THỰC HÀNH CÂU 2

(Cơ bản) Cho bảng dữ liệu ở một cửa hàng tạp hóa có 10 giao dịch như sau:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TID | KÉO | COMPA | THƯỚC | TẬP TRẮNG | BÚT BI | BÚT MÀU | TẨY |
| T1 |  | x |  | x | x |  |  |
| T2 | x |  | x | x | x |  |  |
| T3 |  | x |  | x | x |  |  |
| T4 | x | x |  | x | x |  |  |
| T5 |  |  | x |  |  |  |  |
| T6 |  |  |  |  | x |  |  |
| T7 |  |  |  | x |  |  |  |
| T8 |  |  |  |  |  |  | x |
| T9 |  |  |  |  |  | x | x |
| T10 |  |  |  |  |  | x |  |

Với min\_sup = 30% và min\_conf = 80%

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

* Với min\_sup = 30% và tổng số giao dịch là 10 => Tần số xuất hiện tối thiểu của phần tử để thỏa min\_sup là 3.
* Tập các ứng viên 1 phần tử và tần số xuất hiện của nó (support count) tương ứng như sau:
* C1 = {Kéo: 2, Compa: 3, Thước: 2, Tập trắng: 5, Bút bi: 5, Bút màu: 2, Tẩy:2}
  + - * Các tập phổ biến thỏa min\_sup 3 là: Compa, Tập trắng, Bút bi
      * Các tập phổ biến 1 phần tử (tập các ứng viên thoải min\_support\_count):
* L1 = {{Compa}, {Tập trắng}, {Bút bi}}
* Tập các ứng viên 2 phần tử và tần số xuất hiện của nó
* C2 = L1⋈L1 = {{Compa; Tập trắng}: 3, {Compa; Bút bi}:3, {Bút bi, Tập trắng}:4}
  + - * Các tập phổ biến thỏa min\_sup 3 là: {Compa; Tập trắng}, {Compa; Bút bi}, {Bút bi, Tập trắng}
      * Các tập phổ biến 2 phần tử (tập các ứng viên thoải min\_support\_count):
* L2 = {{Compa; Tập trắng}, {Compa; Bút bi}, {Bút bi, Tập trắng}}
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C3 = L2⋈L2 = {{Compa; Tập trắng; Bút bi}:3}
  + - * Các tập phổ biến thỏa min\_sup 3 là: {Compa; Tập trắng; Bút bi}
      * Các tập phổ biến 3 phần tử (tập các ứng viên thoải min\_support\_count):
* L3 = {{Compa; Tập trắng; Bút bi}}
* Vì L3 chỉ còn 1 tập hợp nên không tiến hành tìm tập ứng cử viên 4 phần tử.

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 |
| Kéo | **2** |
| Compa | **3** |
| Thước | **2** |
| Tập trắng | **5** |
| Bút bi | **5** |
| Bút màu | **2** |
| Tẩy | **2** |

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

|  |  |
| --- | --- |
| Item | Supp |
| Bút bi | **5** |
| Tập trắng | **5** |
| Compa | **3** |

* Điền cột Order Frequent Items vào

|  |  |  |
| --- | --- | --- |
| Transaction ID | Items | Ordered frequent items |
| T1 | Compa, Tập trắng, Bút bi | Bút bi, Tập trắng, Compa |
| T2 | Kéo, Thước, Tập trắng, Bút bi | Bút bi, Tập trắng |
| T3 | Compa, Tập trắng, Bút bi | Bút bi, Tập trắng, Compa |
| T4 | Kéo, Compa, Tập trắng, Bút bi | Bút bi, Tập trắng, Compa |
| T5 | Thước |  |
| T6 | Bút bi | Bút bi |
| T7 | Tập trắng | Tập trắng |
| T8 | Tẩy |  |
| T9 | Bút màu, Tẩy |  |
| T10 | Bút màu |  |

* Vẽ cây FP

Diagram, schematic

Description automatically generated

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

|  |  |
| --- | --- |
| Item | Conditional Pattern Base |
| Compa | {Bút bi; Tập trắng}: 3 |
| Tập trắng | {Bút bi}: 4 |
| Bút bi | {} |

* Xây dựng cây FP điều kiện

|  |  |  |
| --- | --- | --- |
| Item | Conditional Pattern Base | Conditional FP-Tree (supp |
| Compa | {Bút bi; Tập trắng}: 3 | {{Bút bi}:3, {Tập trắng}:3, {Bút bi, Trắng}: 3} |Compa |
| Tập trắng | {Bút bi}:4 | {Bút bi}:4 | Tập trắng |
| Bút bi | {} | {} |

* Kết quả các mẫu thường xuyên

|  |  |  |
| --- | --- | --- |
| Item | Conditional FP-Tree | Frequent Pattern |
| Compa | {{Bút bi}:3, {Tập trắng}:3, {Bút bi, Trắng}: 3} |Compa | {{Compa}:3, {Compa; Bút bi}:3, {Compa: Tập trắng}:3, {Compa, Bút bi, Tập trắng}: 3} |
| Tập trắng | {Bút bi}:4 | Tập trắng | {{Tập trắng}: 4, {Bút bi; Tập trắng}: 4} |
| Bút bi | {} | {{Bút bi}: 5} |

* Lấy những phần tử có 2 phần tử trở lên và thỏa min\_supp ta được các tập hợp sau đây:
* L2 = {{Compa; Bút bi}:3, {Compa; Tập trắng}: 3, {Bút bi; Tập trắng}:4}
* L3 = {{Compa; Bút bi; Tập trắng}:3}

1. **Tìm tất cả các luật kết hợp dựa trên các luật phổ biến ở câu a**

* Xét tập phổ biến {Compa; Bút bi}: 3 có min\_conf = 80%
* {Compa}:3, {Bút bi}: 5, {Compa; Bút bi}:3
* Ta có

|  |  |
| --- | --- |
| Compa 🡪 Bút bi | 3/3 = 100% |
| Bút bi 🡪 Compa | 3/5 = 60% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Compa** 🡪 **Bút bi**
* Xét tập phổ biến {Compa; Tập trắng}: 3 có min\_conf = 80%
* {Compa}:3, {Tập trắng}: 5, {Compa; Tập trắng}:3
* Ta có

|  |  |
| --- | --- |
| Compa 🡪 Tập trắng | 3/3 = 100% |
| Tập trắng 🡪 Compa | 3/5 = 60% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Compa** 🡪 **Tập trắng**
* Xét tập phổ biến {Bút bi; Tập trắng}:4 có min\_conf = 80%
* {Bút bi}: 5, {Tập trắng}: 5, {Bút bi; Tập trắng}: 4
* Ta có

|  |  |
| --- | --- |
| Bút bi🡪 Tập trắng | 4/5 = 80% |
| Tập trắng 🡪 Bút bi | 4/5 = 80% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Bút bi** 🡪 **Tập trắng, Tập trắng** 🡪 **Bút bi**
* Xét tập phổ biến {Compa; Bút bi; Tập trắng}: 3 có min\_conf = 80%
* {Compa}: 3, {Bút bi}: 5, {Tập trắng}: 5, {Compa; Bút bi}: 3, {Compa; Tập trắng}: 3; {Bút bi; Tập trắng}: 4, {Compa; Bút bi; Tập trắng}: 3
* Ta có

|  |  |
| --- | --- |
| Compa 🡪 {Bút bi; Tập trắng} | 3/3 = 100% |
| Bút bi 🡪 {Tập trắng; Compa} | 3/5 = 60% |
| Tập trắng 🡪 {Bút bi; Compa} | 3/5 = 60% |
| {Bút bi; Tập trắng} 🡪 Compa | 3/4 =75% |
| {Tập trắng; Compa} 🡪 Bút bi | 3/3 = 100% |
| {Bút bi; Compa} 🡪 Tập trắng | 3/3 = 100% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Compa** 🡪 **{Bút bi; Tập trắng}, {Tập trắng; Compa}** 🡪 **Bút bi, {Bút bi; Compa}** 🡪 **Tập trắng**