



Pecotche Andres
Lucas Carballo

Minería de datos usando sistemas inteligentes

Práctica 5, parte 2 - Reglas de asociación

Tabla enfermedad.xlsx:

id	Antecedentes	Adulto	Trabajo pesado	Enfermedad
1	1	0	1	1
2	0	1	1	1
3	0	1	1	1
4	0	1	0	0
5	0	1	0	0
6	1	1	0	1
7	0	0	1	0
8	0	1	1	1
9	0	0	0	0
10	0	1	0	0

3.b)

En base a los datos de la tabla enfermedad.xlsx, calcule manualmente el soporte, interés, y confianza de las siguientes:

Regla	Soporte	cobertura	confianza	Interés
Enfermedad -> Antecedentes	2/10	5/10	2/5	2
Enfermedad -> adulto	4/10	5/10	4/5	8/7
adulto -> enfermedad	4/10	7/10	4/7	8/7
Trabajo pesado -> adulto	3/10	5/10	3/5	6/7

El soporte de $X \rightarrow Y$ en una base de datos D está dado por:

$$Sop(X \Rightarrow Y) = \frac{|X \cup Y|}{|D|}$$

El soporte de la regla Enfermedad -> antecedentes será:

|Enfermedad U Antecedentes| = 2 (en dos ocasiones se da que si enfermedad = 1, antecedentes = 1 también)

|D| = 10

Por tanto el soporte es **2/10**.

Los demás soportes para las demás reglas:

Enfermedad -> Adulto = **4/10**

Adulto -> enfermedad = **4/10** (el orden de las reglas es indistinto para los soportes)

Trabajo pesado -> adulto = **3/10**

La cobertura de $X \rightarrow Y$ en una base de datos D está dada por:

$$Cob(X \Rightarrow Y) = \frac{|X|}{|D|}$$

La cobertura de la regla Enfermedad \rightarrow antecedentes será:

$|Enfermedad| / |D| = 5/10$ (esto es equivalente la prob(enfermedad)).

Las demás coberturas son:

Enfermedad \rightarrow Adulto = prob (enfermedad) = **5/10**

Adulto \rightarrow enfermedad = prob (adulto) = **7/10**

Trabajo pesado \rightarrow adulto = **5/10**

La confianza de $X \rightarrow Y$ en una base de datos D está dada por:

$$Conf(X \Rightarrow Y) = \frac{Sop(X \Rightarrow Y)}{Sop(X)} = \frac{|X \cup Y|}{|X|}$$

Esto también es equivalente a soporte / cobertura.

La confianza de la regla Enfermedad \rightarrow antecedentes será:

$|Enfermedad \cup antecedentes| = 2$

$|X| = 5$

Confianza enfermedad \rightarrow antecedentes: 2/5

Las demás confianzas serán:

Enfermedad \rightarrow Adulto = **4/5**

Adulto \rightarrow enfermedad = **4/7**

Trabajo pesado \rightarrow adulto = **3/5**

El interés de una regla $A \rightarrow B$ está dado por:

$$Interes(A \rightarrow B) = \frac{Sop(A \rightarrow B)}{Sop(A) * Sop(B)}$$

Si el interés es >1 , se dice que hay una dependencia positiva (A incide positivamente en B) y la regla es interesante. Por el contrario, si es 1 hay dependencia negativa y la regla no tiene interés.

El interés para la regla enfermedad \rightarrow antecedentes será:

$|Sop(enfermedad \rightarrow antecedentes)| = 2/10$

$|Sop(enfermedad)| = 5/10$

$|Sop(antecedentes)| = 2/10$

$$\text{Interés(enfermedad} \rightarrow \text{antecedentes)} = (2/10) / (5/10 * 2/10) = 2$$

Interés (enfermedad -> adulto):

$$|\text{Sop}(\text{ enfermedad} \rightarrow \text{adulto})| = 4/10$$

$$|\text{Sop}(\text{enfermedad})| = 5/10$$

$$|\text{Sop}(\text{adulto})| = 7/10$$

$$\text{Interés (enfermedad} \rightarrow \text{adulto)} = 8/7 = 1,1428 \text{ (regla de interés).}$$

Interés (adulto -> enfermedad):

$$|\text{Sop}(\text{ adulto} \rightarrow \text{enfermedad})| = 4/10$$

$$|\text{Sop}(\text{adulto})| = 7/10$$

$$|\text{Sop}(\text{enfermedad})| = 5/10$$

$$\text{Interés (adulto} \rightarrow \text{enfermedad)} = 8/7$$

Interés (Trabajo pesado -> adulto):

$$|\text{Sop}(\text{Trabajo pesado} \rightarrow \text{adulto})| = 3/10$$

$$|\text{Sop}(\text{trabajo pesado})| = 5/10$$

$$|\text{Sop}(\text{adulto})| = 7/10$$

$$\text{Interés (Trabajo pesado} \rightarrow \text{adulto)} = 6/7$$

1. *¿Por qué es interesante la regla Enfermedad → Antecedentes, si bien su soporte y confianza son bajos?*

Porque si bien el soporte de la regla es bajo, el soporte de los atributos de forma independiente es bajo también, de hecho el soporte de los antecedentes, tiene el mismo número de apariciones en conjunto con el atributo enfermedad, que el atributo por sí solo, por lo que se podría pensar que está fuertemente correlacionado con el primero.

2. *¿Por qué varía la confianza entre Adulto → Enfermedad y Enfermedad→Adulto?
¿Varía el soporte en esos casos?*

El soporte mide la proporción de ejemplos predecidos correctamente por la regla, sin importar el antecedente, y se ve reflejado en la ecuación que es indistinto del orden de la regla. En cambio, la confianza si contempla el soporte del antecedente de la regla, por lo cual ahora si difiere de qué atributo es antecedente y cual el coneciente. Como adulto tiene un soporte mayor que enfermedad, y el soporte de la regla es el mismo sin importar el orden, la confianza va a ser mayor en “adulto -> enfermedad” porque el soporte de enfermedad, es menor que el de adulto.

3. *¿Que significa que Trabajo Pesado → Adulto tenga una cobertura y confianza relativamente altas, pero un interés bajo?*

Una confianza alta se consigue con un soporte elevado en la regla, y bajo en el antecedente, es decir, si los ejemplos precedidos por la regla tienen una buena proporción de los ejemplos, y la cobertura no es muy alta, eso nos da una buena confianza, sin embargo, el interés de la regla también va a depender de el soporte del consecuente, y si este es elevado el interés va a ser menor. Como el consecuente de la regla es 'adulto' y este es un atributo bastante común en los ejemplos, la regla pierde interés. Por otro lado, el soporte de la regla tampoco es muy elevado, solo que la cobertura tampoco lo es mucho, y por eso parece que la confianza es mejor a la de los demás ejemplos.

4.a) Aplique manualmente el algoritmo APriori para generar los itemsets frecuentes con soporte mayor o igual a 0.3. Ordénelos en base a su soporte.

Soporte mínimo: $(10 \times 30) / 100 = 3$ (soporte)

	Vino	Leche	Miel	Bizcochos	Galletas	Jugo
T1	1	0	0	0	1	1
T2	0	1	0	0	0	0
T3	0	0	1	1	1	0
T4	1	0	1	1	1	1
T5	0	0	0	0	1	0
T6	1	0	0	0	1	1
T7	0	1	1	1	0	0
T8	0	0	1	1	1	1
T9	1	0	0	1	0	1
T10	0	0	0	1	0	0

Tabla C₁

Item	#.
Vino	4
Leche	2
Miel	4
Bizcochos	6
Galletas	6
Jugo	5

Tabla L₁

Item
Vino
Miel
Bizcochos
Galletas
Jugo

Tabla C₂:

Itemset	#
Vino, Miel	1
Vino, Bizcochos	2
Vino Galletas	3
Vino Jugo	4
miel, bizcochos	4
miel, galletas	3
miel jugo	2
bizcochos, galletas	3
bizcochos, jugo	3
galletas, jugo	3

Tabla L₂:

Itemset
Vino, Jugo
Miel, bizcochos
Vino Galletas
Miel, galletas
Bizcochos, galletas
Bizcochos, jugo
Galletas, juego

Tabla C₃: (sin anotar las combinaciones que no existen transacciones)

Itemset	#
<i>Vino, jugo, galletas</i>	3
<i>Vino, jugo, bizcocho</i>	2
<i>Vino, jugo, miel</i>	1
<i>miel, bizcochos, vino</i>	1
<i>miel, bizcocho, galleta</i>	3
<i>miel bizcocho, jugo</i>	2
<i>Vino, galletas, bizcocho</i>	1
<i>Vlno, galletas, miel</i>	1
<i>miel, galletas, jugo</i>	2
<i>Bizcocho, galleta, jugo</i>	2

Tabla L₃:

Itemset
Vino, galletas, jugo
Miel, bizcocho, galletas

Tabla C₄:

Itemset y (L ₄)	#
<i>miel, bizcochos, galletas, vino</i>	1
<i>Vino, galleta, jugo, miel</i>	1
<i>Vino, galleta, jugo, bizcocho</i>	1
<i>Miel, bizcochos, galletas, jugo</i>	2

Tabla L₄:

No existe, ningún itemset tiene soporte = 3

Reglas posibles:

Vino, galletas → jugo

Miel, bizcocho → galletas