

KHAI PHÁ DỮ LIỆU

VO DUC QUANG – VINH UNIVERSITY

QUANGVD@VINHUNI.EDU.VN



Nội dung

- Chương 1: Tổng quan về Data Mining
- Chương 2: Dữ liệu và tiền xử lý dữ liệu
- Chương 3: Bài toán phân lớp dữ liệu
- Chương 4: Bài toán phân cụm dữ liệu
- Chương 5: Khai phá luật kết hợp

Chương 5 – Khai phá luật kết hợp

- Giới thiệu bài toán
- Giải thuật Apriori
- Một số ứng dụng

Khai phá luật kết hợp

- Phân tích hành vi mua hàng trong siêu thị

Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between



Bread and Jam

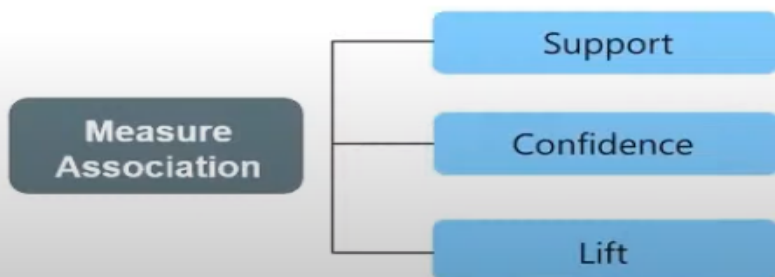
Laptop and Bag



Khai phá luật kết hợp

- Tìm kiếm cách xây dựng luật kết hợp

Association Rule Mining



Khai phá luật kết hợp

- Các công thức

A \Rightarrow B

$$Support = \frac{freq(A, B)}{N}$$

$$Confidence = \frac{freq(A, B)}{freq(A)}$$

$$Lift = \frac{Support}{Supp(A) \times Supp(B)}$$

Khai phá luật kết hợp

■ Ví dụ

Rule	Support	Confidence	Lift
$A \Rightarrow D$	$2/5$	$2/3$	$10/9$
$C \Rightarrow A$	$2/5$	$2/4$	$5/6$
$A \Rightarrow C$	$2/5$	$2/3$	$5/6$
$B, C \Rightarrow A$	$1/5$	$1/3$	$5/9$



Giải thuật Apriori

- Ví dụ minh họa
- $\text{min_support} = 2$

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

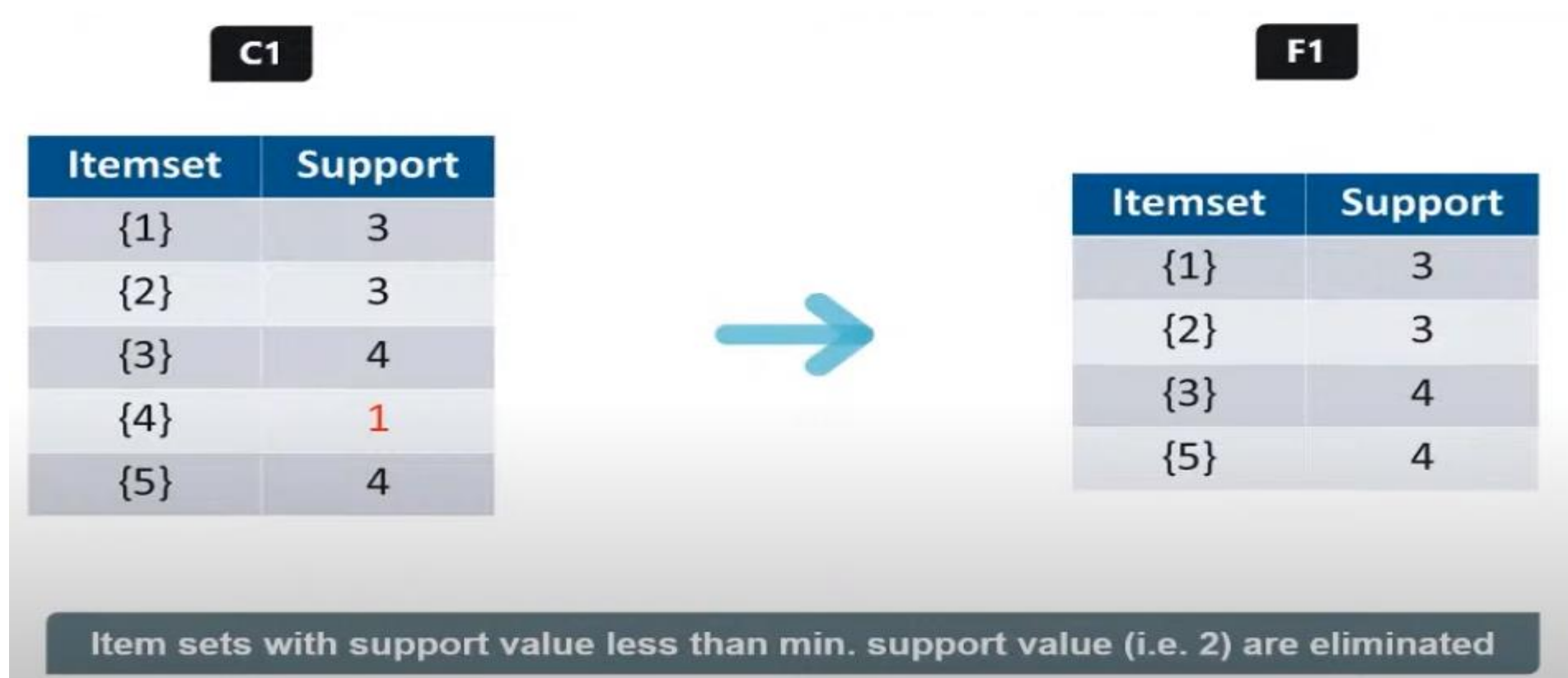


C1

Itemset	Support
{1}	3
{2}	3
{3}	4
{4}	1
{5}	4

Giải thuật Apriori

- Ví dụ minh họa
- $\text{min_support} = 2$



Giải thuật Apriori

- Ví dụ minh họa
- $\text{min_support} = 2$

Only Items present in F1

C2

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5



Itemset	Support
{1,2}	1
{1,3}	3
{1,5}	2
{2,3}	2
{2,5}	3
{3,5}	3



F2

Itemset	Support
{1,3}	3
{1,5}	2
{2,3}	2
{2,5}	3
{3,5}	3

Giải thuật Apriori

- Ví dụ minh họa
- $\text{min_support} = 2$

C3 ?

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5



Itemset	Support
{1,2,3}	
{1,2,5}	
{1,3,5}	
{2,3,5}	

Giải thuật Apriori

- Ví dụ minh họa
- $\text{min_support} = 2$

C3

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5



Itemset	In F2?
{1,2,3}, {1,2}, {1,3}, {2,3}	NO
{1,2,5}, {1,2}, {1,5}, {2,5}	NO
{1,3,5}, {1,5}, {1,3}, {3,5}	YES
{2,3,5}, {2,3}, {2,5}, {3,5}	YES

Giải thuật Apriori

- Ví dụ minh họa

C3

Itemset	In F2?
{1,2,3}, {1,2}, {1,3}, {2,3}	NO
{1,2,5}, {1,2}, {1,5}, {2,5}	NO
{1,3,5}, {1,5}, {1,3}, {3,5}	YES
{2,3,5}, {2,3}, {2,5}, {3,5}	YES



F2

Itemset	Support
{1,3}	3
{1,5}	2
{2,3}	2
{2,5}	3
{3,5}	3

If any of the subsets of these item sets are not there in F2 then we remove that itemset

Giải thuật Apriori

- Ví dụ minh họa

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5



Itemset	Support
{1,3,5}	2
{2,3,5}	2

F3

Giải thuật Apriori

- Ví dụ minh họa

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5



F3

Itemset	Support
{1,3,5}	2
{2,3,5}	2



C3

Itemset	Support
{1,2,3,5}	1



Giải thuật Apriori

■ Ví dụ minh họa

F3

Itemset	Support
{1,3,5}	2
{2,3,5}	2

For $I = \{1,3,5\}$, subsets are $\{1,3\}, \{1,5\}, \{3,5\}, \{1\}, \{3\}, \{5\}$

For $I = \{2,3,5\}$, subsets are $\{2,3\}, \{2,5\}, \{3,5\}, \{2\}, \{3\}, \{5\}$

- For every subsets S of I , output the rule:

$S \rightarrow (I-S)$ (S recommends $I-S$)

if **$\text{support}(I)/\text{support}(S) \geq \text{min_conf value}$**

Giải thuật Apriori

■ Ví dụ minh họa

Applying Rules to Item set F3

1. {1,3,5}

- ✓ Rule 1: **{1,3} → ({1,3,5} - {1,3})** means 1 & 3 → 5
Confidence = $\text{support}(1,3,5) / \text{support}(1,3) = 2/3 = 66.66\% > 60\%$
Rule 1 is selected
- ✓ Rule 2: **{1,5} → ({1,3,5} - {1,5})** means 1 & 5 → 3
Confidence = $\text{support}(1,3,5) / \text{support}(1,5) = 2/2 = 100\% > 60\%$
Rule 2 is selected
- ✓ Rule 3: **{3,5} → ({1,3,5} - {3,5})** means 3 & 5 → 1
Confidence = $\text{support}(1,3,5) / \text{support}(3,5) = 2/3 = 66.66\% > 60\%$
Rule 3 is selected

Giải thuật Apriori

■ Ví dụ minh họa

Applying Rules to Item set F3

1. {1,3,5}

✓ Rule 4: **{1} → ({1,3,5} - {1})** means $1 \rightarrow 3 \ \& \ 5$
Confidence = $\text{support}(1,3,5)/\text{support}(1) = 2/3 = 66.66\% > 60\%$
Rule 4 is selected

✓ Rule 5: **{3} → ({1,3,5} - {3})** means $3 \rightarrow 1 \ \& \ 5$
Confidence = $\text{support}(1,3,5)/\text{support}(3) = 2/4 = 50\% < 60\%$
Rule 5 is rejected

✓ Rule 6: **{5} → ({1,3,5} - {5})** means $5 \rightarrow 1 \ \& \ 3$
Confidence = $\text{support}(1,3,5)/\text{support}(5) = 2/4 = 50\% < 60\%$
Rule 6 is rejected

Khai phá luật kết hợp

- Các công thức

Khai phá luật kết hợp

- Các công thức

Khai phá luật kết hợp

- Các công thức

Why ML & DM?

- Data

Why ML & DM?

- Data

Why ML & DM?

- Data