**Bài 2. Khai thác mẫu phổ biến. Tìm luật kết hợp**

i1 i2 i3 i4 i5 i6 i7 i8

T1 1 0 0 0 0 0 1 1

T2 1 1 0 0 0 1 1 1

T3 1 1 0 0 0 1 1 0

T4 1 0 0 0 0 0 1 1

T5 0 0 1 1 1 1 0 1

T6 1 0 0 1 1 0 0 0

a. Tìm tất cả các tập phổ biến theo ngưỡng minsupp=0,3

b. Tìm tất cả các tập phổ biến tối đại theo ngưỡng minsupp=0,3

c. Tìm tất cả các luật kết hợp hợp lệ theo ngưỡng minsupp=0,3 và ngưỡng minconf=1

được tạo từ các tập phổ biến tối đại của câu 2b.

**Bài làm:**

a. Tìm tất cả các tập phổ biến theo ngưỡng minsupp=0,3:

Để tìm tất cả các tập phổ biến, chúng ta phải đếm số lần xuất hiện của mỗi mục trong tất cả các giao dịch. Sau đó, các mục có số lần xuất hiện lớn hơn hoặc bằng ngưỡng minsupp sẽ được chọn là các tập phổ biến.

Đếm số lần xuất hiện của mỗi mục:

|  |
| --- |
| i1: 5 |
| i2: 3 |
| i3: 1 |
| i4: 1 |
| i5: 2 |
| i6: 2 |
| i7: 4 |
| i8: 4 |

Các tập phổ biến theo ngưỡng minsupp = 0,3:

{ i1 } - support = 5/6 = 0,83

{ i2 } - support = 3/6 = 0,5

{ i7 } - support = 4/6 = 0,67

{ i8 } - support = 4/6 = 0,67

{ i1, i2 } - support = 3/6 = 0,5

{ i1, i7 } - support = 3/6 = 0,5

{ i1, i8 } - support = 3/6 = 0,5

b. Tìm tất cả các tập phổ biến tối đại theo ngưỡng minsupp=0,3:

Để tìm các tập phổ biến tối đại, chúng ta cần xác định những tập phổ biến nào không thể mở rộng thành các tập phổ biến lớn hơn.

Các tập phổ biến tối đa:

{ i1 } { i7 } { i8 } { i1, i7 } { i1, i8 }

c. Tìm tất cả các luật kết hợp hợp lệ theo ngưỡng minsupp=0,3 và ngưỡng minconf=1 được tạo từ các tập phổ biến tối đại của câu 2b:

Để tìm các luật kết hợp, chúng ta cần sử dụng các tập phổ biến tối đa đã tìm được ở bước trước và áp dụng ngưỡng minsupp và minconf cho từng luật kết hợp.

confidence(X -> Y) = support(X union Y) / support(X)

Với ngưỡng minsupp=0,3, các tập itemset có support lớn hơn hoặc bằng 3 là { i1, i7, i8 } và { i5, i7, i8 }. Do đó, ta chỉ có thể tạo ra các luật kết hợp từ các tập này.

Các luật kết hợp hợp lệ với ngưỡng minsupp=0,3 và ngưỡng minconf=1 là:

{ i1, i7 } -> { i8 } (confidence = 2/3 = 0.67)

{ i1, i8 } -> { i7 } (confidence = 2/3 = 0.67)

{ i7, i8 } -> { i1 } (confidence = 2/3 = 0.67)

{ i5, i7 } -> { i8 } (confidence = 1/2 = 0.5)

{ i5, i8 } -> { i7 } (confidence = 1/1 = 1)

{ i7, i8 } -> { i5 } (confidence = 2/3 = 0.67)

{ i1 } -> { i7, i8 } (confidence = 2/5 = 0.4)

{ i7 } -> { i1, i8 } (confidence = 2/4 = 0.5)

{ i8 } -> { i1, i7 } (confidence = 2/4 = 0.5)

{ i5 } -> { i7, i8 } (confidence = 2/2 = 1)