



Universidad Autónoma de Nuevo León

Facultad de Ciencias Físico Matemáticas



Semestre: 7

Materia:
Minería de datos

Maestra:
Mayra Cristina Berrones Reyes

Alumnos :

Andrea López Solís	#1822031
Daniela Govea Serna	#1722714
Francisco García Sánchez Armáss	#1816358
Jesús Eduardo Valencia González	#1630606
Karyme Mayela Gauna Rodríguez	#1819032

Grupo: 003
Aula: AVI2

San Nicolas de los Garza Nuevo León a 18 de septiembre del 2020

Ejercicio de practica

Se procede a identificar los *itemsets* frecuentes y, a partir de ellos, crear reglas de asociación.

Transacción
{A, B, C, D}
{A, B, D}
{A, B}
{B, C, D}
{B, C}
{C, D}
{B, D}

Para este problema se considera que un *item* o *itemset* es frecuente si aparece en un mínimo de 3 transacciones, es decir, su soporte debe de ser igual o superior a $3/7 = 0.43$. Se inicia el algoritmo identificando todos los *items* individuales (*itemsets* de un único *item*) y calculando su soporte.

Itemset (k=1)	Ocurrencias	Soporte
{A}	3	0.43
{B}	6	0.86
{C}	4	0.57
{D}	5	0.71

A continuación, se generan todos los posibles *itemsets* de tamaño $k = 2$ que se pueden crear con los *itemsets* que han superado el paso anterior y se calcula su soporte.

Item k=2

Los *itemsets* {A, B}, {B, C}, {B, D} y {C, D} superan el límite (≥ 0.43) de soporte, por lo que son frecuentes. Los *itemsets* {A, C} y {A, D} no superan el soporte mínimo (≤ 0.43) por lo que se descartan.

Itemset (k=2)	Ocurrencias	Soporte
✓ {A, B}	3	0.43
{A, C}	1	0.14
{A, D}	2	0.29
✓ {B, C}	3	0.43
✓ {B, D}	4	0.57
✓ {C, D}	3	0.43

Item k=2

Se repite el proceso, esta vez creando *itemsets* de tamaño $k = 3$.

Itemset (k=2)	Ocurrencias	Soporte
{A, B}	3	0.43
{B, C}	3	0.43
{B, D}	4	0.57
{C, D}	3	0.43

Item k=3

Los *itemsets* {A, B, C}, {A, B, D} y {C, D, A} contienen subconjuntos infrecuentes, por lo que son descartados. Para los restantes se calcula su soporte.

Itemset (k=3)	Ocurrencias	Soporte
{A, B, C}	0	0
{A, B, D}	1	0.14
{B, C, D}	2	0.29
{C, D, A}	0	0

El *items* {B, C, D} no supera el soporte mínimo por lo que se considera infrecuente. Al no haber ningún nuevo *itemset* frecuente, se detiene el algoritmo.

Como resultado de la búsqueda se han identificado los siguientes *itemsets* frecuentes:

Itemset frecuentes
{A, B}
{B, C}
{B, D}
{C, D}

De todas las posibles reglas, únicamente:

- {C} => {D}
- {C} => {B}
- {A} => {B}

superan el límite de confianza.

Reglas	Confianza	Confianza
{A} => {B}	$\text{soporte}\{A, B\} / \text{soporte } \{A\}$	$0.43 / 0.43 = 1$
{B} => {A}	$\text{soporte}\{A, B\} / \text{soporte } \{B\}$	$0.43 / 0.86 = 0.5$
{B} => {C}	$\text{soporte}\{B, C\} / \text{soporte } \{B\}$	$0.43 / 0.86 = 0.5$
{C} => {B}	$\text{soporte}\{B, C\} / \text{soporte } \{C\}$	$0.43 / 0.57 = 0.75$
{B} => {D}	$\text{soporte}\{B, D\} / \text{soporte } \{B\}$	$0.43 / 0.86 = 0.5$
{D} => {B}	$\text{soporte}\{B, D\} / \text{soporte } \{D\}$	$0.43 / 0.71 = 0.6$
{C} => {D}	$\text{soporte}\{C, D\} / \text{soporte } \{C\}$	$0.43 / 0.57 = 0.75$
{D} => {C}	$\text{soporte}\{C, D\} / \text{soporte } \{D\}$	$0.43 / 0.71 = 0.6$