

## (2) Apriori Algorithm

### 甲、摘要

先驗演算法常用於資料探勘中，可重大量數據的資料庫中，找尋資料間彼此的關聯。其步驟可分為兩個，第一步是統計各集合頻率，第二步則是找關聯規則。可透過計算 Support 與 Confidence 來查看數據間關聯的強弱。

Tid	Items
001	A, C, D
002	B, C, E
003	A, B, C, E
004	B, E

$\text{support}(X \rightarrow Y) = \text{Prob.}(X \cup Y)$   
 $\text{min\_support} = 0.5$

▶ Frequent patterns  
 ◦ A, B, C, E  
 ◦ AC, BC, BE, CE  
 ◦ BCE

圖一、頻率統計範例

Tid	Items
001	A, C, D
002	B, C, E
003	A, B, C, E
004	B, E

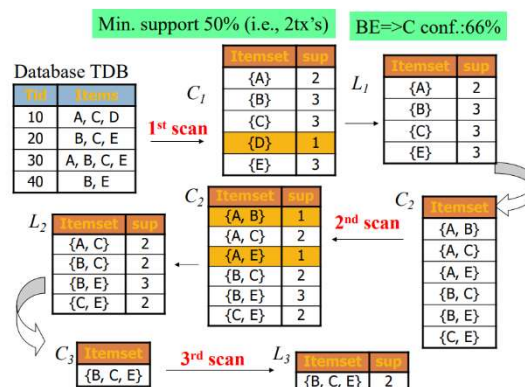
$\text{confidence}(X \rightarrow Y) = \text{Prob.}(X \cup Y / X)$   
 $\text{min\_conf} = 2/3$

▶ Strong association rules  
 ◦  $\{B, E\} \rightarrow C$  (2/3)  
 ◦  $C \rightarrow A$  (2/3)  
 ◦  $A \rightarrow C$  (2/2)

圖二、找關聯規則範例

### 乙、想法

Support: 先利用 BFS 尋找符合 min\_support 的所有集合，接著運用 Downward Closure 理論可以知道，若有被統計到的候選者它的子集也會被統計到。



圖三、計算 support 範例流程

Confidence: 利用條件機率算出 confidence 值，並篩選出有大於 min\_confidence 的候選者。