1. Lý thuyết:
2. Tìm hiểu và trình bày chi tiết một phương pháp cải tiến quá trình tìm luật kết hợp từ tập phổ biến. Giải thích vì sao nó hiệu quả
3. Cho CSDL sau và minsupp = 50%, minconf = 100%

|  |  |
| --- | --- |
| **TID** | **Items\_bought** |
| **100** | **I, B, F, D, E, C, H, J** |
| **200** | **F, C, F, G, A, D, C** |
| **300** | **B, J, D, A, H** |
| **400** | **E, A, B, E, G** |

1. - Sử dụng thuật toán Apriori để tìm tất cả các tập phổ biến:

* {A, B, D, C, E, F, G, H, J, AB, AD, AG, BD, BH, BE, BJ, CD, CF, DF, DH, DJ, HJ, BDH, BDJ, BHJ, CDF, DHJ, BDHJ}.
* Các tập Large k-item phát sinh:
* Sử dụng thuật toán FP-Growth để tìm tất cả các tập phổ biến:
* { A, B, D, C, E, F, G, H, J, AB, AD, AG, BD, BH, BE, BJ, CD, CF, DF, DH, DJ, HJ, BDH, BDJ, BHJ, CDF, DHJ, BDHJ }.
* Cây FP-Growth:
* Bảng:

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Cơ sở mẫu điều kiện** | **FP-Tree điều kiện** | **Các mẫu phổ biến** |
| J | {BDCEFH: 1, ABDH: 1} | {B: 2, D: 2, H: 2} - J | J, BJ, DJ, HJ, BDJ, BHJ, DHJ, BDHJ |
| H | {BDCEF: 1, ABD: 1} | {B: 2, D: 2} - H | H, BH, DH, BDH |
| G | {ADCF: 1, ABE: 1} | {A: 2} - G | G, AG |
| F | {BDCE: 1, ADC: 1} | {C: 2, D: 2} - F | F, CF, DF, CDF |
| E | {BDC: 1, AB: 1} | {B: 2} - E | E, BE |
| C | {BD: 1, AD: 1} | {D: 2} - C | C, CD |
| D | {B: 1, A: 1, AB: 1} | {B: 2, A: 2} - D | D, BD, AD |
| B | {A: 2} | ∅ | B, AB |
| A | ∅ | ∅ | A |

* So sánh kết quả: Kết quả tìm được của 2 thuật toán là như nhau.
* Liệt kê tập phổ biến tối đại (**Maximal frequent itemsets**): AB, AD, AG, BE, CDF, BDHJ.
* Liệt kê các tập phổ biến đóng (**Closed frequent Itemsets**):

+ A: c(A) = i(t(A)) = i(200, 300, 400) = A.

+ B: c(B) = i(t(B)) = i(100, 300, 400) = B.

+ D: c(D) = i(t(D)) = i(100, 200, 300) = D.

+ AB: c(AB) = i(t(AB)) = i(300, 400) = AB.

+ AD: c(AD) = i(t(AD)) = i(200, 300) = AD.

+ AG: c(AG) = i(t(AG)) = i(200, 400) = AG.

+ BE: c(BE) = i(t(BE)) = i(100, 400) = BE.

+ CDF: c(CDF) = i(t(CDF)) = i(100, 200) = CDF.

+ BDHJ: c(BDHJ) = i(t(BDHJ)) = i(100, 300) = BDHJ.

1. Tìm tất cả các luật kết hợp có dạng (item1 ^ item2 -> item3) thỏa mãn minsupp và minconf: BDH, BDJ, BHJ, CDF, DHJ, BDHJ

* BD -> H: Conf(BD -> H) = 2/2 = 1.
* BH -> D:
* DH -> B:
* BD -> J
* BJ -> D
* DJ -> B
* BH -> J
* BJ -> H
* HJ -> B
* CD -> F
* CF -> D
* DF -> C
* DH -> J
* DJ -> H
* HJ -> D

1. Ứng dụng cải tiến của câu 1 vào việc tìm các luật kết hợp ở câu b thỏa mãn ngưỡng minconf. (So sánh hiệu quả về thời gian thực hiện kết quả ở câu b).
2. Thực hành:
3. A