



Tema 2

Actividad 2

Autor

Juan José Méndez Torrero

Patrones frecuentes cerrados en SPMF

Introducción

En conjunto de datos proporcionado cuenta con 5 transacciones:

- f a c d g i m p
- a b c f l m o
- b f h j o
- b c k p
- a f c e l p m n

Para esta tarea, con la ayuda SPMF, ejecutaremos los algoritmos FP-Growth y FP-Growth-Close, sobre el conjunto de datos proporcionado y se hará una comparación de los resultados.

Con el objetivo de poder analizar más fácilmente las diferencias obtenidas entre ambos algoritmos, se utilizará un **soporte mínimo del 0.5** para ambos, ya que el uso de un soporte mínimo menor incrementa bastante el número de patrones frecuentes obtenidos en ambos algoritmos.

FP-Growth

Cómo se puede observar en la Figura 1, tras ejecutar el algoritmo FP-Growth sobre los datos, obtenemos 18 patrones frecuentes, con un soporte máximo de 4. Esto quiere decir que, ese patrón, aparece 4 en el conjunto de datos inicial. Además, se puede observar que se han encontrado patrones con una frecuencia de 3 que tiene un tamaño de 4 ítems (a c f m).

Patterns:	
Pattern	#SUP:
p	3
c p	3
m	3
c m	3
f m	3
c f m	3
a m	3
a c m	3
a f m	3
a c f m	3
b	3
a	3
a c	3
a f	3
a c f	3
f	4
c f	3
c	4

Number of patterns: 18

Figura 1. Resultados del algoritmo FP-Growth

FP-Growth-Close

La Figura 2 muestra el resultado obtenido tras ejecutar el algoritmo FP-Growth-Close sobre el conjunto de datos proporcionado. Como se puede observar, el resultado muestra 5 patrones frecuentes, que, observando la Figura 1, son patrones que también encontró la otra versión del algoritmo. También se puede observar, que los patrones frecuentes cerrados encontrados tienen bastante importancia, ya que ha encontrado el patrón con frecuencia 3 que tiene un tamaño de 4 ítems (a c f m).

Patterns:	
Pattern	#SUP:
c p	3
c f a m	3
b	3
f	4
c	4

Number of patterns: 5

Figura 2. Resultados del algoritmo FP-Growth-Close

Conclusiones

Se puede observar que el algoritmo FP-Growth-Close, resulta mucho más eficiente y que, a pesar de haber encontrado un menor número de patrones frecuentes, se puede observar que la información de dichos patrones es igual que la encontrada en todos los patrones frecuentes resultantes del algoritmo FP-Growth.

Por ejemplo, el patrón frecuente encontrado con el algoritmo FP-Growth-Close, (c f a m), con un soporte de 3, quiere decir que los patrones derivados de este patrón, tales como (c f) o (c a m), también tienen un soporte mínimo de 3, tal y como se puede observar en los resultados del algoritmo FP-Growth.