

BÀI TẬP VỀ LUẬT KẾT HỢP

Bài 1. Cho Cơ sở dữ liệu giao dịch sau:

| TID | Nội dung Giao dịch |
|-----|--------------------|
| 100 | ACDI |
| 200 | ACI |
| 300 | CEI |
| 400 | ABDE |
| 500 | BDI |
| 600 | ABDE |

Cho ngưỡng $\text{minsup} = 0.3$, $\text{minconf} = 1$

- Hãy sử dụng thuật toán Apriori để tìm tất cả các tập phổ biến.
- Hãy sử dụng thuật toán FP-growth để tìm tất cả các tập phổ biến.
- Hãy tìm tất cả các luật kết hợp từ tập phổ biến có nhiều phần tử nhất?

Bài 2. Bài tập về tập mặt hàng và tập giao tác $I = \{i1, \dots, i8\}$, $O = \{o1, \dots, o6\}$

$o1 = \{i1, i7, i8\}$

$o2 = \{i1, i2, i6, i7, i8\}$

$o3 = \{i1, i2, i6, i7\}$

$o4 = \{i1, i7, i8\}$

$o5 = \{i3, i4, i5, i6, i8\}$

$o6 = \{i1, i4, i5\}$

- Tìm ngữ cảnh khai thác dữ liệu được tạo từ I , O .
- Tìm tất cả các tập phổ biến theo ngưỡng $\text{minsupp} = 0.3$
- Tìm tất cả các tập phổ biến tối đại theo ngưỡng $\text{minsupp} = 0.3$
- Tìm tất cả các luật kết hợp hợp lệ theo ngưỡng $\text{minsupp} = 0.3$ và ngưỡng $\text{minconf} = 1$ được tạo từ các tập phổ biến tối đại của câu 2c.
- Anh chị có suy nghĩ gì về một thuật toán tìm tập phổ biến tối đại.

Bài 3. Cho Cơ sở dữ liệu giao dịch sau:

| TID | Nội dung Giao dịch |
|-----|--------------------|
| 100 | ACDI |
| 200 | ACI |
| 300 | CEI |
| 400 | ABDE |
| 500 | BDI |
| 600 | ABDE |

Cho ngưỡng minsup = 0.3, minconf = 1

- Hãy tìm tất cả các tập phổ biến bằng thuật toán FP-growth.
- Hãy sử dụng thuật toán Apriori tìm các tập phổ biến thỏa mãn minsupp.
- Hãy tìm tất cả các luật kết hợp từ tập phổ biến có nhiều phần tử nhất?

Bài 4. Cho CSDL Giao dịch như sau:

| | D | E | T | H | I | M |
|-------------|----------|----------|----------|----------|----------|----------|
| T100 | 1 | 0 | 1 | 1 | 0 | 1 |
| T200 | 1 | 0 | 1 | 0 | 0 | 1 |
| T300 | 0 | 0 | 1 | 0 | 1 | 1 |
| T400 | 1 | 1 | 0 | 1 | 1 | 0 |
| T500 | 0 | 1 | 0 | 1 | 0 | 1 |
| T600 | 1 | 1 | 0 | 1 | 1 | 1 |

Cho ngưỡng minsup = 60%, minconf = 100%

- Hãy sử dụng thuật toán Apriori để tìm tất cả các tập phổ biến.
- Hãy sử dụng thuật toán FP-growth để tìm tất cả các tập phổ biến.
- Hãy tìm tất cả các luật kết hợp từ tập phổ biến có nhiều phần tử nhất?

Bài 5. Cho CSDL nhị phân như sau:

| STT | Lạc | Vùng | Đồ | Bánh mì | Bơ | Gạo | Lúa mì |
|-----|-------|-------|-------|---------|-------|-------|--------|
| 1 | Không | Có | Không | Có | Có | Không | Không |
| 2 | Có | Không | Có | Có | Có | Không | Không |
| 3 | Không | Có | Không | Có | Có | Không | Không |
| 4 | Có | Có | Không | Có | Có | Không | Không |
| 5 | Không | Không | Có | Không | Không | Không | Không |
| 6 | Không | Không | Không | Không | Có | Không | Không |
| 7 | Không | Không | Không | Có | Không | Không | Không |
| 8 | Không | Không | Không | Không | Không | Không | Có |
| 9 | Không | Không | Không | Không | Không | Có | Có |
| 10 | Không | Không | Không | Không | Không | Có | Không |

- Hãy sử dụng thuật toán Apriori để tìm tất cả các tập phổ biến.
- Hãy sử dụng thuật toán FP-growth để tìm tất cả các tập phổ biến.
- Hãy tìm tất cả các luật kết hợp từ tập phổ biến có nhiều phần tử nhất?

Bài 6. Cho CSDL sau với MinSup=3

| TID | Nội dung giao dịch |
|-----|--------------------|
| 100 | f,a,b,d,g,i,m,p |
| 200 | a,b,c,f,l,m,o |
| 300 | a,a,h,j,o |
| 400 | b,c,k,s,p |
| 500 | a,f,b,c,l,p,m,n |

- Tìm tập phổ biến bằng thuật toán Apriori.
- Tìm tập phổ biến bằng thuật toán FP_Growth.
- Tìm các luật kết hợp có ngưỡng MinSup=3 và MinConf=1
- Hãy tìm tất cả các luật kết hợp từ tập phổ biến có nhiều phần tử nhất với MinConf=1

Bài 7. Cho CSDL và minsupp = 60%, minconf = 100%

| TID | A | B | C | D | E | F | G | H | K | M | N |
|-----|---|---|---|---|---|---|---|---|---|---|---|
| 10 | | 1 | | | 1 | 1 | 1 | 1 | | | |
| 20 | | | | 1 | | 1 | | | 1 | 1 | |
| 30 | 1 | | 1 | 1 | | 1 | | 1 | | 1 | 1 |
| 40 | 1 | 1 | | | | | | 1 | 1 | 1 | 1 |
| 50 | 1 | 1 | | | 1 | | 1 | 1 | | 1 | 1 |

- Hãy sử dụng Apriori để tìm tập phổ biến.
- Hãy sử dụng FP_Growth để tìm tập phổ biến.
- Hãy tìm tất cả các luật kết hợp từ tập phổ biến có nhiều phần tử nhất?

Bài 8. Sinh viên làm thực hành trên các bộ dữ liệu lớn lấy từ UCI