Bài tập thực hành:

Khai thác tập phổ biến & Luật kết hợp

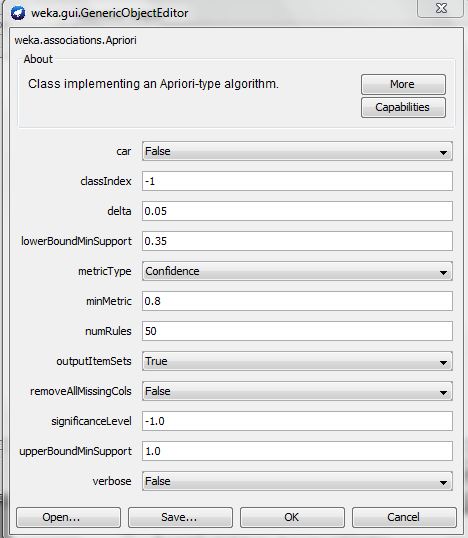
Bài 1: Apriori

1. Hãy sử dụng thuật toán Apriori trong công cụ WEKA để tìm tất cả các tập phổ biến thỏa minsup đã cho
2. Trình bày các tập phổ biến tìm được theo kích thước tập hạng mục

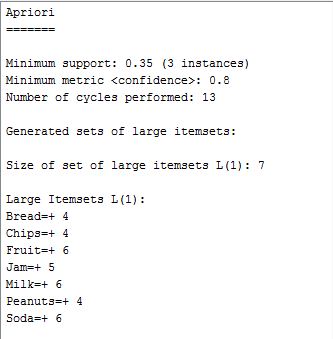
* L1:
* L2:
* L3:
* L4:

1. Chụp màn hình WEKA lúc thiết lập tham số cho thuật toán và màn hình kết quả các tập phổ biến

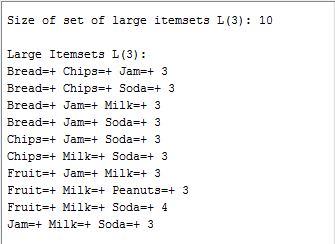
* Thiết lập tham số:

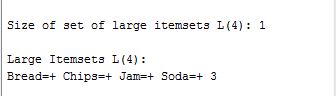


* Kết quả các tập phổ biến:









1. Hãy sử dụng thuật toán Apriori trong WEKA tìm các luật kết hợp thỏa ngưỡng minsup và minconf đã cho:
2. Trình bày các luật tìm được theo từng tập hạng mục

* L1:

1.Bread=+ 4 ==> Jam=+ 4 conf(1)

2.Chips=+ 4 ==> Soda=+ 4 conf(1)

3.Peanuts=+ 4 ==> Fruit=+ 4 conf(1)

22. Milk=+ 6 ==> Fruit=+ 5 conf (0.83)

23. Fruit=+ 6 ==> Milk=+ 5 conf (0.83)

24. Soda=+ 6 ==> Milk=+ 5 conf (0.83)

25. Milk=+ 6 ==> Soda=+ 5 conf (0.83)

26. Jam=+ 5 ==> Bread=+ 4 conf (0.8)

27. Jam=+ 5 ==> Milk=+ 4 conf (0.8)

28. Jam=+ 5 ==> Soda=+ 4 conf (0.8)

* L2:

4. Fruit=+ Soda=+ 4 ==> Milk=+ 4 conf (1)

5. Chips=+ Jam=+ 3 ==> Bread=+ 3 conf (1)

6. Bread=+ Chips=+ 3 ==> Jam=+ 3 conf (1)

7. Bread=+ Soda=+ 3 ==> Chips=+ 3 conf (1)

8. Bread=+ Chips=+ 3 ==> Soda=+ 3 conf (1)

9. Bread=+ Milk=+ 3 ==> Jam=+ 3 conf (1)

10. Bread=+ Soda=+ 3 ==> Jam=+ 3 conf (1)

11. Chips=+ Jam=+ 3 ==> Soda=+ 3 conf (1)

12. Chips=+ Milk=+ 3 ==> Soda=+ 3 conf (1)

13. Fruit=+ Jam=+ 3 ==> Milk=+ 3 conf (1)

14. Milk=+ Peanuts=+ 3 ==> Fruit=+ 3 conf (1)

19. Chips=+ Jam=+ 3 ==> Bread=+ Soda=+ 3 conf (1)

20. Bread=+ Soda=+ 3 ==> Chips=+ Jam=+ 3 conf (1)

21. Bread=+ Chips=+ 3 ==> Jam=+ Soda=+ 3 conf (1)

29. Milk=+ Soda=+ 5 ==> Fruit=+ 4 conf (0.8)

30. Fruit=+ Milk=+ 5 ==> Soda=+ 4 conf (0.8)

* L3:

15. Chips=+ Jam=+ Soda=+ 3 ==> Bread=+ 3 conf (1)

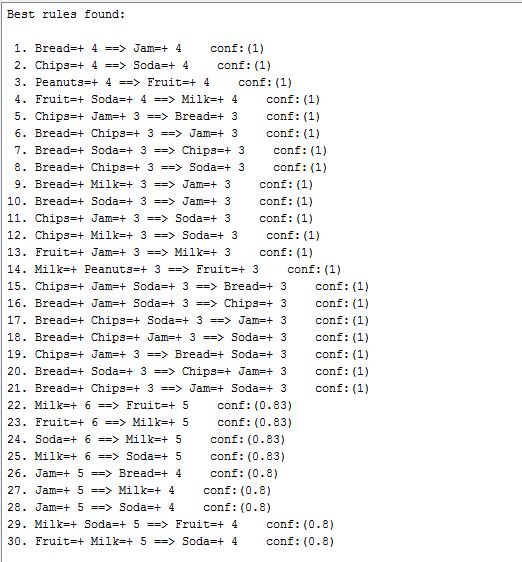
16. Bread=+ Jam=+ Soda=+ 3 ==> Chips=+ 3 conf (1)

17. Bread=+ Chips=+ Soda=+ 3 ==> Jam=+ 3 conf (1)

18. Bread=+ Chips=+ Jam=+ 3 ==> Soda=+ 3 conf (1)

* L4: rỗng

1. Chụp màn hình kết quả luật kết hợp



Bài 2: FP-Growth

1. Hãy trình bày từng bước thuật toán FP-Growth để tìm các tập phổ biến thỏa minsup:
2. Xây dựng cây FP:

* Bước 1: tính độ phổ biến cho các item và sắp xếp các item theo thứ tự giảm dần của độ phổ biến :
* Bảng tất cả các Item:

|  |  |
| --- | --- |
| TID | Hạng mục |
| 1 | Bread,Peanuts,Milk,Fruit,Jam |
| 2 | Bread, Jam,Soda,Chips, Milk,Fruit |
| 3 | Steak, Jam,Soda,Chips, Bread |
| 4 | Jam,Soda, Peanuts,Milk,Fruit |
| 5 | Jam,Soda, Chips, Milk, Bread |
| 6 | Fruit ,Soda,Chips, Milk, |
| 7 | Fruit ,Soda, Peanuts, Milk |
| 8 | Fruit ,Peanuts,Cheese,Yogurt |

* Danh sách các item đã được sắp thứ tự giảm dần theo độ phổ biến:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hạng mục | Fruit | Milk | Soda | Jam | Bread | Chips | Peanuts | Cheese | Steak | Yogurt |
| Supp | 6 | 6 | 6 | 5 | 4 | 4 | 4 | 1 | 1 | 1 |

* Bước 2: Xây dựng cây FP:
* Từ tập dữ liệu ban đầu, ta có được cây FP như sau:

|  |  |  |
| --- | --- | --- |
| TID | Hạng mục | Hạng mục phổ biến  (đã sắp theo thứ tự) |
| 1 | Bread,Peanuts,Milk,Fruit,Jam | Fruit, Milk, Jam,Bread,Peanuts |
| 2 | Bread, Jam,Soda,Chips, Milk,Fruit | Fruit,Milk,Soda,Jam,Bread,Chips |
| 3 | Steak, Jam,Soda,Chips, Bread | Soda,Jam,Bread,Chips |
| 4 | Jam,Soda, Peanuts,Milk,Fruit | Fruit,Milk,Soda,Jam, Peanuts |
| 5 | Jam,Soda, Chips, Milk, Bread | Milk,Soda,Jam,Bread,Chips |
| 6 | Fruit ,Soda,Chips, Milk, | Fruit,Milk,Soda, Chips |
| 7 | Fruit ,Soda, Peanuts, Milk | Fruit,Milk,Soda, Peanuts |
| 8 | Fruit ,Peanuts,Cheese,Yogurt | Fruit ,Peanuts |

* Xét TID = 1:



* Xét TID = 2:



* Xét TID = 3:



* Xét TID = 4:



* Xét TID = 5:



* Xét TID = 6:



* Xét TID = 7:



* Xét TID = 8:



* Từ tập dữ liệu ban đầu, ta xây dựng header table của cây FP như sau:

|  |  |
| --- | --- |
| items | head of node-links |
|  |  |

1. Tìm tập phổ biến: (xem slide 23: FP\_tree slide)

* Xét Fruit:
* Xét Milk:
* Xét Soda:
* Xét Jam:
* Xét Bread:
* Xét Chips:
* Xét Peanuts:

1. Cho nhận xét về thứ tự phát sinh các tập phổ biến của thuật toán Apriori và FP-Growth. Kết quả cuối cùng của chúng có giống nhau không?