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

|  |
| --- |
|  |

Cho ngưỡng minsup = 0.3, minconf = 1.

-> min\_sup\_count= 0.3\*6 = 1.8 2.

## a) Hãy sử dụng thuật toán Apriori để tìm tất cả các tập phổ biến.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C1 | |  | Tập phổ biến L1  *(chỉ lấy các tập ứng viên từ C1 có count >=2)* | |
| Itemset | count | Itemset | count |
| A | 4 | A | 4 |
| C | 3 | C | 3 |
| D | 4 | D | 4 |
| I | 4 | I | 4 |
| E | 3 | E | 3 |
| B | 3 | B | 3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C2  *(tổ hợp từ L1)* | |  | Tập phổ biến L2  *(chỉ lấy các tập ứng viên từ C2 có count >=2)* | |
| Itemset | count |  | Itemset | count |
| AC | 2 | AC | 2 |
| AD | 3 | AD | 3 |
| AI | 2 | AI | 2 |
| AE | 2 | AE | 2 |
| AB | 2 | AB | 2 |
| CD | 1 |  |  |
| CI | 3 | CI | 3 |
| CE | 1 | CE |  |
| CB | 0 |  |  |
| DI | 2 | DI | 2 |
| DE | 2 | DE | 2 |
| DB | 3 | DB | 3 |
| IE | 1 |  |  |
| IB | 1 |  |  |
| EB | 2 | EB | 2 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C3  *(tổ hợp từ L2, chỉ lấy tập ứng viên có 3 phần tử, loại các tập có tập con không thuộc L2 )* | |  | Tập phổ biến L3  *(chỉ lấy các tập ứng viên từ C3 có count >=2)* | |
| Itemset | count | Itemset | count |
| ABC  (BC k có trong L2-> loại) |  |  |  |
| ABD | 2 | ABD | 2 |
| ABE | 2 | ABE | 2 |
| ABCI  (có 4 hạng mục k tính) |  |  |  |
| ACI | 2 | ACI | 2 |
| ADE | 2 | ADE | 2 |
| ADI | 1 |  |  |
| BDE | 2 | BDE | 2 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C4  *(tổ hợp từ L3, chỉ lấy tập ứng viên có 4 phần tử, loại các tập có tập con không thuộc L3 )* | |  | Tập phổ biến L4  *(chỉ lấy các tập ứng viên từ C4 có count >=2)* | |
| Itemset | count | Itemset | count |
| ABDE | 2 | ABDE | 2 |

* Các tập phổ biến: FI= {A,B,C,D,E,I, AB,AC,AD,AE,AI,BE,BD,CI, DE,DI, ABD,ABE,ACI,ADE,BDE,ABDE}

## b) Hãy sử dụng thuật toán FP-growth để tìm tất cả các tập phổ biến.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | | L1  *(chỉ lấy các item có count >=2)* | | L1  *(sắp xếp item theo chiều giảm dần của count)* | |  | |
| items | count | items | count | items | count | TID | Các giao dịch |
| A | 4 | A | 4 | A | 4 | 100 | A,D,I,C |
| B | 3 | B | 3 | D | 4 | 200 | A,I,C |
| C | 3 | C | 3 | I | 4 | 300 | I,C,E |
| D | 4 | D | 4 | B | 3 | 400 | A, D, B,E |
| E | 3 | E | 3 | C | 3 | 500 | D,I,B |
| I | 4 | I | 4 | E | 3 | 600 | A, D, B,E |

|  |
| --- |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Item  *(sắp xếp theo chiều tăng dần của count)* | Cơ sở mẫu đk | FP-Tree điều kiện  Min\_sup=2 | Các tập phổ biến |
| E | {ADB:2, IC:1} | {A:2},{D:2},{B:2}-E | E,EA,ED,EB,EAD,EAB, EDB,EADB |
| C | {ADI:1, AI:1, I:1} | {A:2}, {I:3}-C | C,CA,CI,CAI |
| B | {AD:2, DI:1} | {A:2}, {D:3}-B | B,BA,BD,BAD |
| I | {AD:1, A:1, D:1} | {A:2}, {D:2}-I | I, IA, ID, IAD |
| D | {A:3} | {A:3}-D | D,DA |
| A |  |  | A |

* Các tập phổ biến: FI= {A,B,C,D,E,I, AB,AC,AD,AE,AI,BE,BD,CI, DE,DI, ABD,ABE,ACI,ADE,BDE,ABDE}

## c) Hãy tìm tất cả các luật kết hợp từ tâp phổ biến có nhiều phần tử nhất?

- Tập phổ biến ABDE, tìm luật kết hợp từ tập phổ biến này, minconf=1.

* Xét loại bỏ 1 tập mục:
* Xét A: R1: DBE->A: conf(R1) = sup(ADBE)/sup(DBE) = 2/2 = 1 >=1
  + - R1 là luật kết hợp
* Xét D: R2:ABE->D: conf(R2) = sup(ADBE)/sup(ABE) = 2/2 = 1 >=1
  + - R2 là luật kết hợp
  + Xét B: R3:ADE->B: conf(R3) = sup(ADBE)/sup(ADE) = 2/2 = 1 >=1
    - R3 là luật kết hợp
  + Xét E: R4:ADB->E: conf(R4) = sup(ADBE)/sup(ADB) = 2/2 = 1 >=1
    - R4 là luật kết hợp
* Xét loại bỏ 2 tập mục:
* Xét AD: R5: BE->AD: conf(R5) = sup(ADBE)/sup(BE) = 2/2 = 1 >=1
  + - R5 là luật kết hợp
* Xét AB: R6:DE->AB: conf(R6) = sup(ADBE)/sup(DE) = 2/2 = 1 >=1
  + - R6 là luật kết hợp
  + Xét AE: R7:DB->AE: conf(R7) = sup(ADBE)/sup(DB) = 2/3 = 0.667 < 1
    - R7 không phải là luật kết hợp
  + Xét DB: R8:AE->DB: conf(R8) = sup(ADBE)/sup(AE) = 2/2 = 1 >=1
    - R8 là luật kết hợp
  + Xét BE: R9:AD->BE: conf(R8) = sup(ADBE)/sup(AD) = 2/3 = 0.667 <1
    - R9 không phải là luật kết hợp
  + Xét DE: R10:AB->DE: conf(R8) = sup(ADBE)/sup(AB) = 2/2 = 1 >=1
    - R10 là luật kết hợp
* Xét loại bỏ 3 tập mục:
* Xét ABD: R11: E->ABD: conf(R5) = sup(ADBE)/sup(E) = 2/3 = 0.667 <1
  + - R11 không phải là luật kết hợp
* Xét ABE: R12:D->ABE: conf(R6) = sup(ADBE)/sup(D) = 2/4 = 0.5 < 1
  + - R12 không phải là luật kết hợp
  + Xét ADE: R13:B->ADE: conf(R7) = sup(ADBE)/sup(B) = 2/3 = 0.667 < 1
    - R13 không phải là luật kết hợp
  + Xét BDE: R14:A->BDE: conf(R8) = sup(ADBE)/sup(A) = 2/4 = 0.5 < 1
    - R15 không phải là luật kết hợp

🡺 Tập luật kết hợp là: R1, R2, R3, R4, R5, R6, R8, R10

# **Bài 2.** Bài tập về tập mặt hàng và tập giao tác I = {i1,…….i8}, O = {o1,…..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}

## a. Tìm ngữ cảnh khai thác dữ liệu được tạo từ I, O.

|  |  |
| --- | --- |
| TID(tập danh mục) | Items(hạng mục) |
| O1 | i1,i7,**i**8 |
| 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 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TID(tập danh mục) | I1 | I2 | I3 | I4 | I5 | I6 | I7 | I8 |
| O1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| O2 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| O3 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| O4 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| O5 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| O6 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |

Minsup = 0.3 -> mincount = 2

* ***Thuật toán Apriori:***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C1 | |  | Tập phổ biến L1  *(chỉ lấy các tập ứng viên từ C1 có count >=2)* | |
| Itemset | count | Itemset | count |
| I1 | 5 | I1 | 5 |
| I2 | 2 | I2 | 2 |
| I3 | 1 | I4 | 2 |
| I4 | 2 | I5 | 2 |
| I5 | 2 | I6 | 3 |
| I6 | 3 | I7 | 4 |
| I7 | 4 | I8 | 4 |
| I8 | 4 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C2 | |  | Tập phổ biến L2  *(chỉ lấy các tập ứng viên từ C2 có count >=2)* | |
| Itemset | count | Itemset | count |
| i1i2 | 2 | i1i2 | 2 |
| i1i4 | 1 | i1i6 | 2 |
| i1i5 | 1 | i1i7 | 4 |
| i1i6 | 2 | i1i8 | 3 |
| i1i7 | 4 | i2i6 | 2 |
| i1i8 | 3 | i2i7 | 2 |
| i2i4 | 0 | i4i5 | 2 |
| i2i5 | 0 | i6i7 | 2 |
| i2i6 | 2 | i6i8 | 2 |
| i2i7 | 2 | i7i8 | 3 |
| i2i8 | 1 |  |  |
| i4i5 | 2 |  |  |
| i4i6 | 1 |  |  |
| i4i7 | 0 |  |  |
| i4i8 | 1 |  |  |
| i5i6 | 1 |  |  |
| i5i7 | 0 |  |  |
| i5i8 | 1 |  |  |
| i6i7 | 2 |  |  |
| i6i8 | 2 |  |  |
| i7i8 | 3 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C3 | |  | Tập phổ biến L3  *(chỉ lấy các tập ứng viên từ C3 có count >=2)* | |
| Itemset | count | Itemset | count |
| i1i2i6 | 2 | i1i2i6 | 2 |
| i1i2i7 | 2 | i1i2i7 | 2 |
| i1i6i7 | 2 | i1i6i7 | 2 |
| i1i6i8 | 1 | i1i7i8 | 3 |
| i1i7i8 | 3 | i2i6i7 | 2 |
| i2i6i7 | 2 |  |  |
| i6i7i8 | 1 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C4 | |  | Tập phổ biến L4  *(chỉ lấy các tập ứng viên từ C4 có count >=2)* | |
| Itemset | count | Itemset | count |
| i1i2i6i7 | 2 | i1i2i6i7 | 2 |

* ***Thuật toán Apriotip:***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C1 | C1’ |  | C1 |  | L1 |  |
| Itemset | tid | Các hạng mục | items | count | items | count |
| I1 | O1 | i1,i7,**i**8 | I1 | 5 | I1 | 5 |
| I2 | O2 | i1,i2,i6,i7,i8 | I2 | 2 | I2 | 2 |
| I3 | O3 | i1,i2,i6,i7 | I3 | 1 | I4 | 2 |
| I4 | O4 | i1,i7,i8 | I4 | 2 | I5 | 2 |
| I5 | O5 | i3,i4,i5,i6,i8 | I5 | 2 | I6 | 3 |
| I6 | O6 | i1,i4,i5 | I6 | 3 | I7 | 4 |
| I7 |  |  | I7 | 4 | I8 | 4 |
| I8 |  |  | I8 | 4 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C2 | C2’ |  | C2 |  | L2 |  |
| Itemset | Tid | Các hạng mục | Itemset | count | Itemset | count |
| i1i2 | O1 | I1i7,i1i8,i7i8 | i1i2 | 2 | i1i2 | 2 |
| i1i4 | O2 | I1i2,i1i6,i1i7,i1i8,i2i6,i2i7,i2i8,i6i7,i6i8,i7i8 | i1i4 | 1 | i1i6 | 2 |
| i1i5 | O3 | I1i2,i1i6,i1i7,i2i6,i2i7,i6i7 | i1i5 | 1 | i1i7 | 4 |
| i1i6 | O4 | I1i7,i1i8,i7i8 | i1i6 | 2 | i1i8 | 3 |
| i1i7 | O5 | I4i5,i4i6,i4i8,i5i6,i6i8,i5i8 | i1i7 | 4 | i2i6 | 2 |
| i1i8 | O6 | I1i4,i1i5,i4i5 | i1i8 | 3 | i2i7 | 2 |
| i2i4 |  |  | i2i4 | 0 | i4i5 | 2 |
| i2i5 |  |  | i2i5 | 0 | i6i7 | 2 |
| i2i6 |  |  | i2i6 | 2 | i6i8 | 2 |
| i2i7 |  |  | i2i7 | 2 | i7i8 | 3 |
| i2i8 |  |  | i2i8 | 1 |  |  |
| i4i5 |  |  | i4i5 | 2 |  |  |
| i4i6 |  |  | i4i6 | 1 |  |  |
| i4i7 |  |  | i4i7 | 0 |  |  |
| i4i8 |  |  | i4i8 | 1 |  |  |
| i5i6 |  |  | i5i6 | 1 |  |  |
| i5i7 |  |  | i5i7 | 0 |  |  |
| i5i8 |  |  | i5i8 | 1 |  |  |
| i6i7 |  |  | i6i7 | 2 |  |  |
| i6i8 |  |  | i6i8 | 2 |  |  |
| i7i8 |  |  | i7i8 | 3 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C3 | C3’ |  | C3 | | L3 | |
| Itemset | tid | Các hạng mục | Itemset | count | Itemset | count |
| i1i2i6 | O1 | i1i7i8 | i1i2i6 | 2 | i1i2i6 | 2 |
| i1i2i7 | O2 | i1i2i6,i1i2i7,i1i2i8,i1i6i7,i1i6i8,  i1i7i8,i2i6i7,i2i6i8,i2i7i8,i6i7i8 | i1i2i7 | 2 | i1i2i7 | 2 |
| i1i6i7 | O3 | i1i2i6,i1i2i7,i1i6i7 | i1i6i7 | 2 | i1i6i7 | 2 |
| i1i6i8 | O4 | I1i7i8 | i1i6i8 | 1 | i1i7i8 | 3 |
| i1i7i8 |  |  | i1i7i8 | 3 | i2i6i7 | 2 |
| i2i6i7 |  |  | i2i6i7 | 2 |  |  |
| i6i7i8 |  |  | i6i7i8 | 1 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C3 | C3’ |  | C3 | | L3 | |
| Itemset | tid | Các hạng mục | Itemset | count | Itemset | count |
| i1i2i6i7 | O1 |  | i1i2i6i7 | 2 | i1i2i6i7 | 2 |
|  | O2 | I1i2i6i7,i1i2i7i8,i1i6i7i8 |  |  |  |  |
|  | O3 | I1i2i6i7 |  |  |  |  |
|  | O4 |  |  |  |  |  |

* ***Thuật toán FP-growth***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | | L1  (chỉ lấy các item có count >=2) | | L1  (sắp xếp item theo chiều giảm dần của count) | |  | |
| items | count | items | count | items | count | TID | Các giao dịch |
| I1 | 5 | I1 | 5 | I1 | 5 | O1 | i1,i7,**i**8 |
| I2 | 2 | I2 | 2 | I7 | 4 | O2 | i1,i7,i8,i6,i2 |
| I3 | 1 | I4 | 2 | I8 | 4 | O3 | i1,i7,i6,i2 |
| I4 | 2 | I5 | 2 | I6 | 3 | O4 | i1,i7,i8 |
| I5 | 2 | I6 | 3 | I2 | 2 | O5 | i8,i6,i4,i5 |
| I6 | 3 | I7 | 4 | I4 | 2 | O6 | i1,i4,i5 |
| I7 | 4 | I8 | 4 | I5 | 2 |  |  |
| I8 | 4 |  |  |  |  |  |  |

|  |
| --- |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Item  (*sắp xếp theo chiều tăng dần của count)* | Cơ sở mẫu đk | FP-Tree điều kiện  min\_sup\_count=2 | Các tập phổ biến |
| I5 | {i1i4:1, i8i6i4:1} | {i4}-i5 | I5,i5i4 |
| I4 | {i1:1,i8i6:1} | - i4 | I4 |
| I2 | {i1i7i8i6:1, i1i7i6:1} | {i1:2, i7:2, i6:2}-i2 | I2,i2i1,i2i7,i2i6,i2i1i7,  i2i1i6,i2i7i6,i2i1i7i6 |
| I6 | {i1i7i8:1, i1i7:1, i8:1} | {i1:2, i7:2, i8:2}-i6 | I6, i6i1, i6i7, i6i8, i6i1i7, i6i1i8, i6i7i8,i6i1i7i8 |
| I8 | {i1i7:3} | {i1:3,i7:3}-i8 | I8,i8i1,i8i7,i8i1i7 |
| I7 | {i1:4} | {i1:4}-i7 | I7, i1i7 |
| I1 |  | - i1 | I1 |

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

FI= { I1, I7, i1i7, I8, i8i1, i8i7, i8i1i7, I6, i6i1, i6i7, i6i8, i6i1i7, i6i1i8, i6i7i8, i6i1i7i8, I2, i2i1, i2i7, i2i6, i2i1i7, i2i1i6, i2i7i6, i2i1i7i6, I4, I5,i5i4}

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

* FI= { i1i2i6i7}

## d. 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 2c.

- Tập phổ biến i1i2i6i7, tìm luật kết hợp từ tập phổ biến này, minconf=1.

* Xét loại bỏ 1 tập mục:
* Xét i1: R1: i2i6i7->i1: conf(R1) = sup(i1i2i6i7)/sup(i2i6i7) = 2/2 = 1 >=1
  + - R1 là luật kết hợp
* Xét i2: R2: i1i6i7->i2: conf(R2) = sup(i1i2i6i7)/sup(i1i6i7) = 2/2 = 1 >=1
  + - R2 là luật kết hợp
  + Xét i6: R3: i2i1i7->i6: conf(R3) = sup(i1i2i6i7)/sup(i2i1i7) = 2/2 = 1 >=1
    - R3 là luật kết hợp
  + Xét i7: R4: i2i6i1->i7: conf(R4) = sup(i1i2i6i7)/sup(i2i6i1) = 2/2 = 1 >=1
    - R4 là luật kết hợp

….

## e. Anh chị có suy nghĩ gì về một thuật toán tìm tập phổ biến tối đại.

# **Bài 4. Cho CSDL Giao dịch như sau:**

|  |
| --- |
|  |

Cho ngưỡng minsup = 60%, minconf = 100%.

=> min\_sup\_count = 0.6\*6 = 3.6

## a) Hãy sử dụng thuật toán Apriori để tìm tất cả các tập phổ biến.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C1 | |  | Tập phổ biến L1 | |
| Itemset | count | So sánh count với min\_sup\_count = 4 | Itemset | count |
| D | 4 | D | 4 |
| E | 3 | H | 4 |
| T | 3 | M | 5 |
| H | 4 |  |  |
| I | 3 |  |  |
| M | 5 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C2 | |  | Tập phổ biến L2 | |
| Itemset | count | So sánh count với min\_sup\_count = 4 | Itemset | count |
| DH | 3 |  |  |
| DM | 3 |  |  |
| HM | 3 |  |  |

* Các tập phổ biến: FI= {D, H ,M}

## b) Hãy sử dụng thuật toán FP-growth để tìm tất cả các tập phổ biến.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | | L1  (chỉ lấy các item có count >=4) | | L1  (sắp xếp item theo chiều giảm dần của count) | |  | |
| items | count | items | count | items | count | TID | Các giao dịch |
| D | 4 | D | 4 | M | 5 | T100 | MDH |
| E | 3 | H | 4 | D | 4 | T200 | MD |
| T | 3 | M | 5 | H | 4 | T300 | M |
| H | 4 |  |  |  |  | T400 | DH |
| I | 3 |  |  |  |  | T500 | MH |
| M | 5 |  |  |  |  | T600 | MDH |

|  |
| --- |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Item  *(sắp xếp theo chiều tăng dần của count)* | Cơ sở mẫu đk | FP-Tree điều kiện  min\_sup\_count=4 | Các tập phổ biến |
| H | {MDH:2, M:1,S:1} | - H | H |
| D | {M:3} | - D | D |
| M |  | - M | M |

* Các tập phổ biến: FI= {D, H ,M}

## c) 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:**

|  |
| --- |
|  |

Minsup = 0.4 => min\_sup\_count = 0.4\*10 = 4

## a) Hãy sử dụng thuật toán Apriori để tìm tất cả các tập phổ biến.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C1 | |  | Tập phổ biến L1 | |
| Itemset | count | So sánh count với min\_sup\_count = 4 | Itemset | count |
| Lạc | 2 | Bánh mì | 5 |
| Vừng | 3 | Bơ | 5 |
| Đỗ | 2 |  |  |
| Bánh mì | 5 |  |  |
| Bơ | 5 |  |  |
| Gạo | 2 |  |  |
| Lúa mì | 2 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C2 | |  | Tập phổ biến L2 | |
| Itemset | count | So sánh count với min\_sup\_count = 4 | Itemset | count |
| Bánh mì- Bơ | 4 | Bánh mì-Bơ | 4 |

* Các tập phổ biến: FI= { Bánh mì ; Bơ; Bánh mì-Bơ}

## b) Hãy sử dụng thuật toán FP-growth để tìm tất cả các tập phổ biến.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | | L1  (chỉ lấy các item có count >=4) | | L1  (sắp xếp item theo chiều giảm dần của count) | |  | |
| items | count | items | count | items | count | TID | Các giao dịch |
| Lạc | 2 | Bánh mì | 5 | Bánh mì | 5 | 1 | Bánh mì, Bơ |
| Vừng | 3 | Bơ | 5 | Bơ | 5 | 2 | Bánh mì, Bơ |
| Đỗ | 2 |  |  |  |  | 3 | Bánh mì, Bơ |
| Bánh mì | 5 |  |  |  |  | 4 | Bánh mì, Bơ |
| Bơ | 5 |  |  |  |  | 5 |  |
| Gạo | 2 |  |  |  |  | 6 | Bơ |
| Lúa mì | 2 |  |  |  |  | 7 | Bánh mì |
|  |  |  |  |  |  | 8 |  |
|  |  |  |  |  |  | 9 |  |
|  |  |  |  |  |  | 10 |  |

|  |
| --- |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Item  (*sắp xếp theo chiều tăng dần của count)* | Cơ sở mẫu đk | FP-Tree điều kiện  min\_sup\_count=4 | Các tập phổ biến |
| Bánh mì |  | - Bánh mì | Bánh mì |
| Bơ | { Bánh mì:4} | { Bánh mì :4}-Bơ | Bơ, Bơ-Bánh mì |

* Các tập phổ biến: FI= { Bánh mì ; Bơ; Bánh mì-Bơ}

## c) 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**

|  |
| --- |
|  |

Min\_sup\_count =

## a) Tìm tập phổ biến bằng thuật toán Apriori.

## b) Tìm tập phổ biến bằng thuật toán FP\_Growth.

## c) Tìm các luật kết hợp có ngưỡng MinSup=3 và MinConf=1

## d) 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%**

|  |
| --- |
|  |

min\_sup\_count = 0.6\*5 = 3

## a) Hãy sử dụng Apriori để tìm tập phổ biến.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C1 | |  | Tập phổ biến L1  *(chỉ lấy các tập ứng viên từ C1 có count >=3)* | |
| Itemset | count | Itemset | count |
| A | 3 | A | 3 |
| B | 3 | B | 3 |
| C | 1 | F | 3 |
| D | 2 | H | 4 |
| E | 2 | M | 4 |
| F | 3 | N | 3 |
| G | 2 |  |  |
| H | 4 |  |  |
| K | 2 |  |  |
| M | 4 |  |  |
| N | 3 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C2  (tổ hợp từ L1) | |  | Tập phổ biến L2  *(chỉ lấy các tập ứng viên từ C2 có count >=3)* | |
| Itemset | count | Itemset | count |
| AB | 2 | AH | 3 |
| AF | 1 | AM | 3 |
| AH | 3 | AN | 3 |
| AM | 3 | BH | 3 |
| AN | 3 | HM | 3 |
| BF | 1 | HN | 3 |
| BH | 3 | MN | 3 |
| BM | 2 |  |  |
| BN | 2 |  |  |
| FH | 2 |  |  |
| FM | 2 |  |  |
| FN | 1 |  |  |
| HM | 3 |  |  |
| HN | 3 |  |  |
| MN | 3 |  |  |
|  |  |  |  |  |
| Tập ứng viên C3  (tổ hợp từ L2) | |  | Tập phổ biến L3  *(chỉ lấy các tập ứng viên từ C3 có count >=3)* | |
| Itemset | count | Itemset | count |
| AHM | 3 | AHM | 3 |
| AHN | 3 | AHN | 3 |
| AMN | 3 | AMN | 3 |
| HMN | 3 | HMN | 3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập ứng viên C4  (tổ hợp từ L3) | |  | Tập phổ biến L4  *(chỉ lấy các tập ứng viên từ C4 có count >=3)* | |
| Itemset | count | Itemset | count |
| AHMN | 3 | AHMN | 3 |

=> Tập phổ biến FI = {A,B,F,H,M,N,AH,AM,AN,BH,HM,HN,MN,AHM,AHN,AMN,HMN,AHMN}

## b) Hãy sử dụng FP\_Growth để tìm tập phổ biến.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | | L1  (chỉ lấy các item có count >=3) | | L1  (sắp xếp item theo chiều giảm dần của count) | |  | |
| items | count | items | count | items | count | TID | Các giao dịch |
| A | 3 | A | 3 | H | 4 | 10 | H,B,F |
| B | 3 | B | 3 | M | 4 | 20 | M,F |
| C | 1 | F | 3 | A | 3 | 30 | H,M,A,F,N |
| D | 2 | H | 4 | B | 3 | 40 | H,M,A,B,N |
| E | 2 | M | 4 | F | 3 | 50 | H,M,A,B,N |
| F | 3 | N | 3 | N | 3 |  |  |
| G | 2 |  |  |  |  |  |  |
| H | 4 |  |  |  |  |  |  |
| K | 2 |  |  |  |  |  |  |
| M | 4 |  |  |  |  |  |  |
| N | 3 |  |  |  |  |  |  |

|  |
| --- |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Item  *(sắp xếp theo chiều tăng dần của count)* | Cơ sở mẫu đk | FP-Tree điều kiện  min\_sup\_count=4 | Các tập phổ biến |
| A | {HM:3} | {H:3},{M:3} - A | A,AH,AM ,AHM |
| B | {H:1,HMA:2} | {H:3} - B | B,BH |
| F | {HB:1,HMA:1, M:1} | - F | F |
| N | {HMAF:1, HMAB:2} | {H:3},{M:3},{A:3} - N | N,NH,NM,NA,NHM,NHA,NMA,NHMA |
| H |  |  | H |
| M | {H:3} | {H:3} - M | M,MH |

=> Tập phổ biến FI = {A,B,F,H,M,N,AH,AM,AN,BH,HM,HN,MN,AHM,AHN,AMN,HMN,AHMN}

## c) Hãy tìm tất cả các luật kết hợp từ tập phổ biến có nhiều phần tử nhất?