

Ciencia de datos, práctica 2

Juan Casado Ballesteros, Samuel García Gonzalez, Iván Anaya Martín

October 16, 2019

Abstract

Contents

1	Asociación sobre los datos de las cestas de la compra	3
1.1	Cargar los datos de un .txt	3
1.2	Llamar a apriori	3
1.3	Calcular asociación	3
2	Asociación sobre los datos de los vehículos	4
3	Creación de un algoritmo apriori	4

1 Asociación sobre los datos de las cestas de la compra

Hemos introducido los datos de las cestas de la compra en un fichero .txt para evitar tener que escribirlos varias veces. Utilizamos funciones que hemos creado para automatizar la lectura del fichero y también para utilizar el algoritmo apriori. En estas funciones solo realizamos una lectura y una ejecución del algoritmo con los parámetros que nosotros hemos elegido. En el caso del fichero txt guardamos los elementos como listas de la compra en horizontal y los transformamos a como la función que llama a apriori espera recibirlos. En la función que llama a apriori configuramos como queremos utilizar el algoritmo de modo que no se nos impriman los detalles de su ejecución ni se calculen las asociaciones con conjuntos vacíos.

1.1 Cargar los datos de un .txt

Convierte la matriz a datos booleanos y la transpone, nos es más fácil escribir los datos en horizontal y transponer que no escribirlos ya como se espera que estén.

```
> readAprioriFile

function(file){
  muestra<-Matrix(as.matrix(read.table(file)), sparse=T)
  muestrangCMatrix<-as(muestra,"nsparseMatrix")
  t(muestrangCMatrix)
}
```

1.2 Llamar a apriori

```
> calapriori

function(matrix,soporte,confianza){
  transacciones<-as(matrix,"transactions")
  asociaciones <- apriori(transacciones,
    parameter=list(minlen=2, support=soporte,confidence=confianza),
    control=list(verbose=F))
  inspect(asociaciones)
}
```

1.3 Calcular asociación

Calculamos la asociación con soporte 0.5 y confianza 0.8 para los datos de las cestas de la compra.

```
> calapriori(readAprioriFile("datos1.txt"),0.5,0.8)
```

	lhs	rhs	support	confidence	lift	count
[1]	{Agua}	=> {Pan}	0.6666667	1.0	1.20	4
[2]	{Pan}	=> {Agua}	0.6666667	0.8	1.20	4
[3]	{Leche}	=> {Pan}	0.6666667	0.8	0.96	4
[4]	{Pan}	=> {Leche}	0.6666667	0.8	0.96	4
[5]	{Agua,Leche}	=> {Pan}	0.5000000	1.0	1.20	3

2 Asociación sobre los datos de los vehículos

Ya habíamos creado la función para leer datos de un .txt y suministrarlos a apriori. Repetimos el proceso ahora con los datos de los automóviles obteniendo los siguientes resultados para un soporte de 0.4 y una confianza de 0.9.

```
> calapriori(readAprioriFile("datos2.txt"),0.4,0.9)
```

	lhs	rhs	support	confidence	lift	count
[1]	{Control_de_Velocidad}	=> {Faros_de_Xenon}	0.625	1	1.333333	5
[2]	{Bluetooth}	=> {Faros_de_Xenon}	0.625	1	1.333333	5
[3]	{Bluetooth,Control_de_Velocidad}	=> {Faros_de_Xenon}	0.500	1	1.333333	4

3 Creación de un algoritmo apriori

Hemos programado una versión simplificada del algoritmo eliminando algunas de las optimizaciones que este realiza. Para implementarlo lo hemos hecho utilizando tres funciones.

Repetimos el cálculo de la asociación para los datos de las cestas de la compra y de los automóviles comprobando que nuestro algoritmo proporciona los mismos resultados que apriori nos había dado anteriormente. Comprobamos que el algoritmo se comporta como esperábamos.

```
> print(toTable(f_apriori(readAprioriFile("datos1.txt"),0.5,0.8)),right=F)
```

	lhs	rhs	support	confidence	lift	count
1	{Pan}	=> {Agua}	0.6666667	0.8	1.20	4
2	{Agua}	=> {Pan}	0.6666667	1.0	1.20	4
3	{Pan}	=> {Leche}	0.6666667	0.8	0.96	4
4	{Leche}	=> {Pan}	0.6666667	0.8	0.96	4
5	{Agua,Leche}	=> {Pan}	0.5000000	1.0	1.20	3

```
> print(toTable(f_apriori(readAprioriFile("datos2.txt"),0.4,0.9)),right=F)
```

	lhs	rhs	support	confidence	lift	count
1	{Bluetooth}	=> {Faros_de_Xenon}	0.625	1	1.333333	5
2	{Control_de_Velocidad}	=> {Faros_de_Xenon}	0.625	1	1.333333	5
3	{Bluetooth,Control_de_Velocidad}	=> {Faros_de_Xenon}	0.500	1	1.333333	4