A partir de estos, se pueden construir reglas de asociación.

Algoritmo APRIORI

Algoritmo Apriori

Recibe: Conjunto de transacciones, $S = \{T_1, T_2, \dots, T_n\}$; Umbral de soporte, ϵ

---

1. $L_1 \leftarrow$ Todos los 1-*itemsets* (dado ϵ)

2. Para $k = 2, 3, \dots$

   2.1. Se crea un conjunto de k-*itemsets* candidatos a partir de los $(k-1)$-*itemsets* fuertes obtenidos en el paso anterior:

$$C_k \leftarrow \left\{T = T^{k-1} \cup \{i\} : \left(T^{k-1} \in L_{k-1}\right) \wedge \left(i \notin T^{k-1}\right) \wedge \left(\forall j \in T, (T \setminus \{j\}) \in L_{k-1}\right)\right\}$$

   2.2. Contar las apariciones de los k-*itemsets* candidatos de C_k en el conjunto de transacciones S

   2.3. Filtrar los k-*itemsets* de C_k que son realmente fuertes o frecuentes:

$$L_k = \{T \in C_k : soporte(T) > \epsilon\}$$

Parar si $L_k = \emptyset$

Devuelve: k-*itemsets* frecuentes, para todo k

Algoritmo APRIORI

Problemas

- El listado de todos los posibles candidatos en cada paso, es un procedimiento exhaustivo (coste computacional).

  Es habitual buscar los itemsets candidatos solamente como una combinación de los itemsets frecuentes de la iteración anterior $(k-1)$

- Recorre múltiples veces el conjunto de transacciones de referencia S

  Se puede ir reduciendo el conjunto de transacciones S (no se encontrarán itemsets de tamaño k en una transacción que no de tamaño $k-1$).

Aprendizaje no supervisado

VC10: Análisis de transacciones y reglas de asociación

Félix José Fuentes Hurtado
felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia