

Họ và tên: Hà Huy

Son MSSV:

18574802010055

Bài 1:

Thuật toán Apriori



□ Bài tập

- $min_sup = 2$
- $min_conf = 60\%$

TID	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

Trả lời:

C1	
Itemset	Sup. count
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3



L1	
Itemset	Sup. count
{A}	2
{B}	3
{C}	3
{E}	3

C2	
Itemset	Sup. count
{A B}	1
{A C}	2
{A E}	1
{B C}	2
{B E}	3
{C E}	2



L2	
Itemset	Sup. count
{A C}	2
{B C}	2
{B E}	3
{C E}	2

Join C3 = L2 X L2 = {{A C}, {B C}, {B E}, {C E}} X {{A C}, {B C}, {B E}, {C E}}
 = {{A B C}, {A B E}, {B C E}}.

C3	
Itemset	Sup. count
{A B C}	1
{A B E}	1
{B C E}	2

→

L3	
Itemset	Sup. count
{B C E}	2

Nhận thấy C4 = L4, và thuật toán kết thúc.

W = {B C E}

Vậy các tập con thực sự không rỗng của W: {B C}, {B E}, {C E}, {B}, {C}, {E}

Luật 1: {B C} => {{B C E} - {B C}}

conf = sup(BCE) / sup(BC) = 2/2 = 100% > 60%

=> Luật 1 được chọn

Luật 2: {B E} => {{B C E} - {B E}}

conf = sup(B C E) / sup(B E) = 2/3 = 66,66% > 60%

=> Luật 2 được chọn

Luật 3: {C E} => {{B C E} - {C E}}

conf = sup(B C E) / sup(C E) = 2/2 = 100% > 60%

=> Luật 3 được chọn

Luật 4: {B} => {{B C E} - {B}}

conf = sup(B C E) / sup(B) = 2/3 = 66,66% > 60%

=> Luật 4 được chọn

Luật 5: {C} => {{B C E} - {C}}

conf = sup(B C E) / sup(C) = 2/3 = 66,66% > 60%

=> Luật 5 được chọn

Luật 6: {E} => {{B C E} - {E}}

conf = sup(B C E) / sup(E) = 2/3 = 66,66% > 60%

=> Luật 6 được chọn

=====

$\{B\ C\} \Rightarrow \{E\}$, $\text{conf} = 2/2 = 100\%$

$\{B\ E\} \Rightarrow \{C\}$, $\text{conf} = 2/3 = 66\%$

$\{C\ E\} \Rightarrow \{B\}$, $\text{conf} = 2/2 = 100\%$

$\{B\} \Rightarrow \{C\ E\}$, $\text{conf} = 2/3 = 66\%$

$\{C\} \Rightarrow \{B\ E\}$, $\text{conf} = 2/3 = 66\%$

$\{E\} \Rightarrow \{B\ C\}$, $\text{conf} = 2/3 = 66\%$

Bài 2:

		C1	
TID	Items	Itemset	Support
T1	1 3 4	{1}	3
T2	2 3 5	{2}	3
T3	1 2 3 5	{3}	4
T4	2 5	{4}	1
T5	1 3 5	{5}	4

Trả lời:

C1		L1	
Itemset	Sup. count	Itemset	Sup. count
{1}	3	{1}	3
{2}	3	{2}	3
{3}	4	{3}	4
{4}	1	{5}	4
{5}	4		

C2		L2	
Itemset	Sup. count	Itemset	Sup. count
{1 2}	1	{1 3}	3
{1 3}	3	{1 5}	2
{1 5}	2	{2 3}	2
{2 3}	2	{2 5}	3
{2 5}	3	{3 5}	3
{3 5}	3		

Join C3 = L2 X L2 = {{1 3}, {1 5}, {2 3}, {2 5}, {3 5}} X {{1 3}, {1 5}, {2 3}, {2 5}, {3 5}} = {{1 3 5}, {2 3 5}, {1 2 3}, {1 2 5}}.

C3			L3	
Itemset	Sup. count		Itemset	Sup. count
{1 3 5}	2	→	{1 3 5}	2
{2 3 5}	2		{2 3 5}	2
{1 2 3}	1			
{1 2 5}	1			

Join C4 = L3 X L3 = {{1 3 5}, {2 3 5}} X {{1 3 5}, {2 3 5}} = {{1 2 3 5}},

C4 = ∅ và thuật toán kết thúc.

W1 = {1 3 5}

W2 = {2 3 5}

Vậy các tập con thực sự không rỗng của W1: {1 3}, {1 5}, {3 5}, {1}, {3}, {5}

Vậy các tập con thực sự không rỗng của W2: {2 3}, {2 5}, {3 5}, {2}, {3}, {5}

- W1 = {1 3 5}

Luật 1: {1 3} => {{1 3 5} - {1 3}}

conf = 2/3 = 66.66% > 60%

=> Luật 1 được chọn

Luật 2: {1 5} => {{1 3 5} - {1 5}}

conf = 2/2 = 100% > 60%

=> Luật 2 được chọn

Luật 3: {3 5} => {{1 3 5} - {3 5}}

conf = 2/3 = 66.66% > 60%

=> Luật 3 được chọn

Luật 4: {1} => {{1 3 5} - {1}}

conf = 2/3 = 66.66% > 60%

=> Luật 4 được chọn

Luật 5: {3} => {{1 3 5} - {3}}

conf = 2/4 = 50% < 60%

=> Luật 5 không được chọn

Luật 6: $\{5\} \Rightarrow \{\{1\ 3\ 5\} - \{5\}\}$

$\text{conf} = 2/4 = 50\% < 60\%$

=> Luật 6 không được chọn

=====

$\{1\ 3\} \Rightarrow \{5\}$, $\text{conf} = 2/3 = 66.66\% > 60\%$

$\{1\ 5\} \Rightarrow \{3\}$, $\text{conf} = 2/2 = 100\%$

$\{3\ 5\} \Rightarrow \{1\}$, $\text{conf} = 2/3 = 66\%$

$\{1\} \Rightarrow \{3\ 5\}$, $\text{conf} = 2/3 = 66\%$

$\{3\} \Rightarrow \{1\ 5\}$, $\text{conf} = 2/4 = 50\%$

$\{5\} \Rightarrow \{1\ 3\}$, $\text{conf} = 2/4 = 50\%$

• $W2 = \{2\ 3\ 5\}$

Luật 1: $\{2\ 3\} \Rightarrow \{\{2\ 3\ 5\} - \{2\ 3\}\}$

$\text{conf} = 2/2 = 100\% > 60\%$

=> Luật 1 được chọn

Luật 2: $\{2\ 5\} \Rightarrow \{\{2\ 3\ 5\} - \{2\ 5\}\}$

$\text{conf} = 2/3 = 66.66\% > 60\%$

=> Luật 2 được chọn

Luật 3: $\{3\ 5\} \Rightarrow \{\{2\ 3\ 5\} - \{3\ 5\}\}$

$\text{conf} = 2/3 = 66.66\% > 60\%$

=> Luật 3 được chọn

Luật 4: $\{2\} \Rightarrow \{\{2\ 3\ 5\} - \{2\}\}$

$\text{conf} = 2/3 = 66.66\% > 60\%$

=> Luật 4 được chọn

Luật 5: $\{3\} \Rightarrow \{\{2\ 3\ 5\} - \{3\}\}$

$\text{conf} = 2/4 = 50\% < 60\%$

=> Luật 5 không được chọn

Luật 6: $\{5\} \Rightarrow \{\{2\ 3\ 5\} - \{5\}\}$

$\text{conf} = 2/4 = 50\% < 60\%$

=> Luật 6 không được chọn

=====

$\{2\ 3\} \Rightarrow \{5\}, \text{conf} = 2/2 = 100\%$

$\{2\ 5\} \Rightarrow \{3\}, \text{conf} = 2/3 = 66\%$

$\{3\ 5\} \Rightarrow \{2\}, \text{conf} = 2/3 = 66\%$

$\{2\} \Rightarrow \{3\ 5\}, \text{conf} = 2/3 = 66\%$

$\{3\} \Rightarrow \{2\ 5\}, \text{conf} = 2/4 = 50\%$

$\{5\} \Rightarrow \{2\ 3\}, \text{conf} = 2/4 = 50\%$