

BÀI TẬP CHƯƠNG 4

KHAI THÁC LUẬT KẾT HỢP

Bài 1: Cho bảng dữ liệu có 5 giao dịch như sau:

Transaction	ID Items
T001	{I7, I9, I8, I6, I4, I11}
T002	{I3, I9, I8, I6, I4, I11}
T003	{I7, I1, I6, I4}
T004	{I7, I10, I2, I6, I11}
T005	{I2, I9, I6, I5, I4}

Với $\text{min_sup} = 60\%$ và $\text{min_conf} = 80\%$

- Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng giải thuật Apriori.
- Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng giải thuật FP-growth.
- Tìm tất cả các luật kết hợp dựa trên các tập phổ biến tìm được ở câu 1.

Bài 2: Cho bảng dữ liệu ở một cửa hàng văn phòng phẩm như sau:

TID	KÉO	COMPA	THƯỚC	TẬP TRẮNG	BÚT BI	BÚT MÀU	TẤY
T1		x		x	x		
T2	x		x	x	x		
T3		x		x	x		
T4	x	x		x	x		
T5			x				
T6					x		
T7				x			
T8							x
T9						x	x
T10						x	

Với $\text{min_sup} = 30\%$ và $\text{min_conf} = 80\%$, sinh viên thực hiện lại các yêu cầu trên.

Bài 3: Cho CSDL về Nhân viên được cho trong bảng sau:

	Giới tính (GT)	Tuổi (T)	Năng lực làm việc (NL)	Đã lập gia đình (LGD)	Thu nhập (TN)	Thăng chức (TC)
1	Nữ	20..25	Giỏi	Rồi	Rất cao	Có
2	Nam	20..25	Khá	Chưa	Khá	Không
3	Nữ	26..30	Giỏi	Chưa	Khá	Có
4	Nữ	31..40	T.Bình	Chưa	T.Bình	Có
5	Nam	26..30	T.Bình	Rồi	Rất cao	Không
6	Nữ	26..30	Khá	Chưa	Cao	Không
7	Nữ	31..40	Khá	Chưa	T.Bình	Không
8	Nam	26..30	Khá	Rồi	Cao	Có
9	Nữ	>40	Giỏi	Rồi	T.Bình	Không
10	Nữ	26..30	Giỏi	Chưa	Khá	Có

Cho $B = \{\text{Tuổi, Năng lực làm việc, Thăng chức}\}$. Hãy tìm tất cả các luật kết hợp có vẻ phải chỉ gồm thuộc tính Thăng chức (TC) thỏa ngưỡng minsup=30% và minconf = 80%.

Bài 4: Một cơ sở dữ liệu có 4 giao dịch. Cho min sup (ngưỡng hỗ trợ tối thiểu) = 60%, min conf (ngưỡng độ tin cậy tối thiểu) = 80%. Dữ liệu giao dịch như sau:

CustID	TID	Items bought (brand–item category)
01	T100	{King’s-Crab, Sunset-Milk, Dairyland-Cheese, Best-Bread}
02	T200	{Best-Cheese, Dairyland-Milk, Goldenfarm-Apple, Tasty-Pie, Wonder-Bread}
01	T300	{Westcoast-Apple, Dairyland-Milk, Wonder-Bread, Tasty-Pie}
03	T400	{Wonder-Bread, Sunset-Milk, Dairyland-Cheese}

a) Ở mức item category (ví dụ: Milk thay vì Sunset-Milk), xét mẫu luật:

$$\forall X \in \text{transaction}, \text{buys}(X, \text{item1}) \wedge \text{buys}(X, \text{item2}) \Rightarrow \text{buys}(X, \text{item3})[s, c]$$

- Liệt kê tập mục phổ biến (frequent itemset) với k lớn nhất (tức là tập có nhiều mục nhất nhưng vẫn thỏa min sup = 60%).
- Viết ra tất cả luật kết hợp mạnh (strong association rules) có chứa tập mục phổ biến lớn nhất đó, kèm theo giá trị support s và confidence c.

b) Ở mức brand–item category (ví dụ: *Sunset-Milk* thay vì *Milk*), xét mẫu luật:

$$\forall X \in \text{customer}, \text{buys}(X, \text{item1}) \cap \text{buys}(X, \text{item2}) \Rightarrow \text{buys}(X, \text{item3})$$

Liệt kê **frequent itemset** với k lớn nhất, nhưng không cần in ra luật.

Bài 4: Cho mẫu dữ liệu tại [liên kết](#) sau về các giao dịch trong một tháng của một cửa hàng outlet, gồm 9835 giao dịch và 169 items. a) SV Chọn một ngôn ngữ lập trình, cài đặt giải thuật Apriori. để tìm tất cả các luật kết hợp (sinh viên tự chọn min_sup và min_conf).