THỰC HÀNH CÂU 4

(Cơ bản) Cho bảng dữ liệu ở một công ty có các giao dịch như sau:

|  |  |
| --- | --- |
| TID | Items |
| T1 | A, B, C, D |
| T2 | A, B, C |
| T3 | A, B, C |
| T4 | A, B, D |
| T5 | A, B |
| T6 | A, C, D |
| T7 | A, D |
| T8 | B, C, D |

Với min\_sup\_count > 1

1. **Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng thuật toán Apriori**

* Với min\_sup \_count > 1
* Tập các ứng viên 1 phần tử và tần số xuất hiện của nó (support count) tương ứng như sau:
* C1 = {A:7, B:6, C:5, D:5}
  + - * Các tập phổ biến thỏa min\_sup 2 là: A, B, C, D
      * Các tập phổ biến 1 phần tử (tập các ứng viên thoải min\_support\_count):
* L1 = {{A}, {B}, {C}, {D}}
* Tập các ứng viên 2 phần tử và tần số xuất hiện của nó
* C2 = L1⋈L1 = {{A, B}: 5, {A, C}: 4, {A, D}: 3, {B, C}: 4, {B, D}: 3, {C, D}: 3}
  + - * Các tập phổ biến thỏa min\_sup 2 là: {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}
      * Các tập phổ biến 2 phần tử (tập các ứng viên thoải min\_support\_count):
* L2 = {{A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}}
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C3 = L2⋈L2 = {{A, B, C}: 3, {A, B, D}: 2, {A, C, D}: 2, {B, C, D}: 2}
  + - * Các tập phổ biến thỏa min\_sup 2 là: {A, B, C}, {A, B, D}, {A, C, D}, {B, C, D}
      * Các tập phổ biến 3 phần tử (tập các ứng viên thoải min\_support\_count):
* L3 = {{A, B, C}, {A, B, D}, {A, C, D}, {B, C, D}}
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C4 = L3⋈L3 = {{A, B, C, D}: 1}
  + - * Không có tập phổ biến 4 phần tử thỏa min\_sup 2 là
      * Các tập phổ biến 4 phần tử (tập các ứng viên thoải min\_support\_count):
* L4 =

1. **Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng thuật toán FP-Growth**

* Xây dựng cây FP
* Tập phổ biến 1 phần tử và tần suất xuất hiện của nó

|  |  |
| --- | --- |
| Item | Supp |
| A | **7** |
| B | **6** |
| C | **5** |
| D | **5** |

* Tập phổ biến 1 phần tử và tần xuất thỏa điều kiện min\_supp và đã được sắp xếp giảm dần

|  |  |
| --- | --- |
| Item | Supp |
| A | **7** |
| B | **6** |
| C | **5** |
| D | **5** |

* Điền cột Order Frequent Items vào

|  |  |  |
| --- | --- | --- |
| Transaction ID | Items | Ordered frequent items |
| T1 | A, B, C, D | A, B, C, D |
| T2 | A, B, C | A, B, C |
| T3 | A, B, C | A, B, C |
| T4 | A, B, D | A, B, D |
| T5 | A, B | A, B |
| T6 | A, C, D | A, C, D |
| T7 | A, D | A, D |
| T8 | B, C, D | B, C, D |

* Vẽ cây FP

Diagram

Description automatically generated

* Xây dựng cơ sở điều kiện cho mỗi mục phổ biến

|  |  |
| --- | --- |
| Item | Conditional Pattern Base |
| D | {A, B, C}: 1, {A}: 1, {A, B}: 1, {A, C}: 1, {B, C}: 1 |
| C | {A, B}: 3, {A}: 1, {B}: 1 |
| B | {A}: 5 |
| A | {} |

* Xây dựng cây FP điều kiện

|  |  |  |
| --- | --- | --- |
| Item | Conditional Pattern Base | Conditional FP-Tree (supp |
| D | {A, B, C}: 1, {A}: 1, {A, B}: 1, {A, C}: 1, {B, C}: 1 | {{A}: 4, {B}: 3, {C}: 3, {AB}: 2, {A, C}: 2, {B, C}: 2} |D |
| C | {A, B}: 3, {A}: 1, {B}: 1 | {{A}: 4, {B}: 4, {A, B}: 3} |C |
| B | {A}: 5 | {{A}:5} |B |
| A | {} | {} |

* Kết quả các mẫu thường xuyên

|  |  |  |
| --- | --- | --- |
| Item | Conditional FP-Tree | Frequent Pattern |
| D | {{A}: 4, {B}: 3, {C}: 3, {AB}: 2, {A, C}: 2, {B, C}: 2} |D | {{D}: 5, {A, D}: 4, {B, D}: 3, {C, D}: 3, {A, B, D}: 2, {A, C, D}: 2, {B, C, D}: 2} |
| C | {{A}: 4, {B}: 4, {A, B}: 3} |C | {{C}: 5, {A, C}: 4, {B, C}: 4, {A, B, C}: 3} |
| B | {{A}:5} |B | {{B}: 6, {A, B}: 5} |
| A | {} | {A: 7} |

* Lấy những phần tử có 2 phần tử trở lên và thỏa min\_supp ta được các tập hợp sau đây:
* L2 = {{A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}}
* L3 = {{A, B, C}, {A, B, D}, {A, C, D}, {B, C, D}}