

Khai thác Tập luật phổ biến

Giảng viên: Tiến sĩ Bùi Thanh Hùng
Email: tg_buithanhhung@tdtu.edu.vn
Website: <https://sites.google.com/site/hungthanhbui1980>

Bài 1: Tìm luật sử dụng giải thuật Apriori với bộ dữ liệu sau, cho biết Support threshold=50%. Hãy tính xác suất Confidence của tập kết quả.

Transaction	List of items
T1	I1,I2,I3
T2	I2,I3,I4
T3	I2,I4,I5
T4	I1,I2,I4
T5	I1,I2,I3,I5
T6	I1,I2,I3,I4

Bài 2: Sử dụng Weka của bộ dữ liệu cung cấp, lựa chọn các tham số đầu vào khác nhau và đưa ra tập luật kết quả.



Bài 3: Cài đặt giải thuật Apriori, so sánh kết quả với thư viện Weka trên bộ dữ liệu được cung cấp ở bài 2, biết rằng framework của giải thuật như sau:

Thuật toán Apriori [WKQ08]:

Input: - Cơ sở dữ liệu giao dịch $D = \{t \mid t \text{ giao dịch}\}$
 - Độ hỗ trợ tối thiểu $\text{minsup} > 0$

Output: - Tập hợp tất cả các tập phổ biến.

```
0: mincount = minsup * |D|;
1.   $F_1 = \{\text{các tập phổ biến có độ dài 1}\}$ 
2.  for ( $k=1$ ;  $F_k \neq \emptyset$ ;  $k++$ ) do begin
3.       $C_{k+1} = \text{apriori-gen}(F_k)$ ; // sinh mọi ứng viên độ dài k+1
4.      for  $t \in D$  do begin
5.           $C_t = \{c \in C_{k+1} \mid c \subseteq t\}$ ; //mọi ứng viên chứa trong t
6.          for  $c \in C_t$  do
7.               $c.\text{count}++$ ;
8.      end
9.       $F_{k+1} = \{c \in C_{k+1} \mid c.\text{count} \geq \text{mincount}\}$  ;
10. end
11. Answer  $\cup_k F_k$  ;
```