

**关联：**自然界中某种事物发生时其他事物也会发生，则这种联系称之为关联。反映事件之间依赖或关联的知识称为关联型知识（又称依赖关系）。

**关联规则：**关联是两个或多个变量取值之间存在的一类重要的可被发现的某种规律性。

- **关联规则的数学定义**

- 先设  $I = \{i_1, i_2, \dots, i_m\}$  是一个以  $m$  个不同项为元素的集合， $T$  是针对  $I$  的交易集合，每一笔交易包含若干个属于  $I$  的项。关联规则可表示为  $X \Rightarrow Y$ ，其中  $X, Y \subset I$  且  $X \cap Y = \emptyset$
- $X$  称为规则的前提或前项， $Y$  称为结果或后项。每一规则有两个度量标准，即**支持度 (Support)**和**可信度 (Confidence)**
- 规则的支持度定义为:  $support(X \Rightarrow Y) = support(X \cup Y)$
- 规则的可信度定义为:  $confidence(X \Rightarrow Y) = support(X \cup Y) / support(X)$

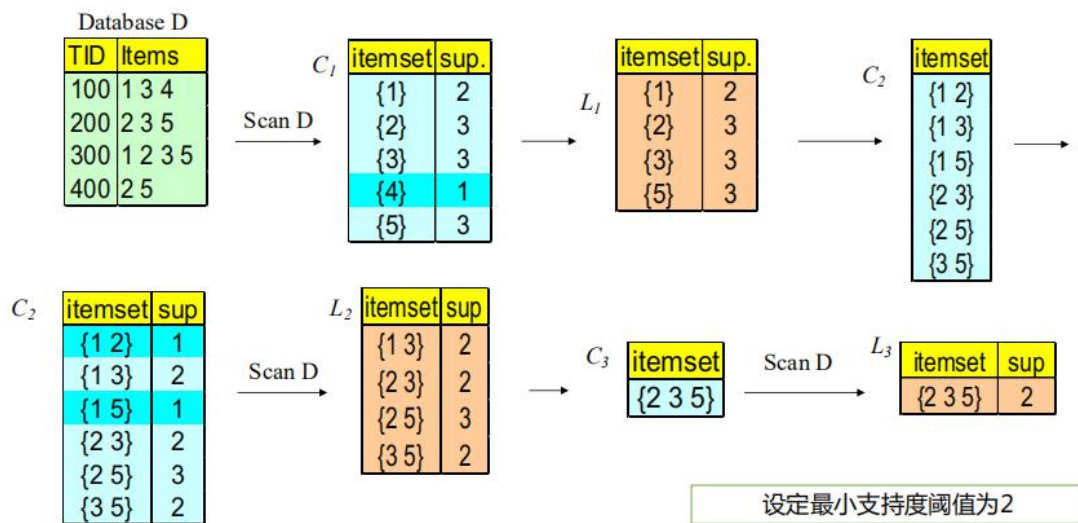
- **关联规则的形式**

- $R: X \Rightarrow Y$
- 其中， $X$  及  $Y$  是两个不相交的集合，即  $X, Y \subset I$  且  $X \cap Y = \emptyset$
- 关联规则可以理解为一个命题，即如果一个交易支持项集  $X$ ，则它也以一定的可能性支持项集  $Y$ ，这一可能性称之为规则的可信度，记为  $conf(R)$  或  $C(R)$

- **举例**

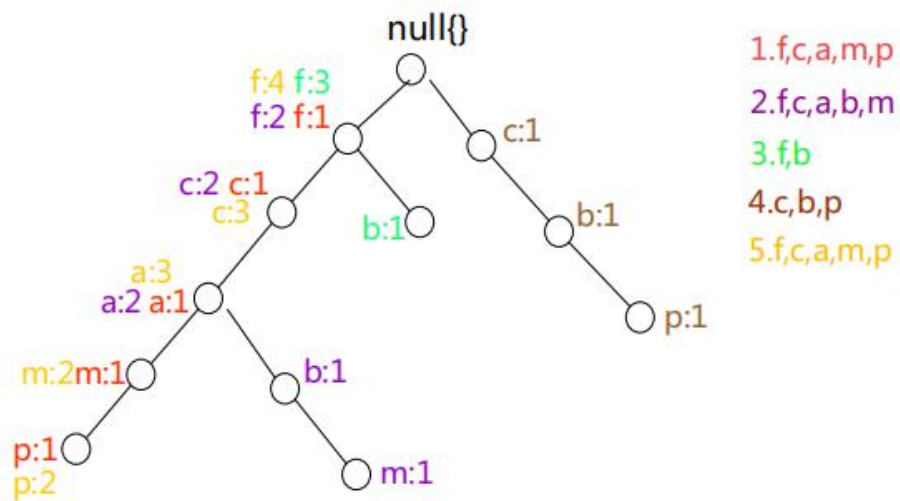
- 规则形式:  $Body \Rightarrow Head [support, confidence]$
- $buys(x, "diapers") \Rightarrow buys(x, "beers") [0.5\%, 60\%]$
- $major(x, "CS") \wedge takes(x, "DB") \Rightarrow grade(x, "A") [1\%, 75\%]$

## Apriori 算法实现流程



## Frequent Patterns Growth (FP-growth)

## FP-growth 算法实例



FP-growth算法树的构造