

Reglas de asociación – Zoo

1. Análisis de la Base de Datos y un primer análisis de reglas (Líneas 1-165 .R)

El trabajo sobre reglas de asociación se ha realizado sobre el dataset Zoo. Lo primero que se ha realizado ha sido un análisis de nuestra base de datos, podemos encontrar más información en: <http://sci2s.ugr.es/keel/dataset.php?cod=69>

Podemos observar que esta base de datos tiene 101 filas y 17 atributos. El tratamiento que se le ha realizado a la base de datos antes de comenzar a trabajar ha sido cambiar la variable legs para tener dos niveles, tiene piernas o no. Además, el resto de variables, salvo la variable type, las hemos pasado a factor para tener los niveles TRUE o FALSE.

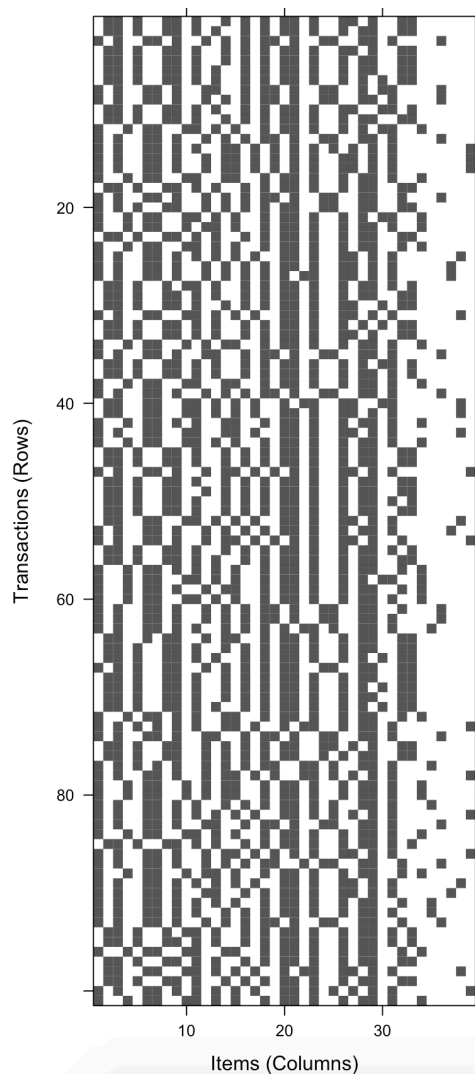


Figura 1

Una vez que obtenemos las transacciones, podemos ver en la Figura 1 la representación gráfica de la distribución de los ítems en las transacciones.

Además, en la Figura 2 se pueden observar los ítems más importantes con un mínimo soporte del 0.4.

Si realizamos un primer análisis con a priori, solamente con un soporte del 0.4 podríamos obtener como se observa en la Figura 3, el diagrama de barras del tamaño de las reglas.

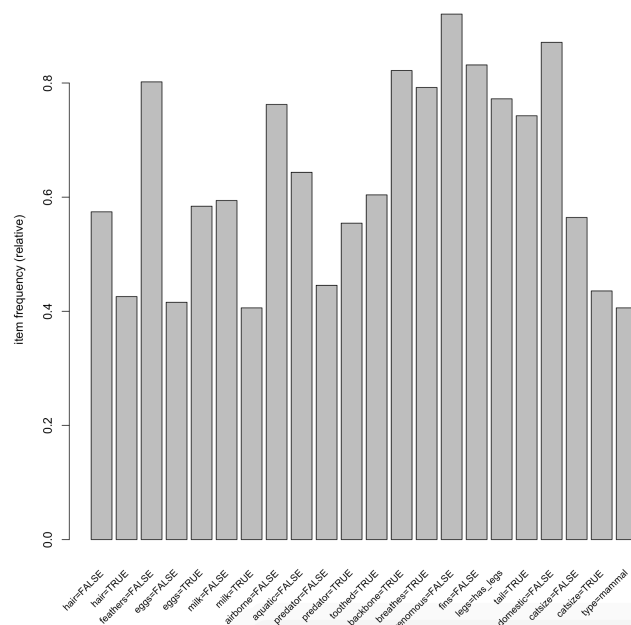


Figura 2

Si analizamos los itemsets maximales y cerrados, podemos ver la cantidad de cada uno de ellos, además de los frecuentes en la Figura 4.

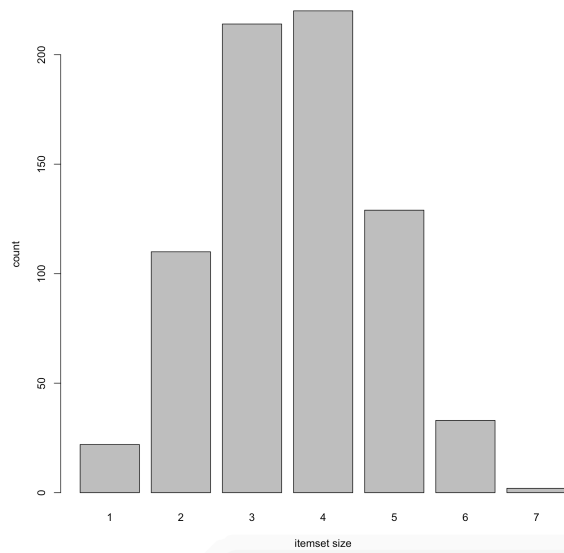


Figura 3

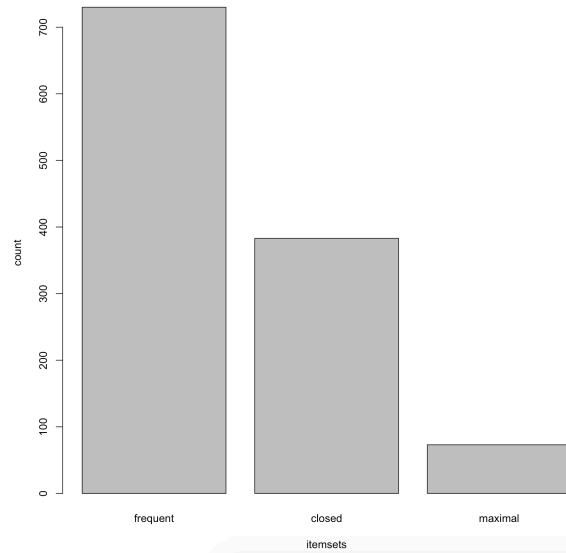


Figura 4

Una vez sacada esta información, se ha procedido a hacer a priori con un soporte del 0.4 y una confianza del 0.8, siendo además una mínima longitud de 2.

Con estas reglas se ha realizado un análisis de las reglas:

Nº	lhs	rhs	Soporte regla	Confianza	Lift	Soporte consecuente
1	Domestic=False	Venemous=False	0.802	0.92	0.999	0.921
2	Eggs=True	Milk=False	0.574	0.983	1.654	0.594
3	Milk=True	Type=Mammal	0.406	1	2.463	0.406
4	Milk=False, Backbone=True	Hair=False	0.416	1	1.74	0.574
5	Milk=False, Venomous=False	Eggs=True	0.515	1	1.71	0.584
6	Eggs=True, Backbone=True	Hair=False	0.406	0.976	1.699	0.574
7	Hair=False, Catsize=False	Milk=False	0.436	1	1.683	0.594

- Si analizamos la regla 1: podemos ver que la confianza de la regla y el soporte del consecuente son iguales, además de que el lift es menor de 1, por lo que esta regla no nos aporta nada.
- La regla 2 puede ser muy interesante al pasar de un soporte en el consecuente del 0.59 a un 0.98 de confianza, aportando información. De esta forma deduciríamos que, si pone huevos, no da leche. De tal forma, si no fuese una base de datos de la que normalmente estamos muy informados, sería una regla muy interesante. Algo interesante que realizar con esta regla, puede ser ver que animales cumplen estas condiciones. En total tenemos 58 animales que cumplen esta regla, de entre los que puede ser el pollo, o muchos de tipo pez.

- De las reglas 3 a 7, tenemos un conjunto de reglas que pasan de un soporte del consecuente de entre el 0.40 al 0.59, a una confianza de las reglas de entre el 0.97 a 1. De esta forma tenemos que con estas reglas aportamos mucha información a los consecuentes y son reglas útiles.

2. Análisis de reglas por grupos (Líneas 169-313 .R)

Para este análisis de reglas en grupos, lo primero que hago es volver a lanzar a priori, de forma que ahora lo lanzaremos con un soporte de 0.1, una confianza del 0.1, una longitud mínima de 2 y una longitud máxima de 3. En total tenemos 7500 reglas con estas características. Una vez realizado, eliminamos las reglas redundantes para hacer un poco de limpieza de reglas.

Un tipo de reglas que he querido analizar han sido las reglas con consecuente Fins=TRUE.

De esta forma he buscado las reglas que cumplen esta condición, y que además tienen tamaño 2, siendo en total 3:

Nº	lhs	rhs	Soporte regla	Confianza	Lift	Soporte consecuente
1	Type=Fish	Fins=True	0.129	1	5.941	0.168
2	Breathes=False	Fins=True	0.129	0.619	3.678	0.168
3	Hair=False	Fins=True	0.148	0.258	1.536	0.168

Por lo que ahora centraré mi búsqueda en encontrar reglas con el consecuente negado, es decir, Fins=FALSE, obteniendo un total de 42 posibles. Para realizar un análisis de reglas en grupo necesito que en el antecedente aparezcan alguno de los anteriores descritos. Así he realizado la búsqueda de estos antecedentes y la única regla que he podido encontrar de tamaño mayor a 2 ha sido:

Nº	lhs	rhs	Soporte regla	Confianza	Lift	Soporte consecuente
1	Hair=False, Legs=True	Fins=False	0.356	1	1.20	0.831
Anterior	Hair=False	Fins=True	0.148	0.258	1.536	0.168

Hemos conseguido el análisis de una regla por grupos, de forma que podríamos decir que: si no tienen pelo tienen aletas, salvo que, si tienen piernas, entonces no tendrán aletas.

Como quería sacar más reglas para analizar en grupos, he realizado una función para automatizar el proceso. Esta función está descrita en el fichero .R que se ha entregado junto a este documento. NOTA: La función puede llegar a tardar demasiado ya que se buscan muchas combinaciones.

La salida que puede proporcionar la función si tenemos los parámetros de a priori con soporte 0.1, confianza 0.1, mínima longitud 2, máxima longitud 3 puede llegar a ser muy grande, al analizar por fuerza bruta. Por lo tanto, vamos a analizar algunas de las reglas:

Nº	lhs	rhs	Soporte regla	Confianza	Lift	Soporte consecuente
1	Toothed=True	Hair=True	0.376	0.622	1.463	0.475
2	Eggs=True, Toothed=True	Hair=False	0.198	1	1.74	0.574
3	Milk=False, Toothed=True	Hair=False	0.207	1	1.74	0.574

La regla nos diría que los animales con dientes tienen pelo, salvo los que ponen huevos o no dan leche.

Nº	lhs	rhs	Soporte regla	Confianza	Lift	Soporte consecuente
1	Toothed=True	Milk=True	0.396	0.6557	1.615	0.405
2	Eggs=True, Toothed=True	Milk=False	0.198	1	1.683	0.594

La regla nos diría que los animales con dientes dan leche, salvo los que ponen huevos. Por lo tanto, es una regla que sería de interés.

Hay que comentar que las reglas analizadas nos pueden parecer triviales, y es debido a que tenemos un gran conocimiento sobre animales, pero en caso de que no supiésemos nada, sería una información importante que extraer.

3. Paquete MOPNAR

En el repositorio de RKEEL se dispone de un paquete llamado MOPNAR específico para las reglas de asociación. Haciendo uso de este paquete podemos tener resultados como los siguientes:

Nº	lhs	rhs	Soporte regla	Confianza	Lift	Soporte consecuente
1	Type=Fish	Fins=True	0.13	1	5.88	0.17
2	Milk=True	Type=Not Fish	0.41	1	2.43	0.41
3	Type=Not mammal	Milk=False	0.6	1	1.66	0.6

- Las 3 reglas serían interesantes al pasar de un soporte en el consecuente de entre 0.17 y 0.6 a una confianza de 1.
- Las reglas tienen un Lift mínimo del 1.66.
- La regla 1 sería que, si es de tipo pez, tiene aletas.
- La regla 2 nos diría que, si da leche, no es de tipo pez.
- La regla 3 dice que los de tipo no mamífero, no dan leche.

Toda esta información proporcionada por el procedimiento, nos daría información interesante, pero como se ha comentado anteriormente, tenemos mucho conocimiento sobre este problema y no nos resulta relevante a simple vista las reglas.