



CHƯƠNG 3

Khai phá tập phổ biến



NỘI DUNG CHƯƠNG 3

- Giới thiệu tập phổ biến
- Các thuật toán tìm tập phổ biến



I. Giới thiệu

Cơ sở dữ liệu bán hàng (transaction)

Tid	Items bought
10	Beer, Nuts, Diaper
20	Beer, Coffee, Diaper
30	Beer, Diaper, Eggs
40	Nuts, Eggs, Milk
50	Nuts, Coffee, Diaper, Eggs, Milk

Mã hóa ? Để lưu trữ và xử lý (biến đổi CSDL về dạng nhị phân)



I. Giới thiệu

Cho tập hạng mục (**itemset**) $I = \{i_1, i_2, \dots, i_n\}$; $i_i \in I$ được gọi là một hạng mục (**item**).

Cho CSDL giao dịch (**transaction**) $T = \{t_1, t_2, \dots, t_m\}$; trong đó mỗi t_i là một giao dịch và là một tập con của I .



I. Giới thiệu

Số lượng các giao dịch T ký hiệu là $|T|$ hay $\text{card}(T)$.

Tập k hạng mục (k - **itemset**) = $\{i_1, i_2, \dots, i_k\}$; gồm k item trong I .



I. Giới thiệu

Ví dụ: CSDL giao dịch là bảng có dạng như sau:

Mã giao dịch	Nội dung giao dịch
1	A, C, T, W
2	C, D, W
3	A, C, T, W
4	A, C, D, W
5	A, C, D, T, W
6	C, D, T



I. Giới thiệu

Gọi $I = \{i_1, i_2, \dots, i_n\}$ là tập tất cả các phần tử riêng biệt.

Luật kết hợp: (*association rule*) được ký hiệu là $X \rightarrow Y$.

Với X, Y là hai tập con của I .

- Trong đó X và Y là hai tập không giao nhau, thể hiện sự ràng buộc của tập hạng mục Y theo tập hạng mục X .
- Nếu A xuất hiện trong một giao dịch thì nhiều khả năng B cũng xuất hiện trong cùng một giao dịch.



I. Giới thiệu

■ Độ đo cho luật kết hợp:

◦ Độ hỗ trợ (*support*) là xác suất mà một giao dịch chứa $X \cup Y$.

$$\text{support}("X \rightarrow Y") = P(X \cup Y) = N(X \cup Y) / |T|$$

◦ Độ tin cậy (*confidence*) là xác suất có điều kiện của B nếu biết A .

$$\begin{aligned} \text{confidence}("X \rightarrow Y") &= P(Y|X) = N(X \cup Y) / N(X) \\ &= \text{supp}(X \rightarrow Y) / \text{supp}(X) \end{aligned}$$



I. Giới thiệu

Ví dụ: luật kết hợp

Computer \rightarrow *antivirus* [*support*=2%, *confidence*=60%].

Nghĩa là:

- 2% trong toàn bộ CSDL giao dịch mua computer và antivirus cùng nhau.
- 60% trong số những người mua computer có mua antivirus.



I. Giới thiệu

Luật kết hợp $X \rightarrow Y$ được coi là một mẫu có giá trị nếu xảy ra đồng thời **$\text{supp}(X \rightarrow Y) \geq \text{minsup}$** và **$\text{conf}(X \rightarrow Y) \geq \text{minconf}$** .

Trong đó **minsup** và **minconf** là hai giá trị ngưỡng cho trước.



I. Giới thiệu

Transaction-id	Items bought
1	A, B, C
2	A, C
3	A, D
4	B, E, F

$min_support = 50\%$, $min_conf = 50\%$:

- $B \rightarrow E$ (25%, 50%),... \rightarrow loại
- $A \rightarrow C$ (50%, 66.7%)
- $C \rightarrow A$ (50%, 100%)



I. Giới thiệu

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

Một số ví dụ luật kết hợp:

$\{\text{Milk, Diaper}\} \rightarrow \{\text{Beer}\}$

$\{\text{Diaper, Beer}\} \rightarrow \{\text{Milk}\}$

$\{\text{Diaper}\} \rightarrow \{\text{Milk, Beer}\}$

$\{\text{Milk, Beer}\} \rightarrow \{\text{Diaper}\}$

$\{\text{Beer}\} \rightarrow \{\text{Milk, Diaper}\}$

$\{\text{Milk}\} \rightarrow \{\text{Diaper, Beer}\}$



I. Giới thiệu

Tập phổ biến (frequent itemsets - FIs)

- Một tập phần tử có độ hỗ trợ (support) lớn hơn hoặc bằng một ngưỡng (minimum support) được gọi là tập phổ biến.

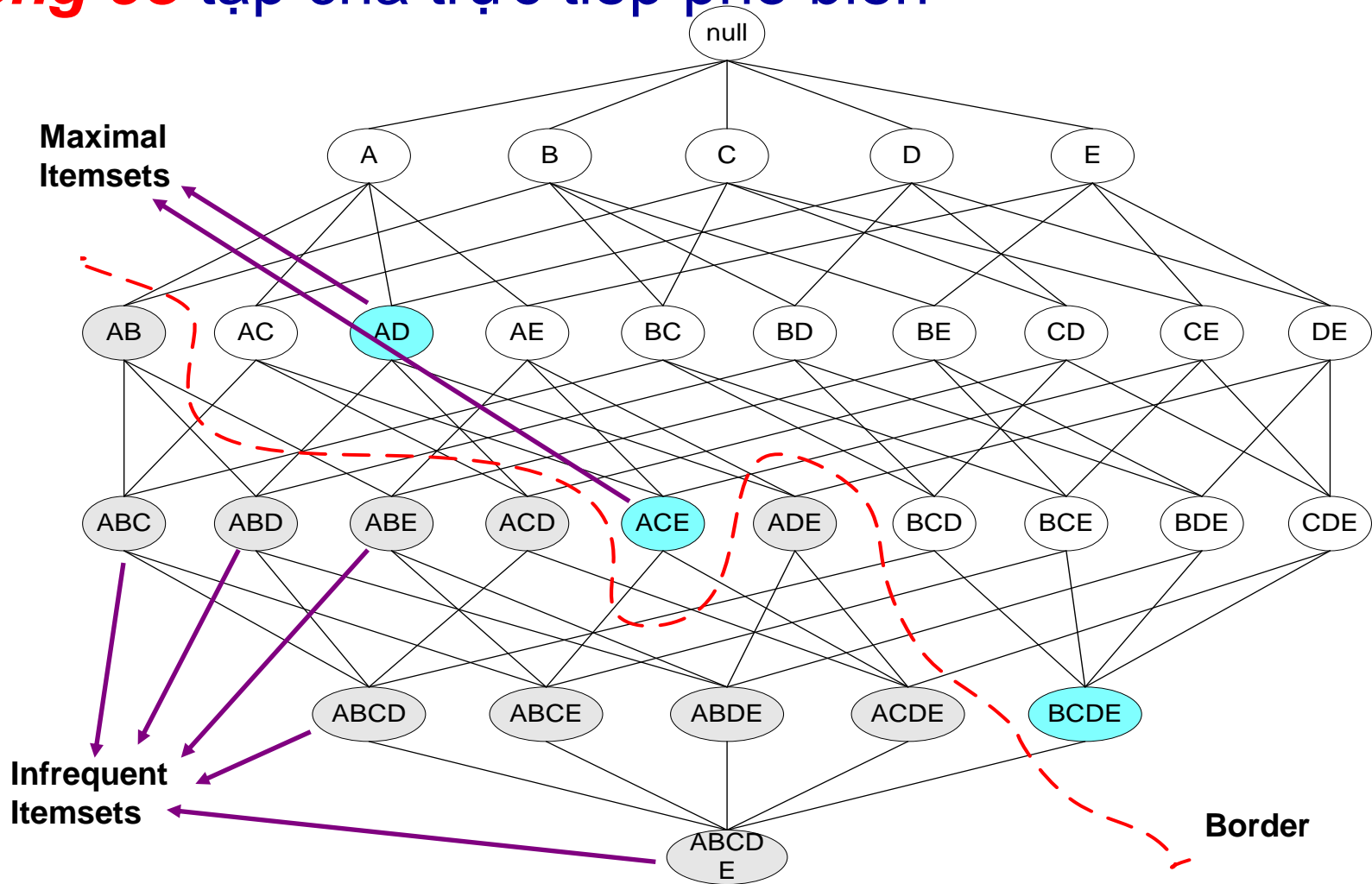
Ví dụ: CSDL giao dịch

- Hãy liệt kê các tập phổ biến 1 phần tử
thỏa $\text{minsupp} = 2$ (66,67%)

Tid	Items
1	A, B, C, D, E
2	B, C, D, E
3	A, C, D, F

TẬP PHỔ BIẾN CỰC ĐẠI

Một tập là phổ biến cực đại nếu tập đó là phổ biến và **không có** tập cha trực tiếp phổ biến





I. Giới thiệu

Tập phổ biến tối đại (Max Pattern/Max Frequent Itemsets) X nếu thỏa:

1. $\text{Supp}(X) \geq \text{minsupp}$ và
2. $\nexists |X'| \supset |X|$ mà X' cũng phổ biến.

Ví dụ: Liệt kê các MFIs thỏa $\text{minsupp} = 2$:

Các MFIs là $\{B, C, D, E\}$ và $\{A, C, D\}$.

Tid	Items
1	A, B, C, D, E
2	B, C, D, E
3	A, C, D, F

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count): 5
Frequent itemsets: ?

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count): 5

Frequent itemsets: {F}

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: ?

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

Support threshold (by count): 3

Frequent itemsets: ?

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

Support threshold (by count): 3

Frequent itemsets:

All subsets of {C,D,E,F} + {J}

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Maximal itemsets: ?

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

Maximal itemsets: ?

Support threshold (by count): 3

Frequent itemsets:

All subsets of {C,D,E,F} + {J}

Maximal itemsets: ?

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Maximal itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

Maximal itemsets: ?

Support threshold (by count): 3

Frequent itemsets:

All subsets of {C,D,E,F} + {J}

Maximal itemsets: ?

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Maximal itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

Maximal itemsets: {E,F}, {J}

Support threshold (by count): 3

Frequent itemsets:

All subsets of {C,D,E,F} + {J}

Maximal itemsets: ?

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Maximal itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

Maximal itemsets: {E,F}, {J}

Support threshold (by count): 3

Frequent itemsets:

All subsets of {C,D,E,F} + {J}

Maximal itemsets:

{C,D,E,F}, {J}

TẬP PHỒ BIẾN CỰC ĐẠI

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Support threshold (by count) : 5

Frequent itemsets: {F}

Maximal itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: {E}, {F}, {E,F}, {J}

Maximal itemsets: {E,F}, {J}

Support threshold (by count): 3

Frequent itemsets:

All subsets of {C,D,E,F} + {J}

Maximal itemsets:

{C,D,E,F}, {J}



I. Giới thiệu

Tập phổ biến đóng (Closed Pattern/Frequent Closed Itemsets) X nếu thỏa:

1. $\text{Supp}(X) \geq \text{minsupp}$ và
2. $\nexists |X'| \supset |X|$ mà $\text{supp}(X') = \text{supp}(X)$.

Ví dụ: Liệt kê các FCIs thỏa $\text{minsupp} = 2$.

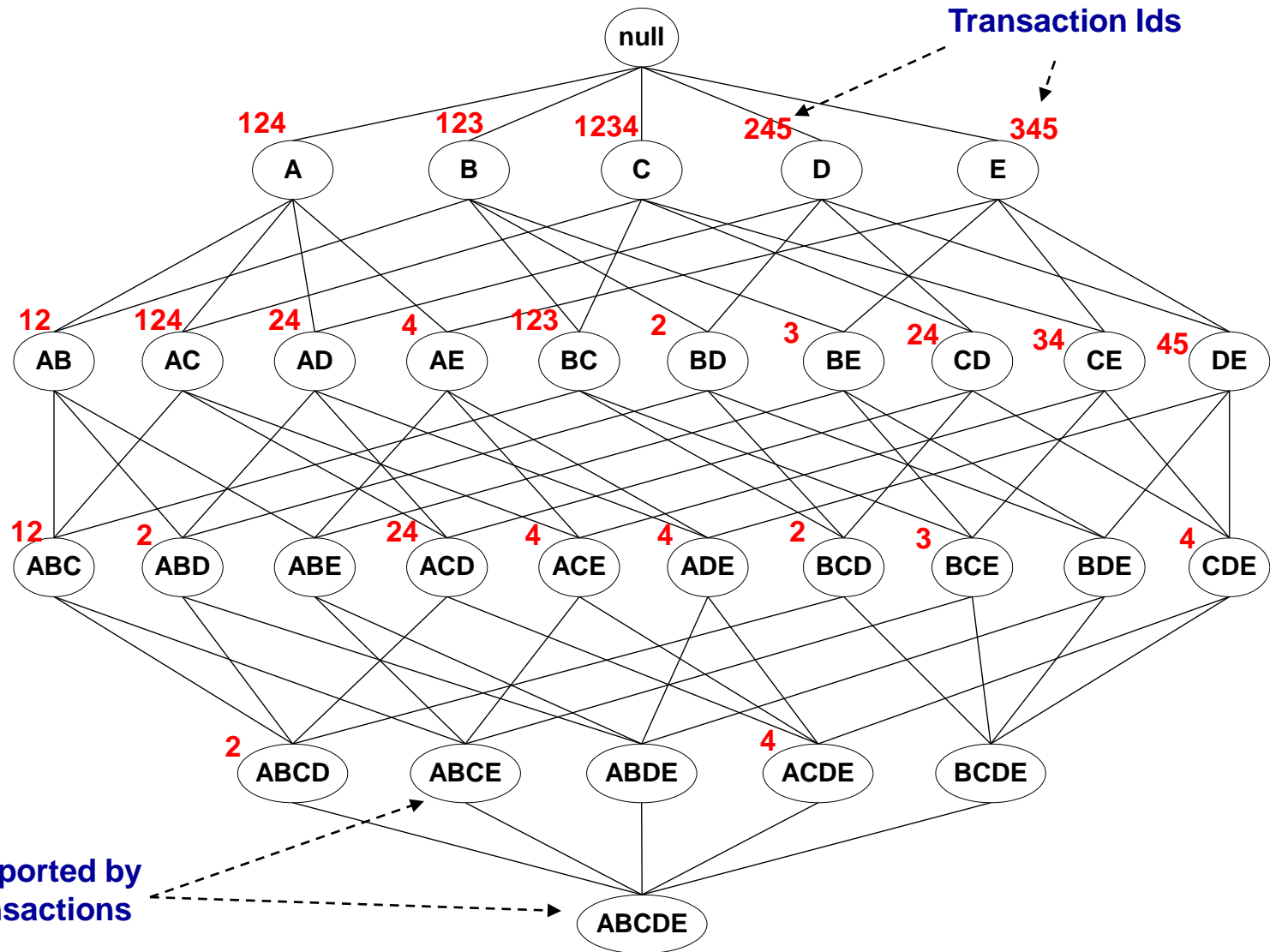
Các FCIs $\{A,B,C\}$, $\{A,B,D\}$ và $\{A,B\}$.

Hỏi: $\{A,B\}$ là MFI ?

Tid	Items
1	A, B, C
2	A, B, C
3	A, B, D
4	A, B, D
5	C, E, F

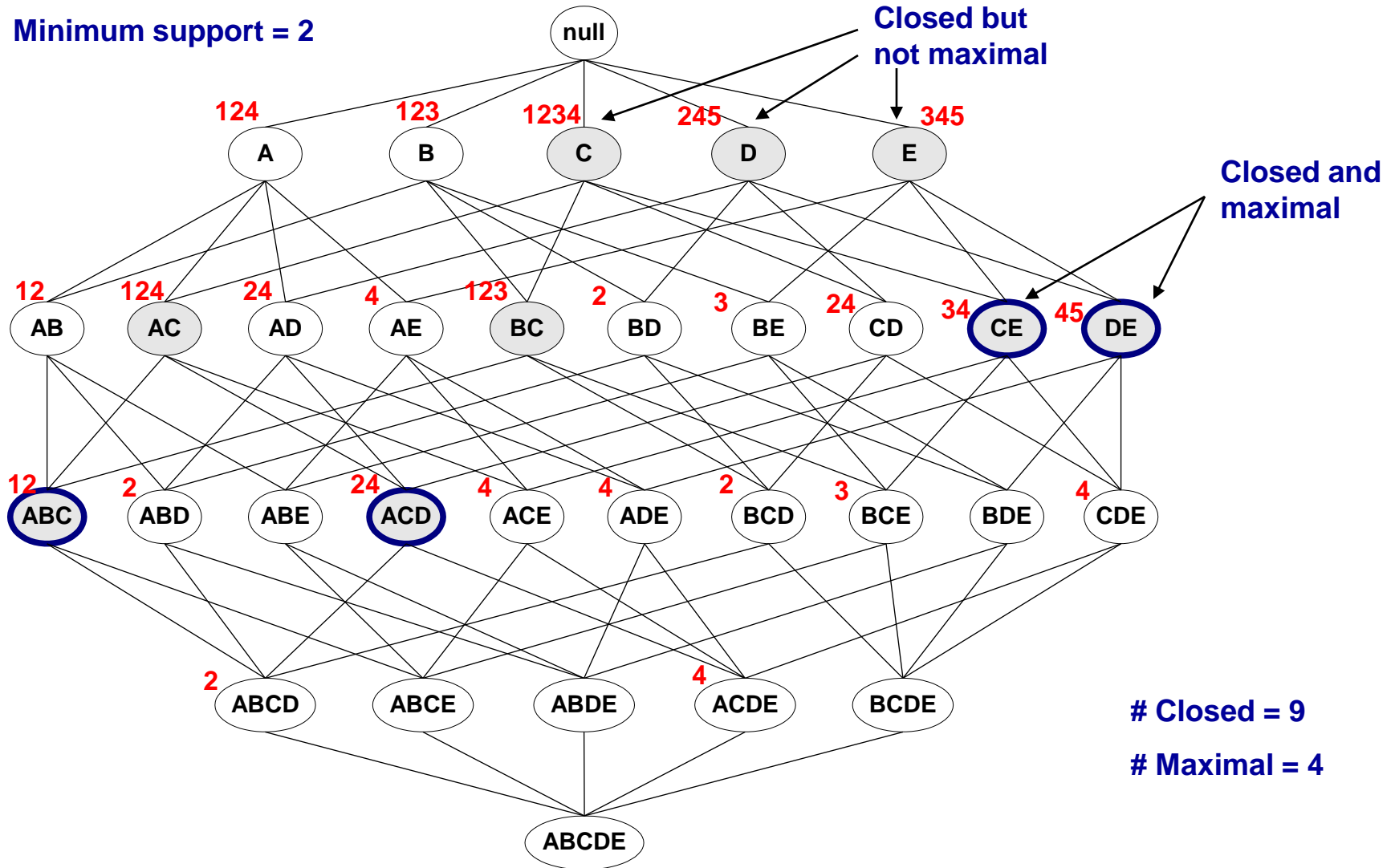
TẬP PHẦN TỬ ĐỒNG

TID	Items
1	ABC
2	ABCD
3	BCE
4	ACDE
5	DE



TẬP PHẦN TỬ ĐỒNG

Minimum support = 2



TẬP PHẦN TỬ ĐÓNG

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Itemsets	Support (counts)	Closed itemsets
{C}	3	
{D}	2	
{C,D}	2	

TẬP PHẦN TỬ ĐÓNG

		Items									
Transactions		A	B	C	D	E	F	G	H	I	J
	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Itemsets	Support (counts)	Closed itemsets
{C}	3	✓
{D}	2	
{C,D}	2	✓

TẬP PHẦN TỬ ĐỒNG

Transactions	Items									
	A	B	C	D	E	F	G	H	I	J
	1									
	2									
	3									
	4									
	5									
	6									
	7									
	8									
	9									
	10									

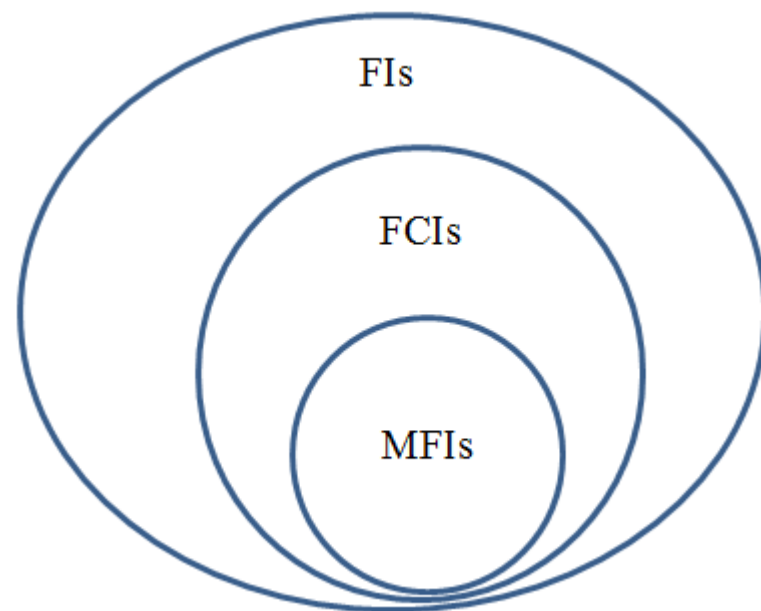
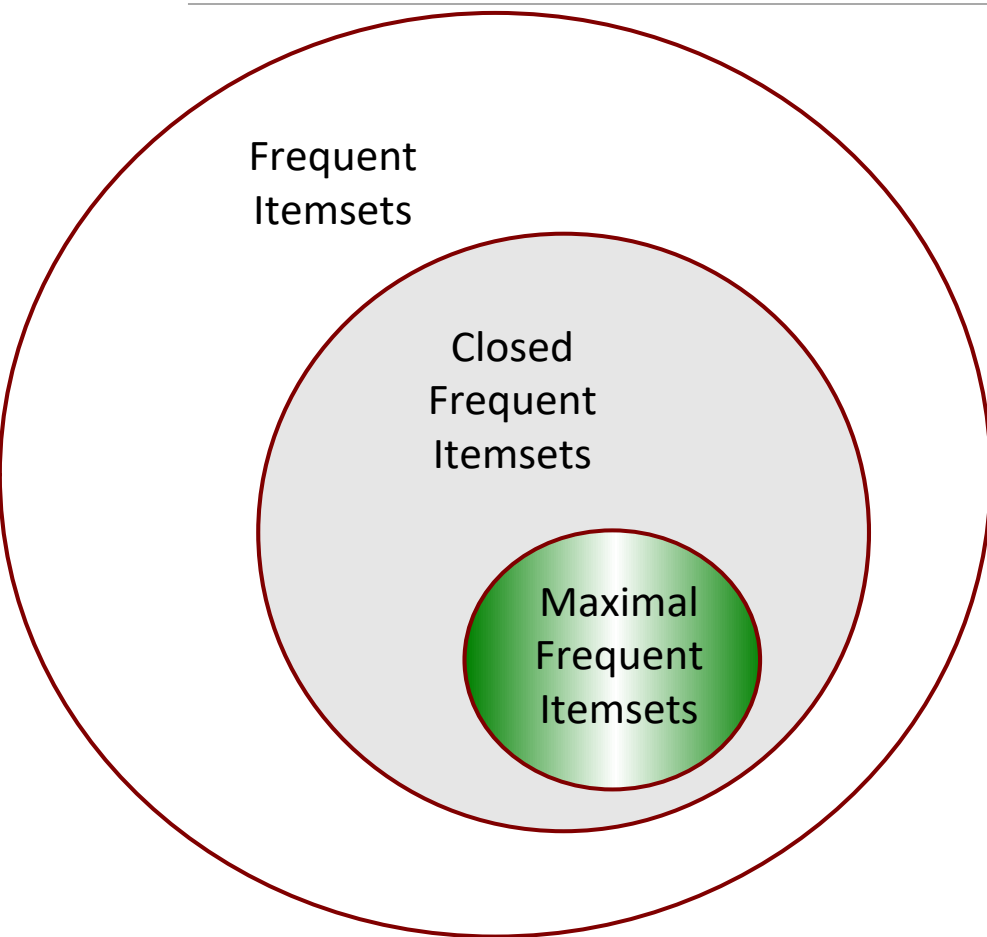
Itemsets	Support (counts)	Closed itemsets
{C}	3	
{D}	2	
{E}	2	
{C,D}	2	
{C,E}	2	
{D,E}	2	
{C,D,E}	2	

TẬP PHẦN TỬ ĐÓNG

Transactions	Items									
	A	B	C	D	E	F	G	H	I	J
	1									
	2									
	3									
	4									
	5									
	6									
	7									
	8									
	9									
	10									

Itemsets	Support (counts)	Closed itemsets
{C}	3	✓
{D}	2	
{E}	2	
{C,D}	2	
{C,E}	2	
{D,E}	2	
{C,D,E}	2	✓

I. Giới thiệu (Quan hệ giữa các phần tử)



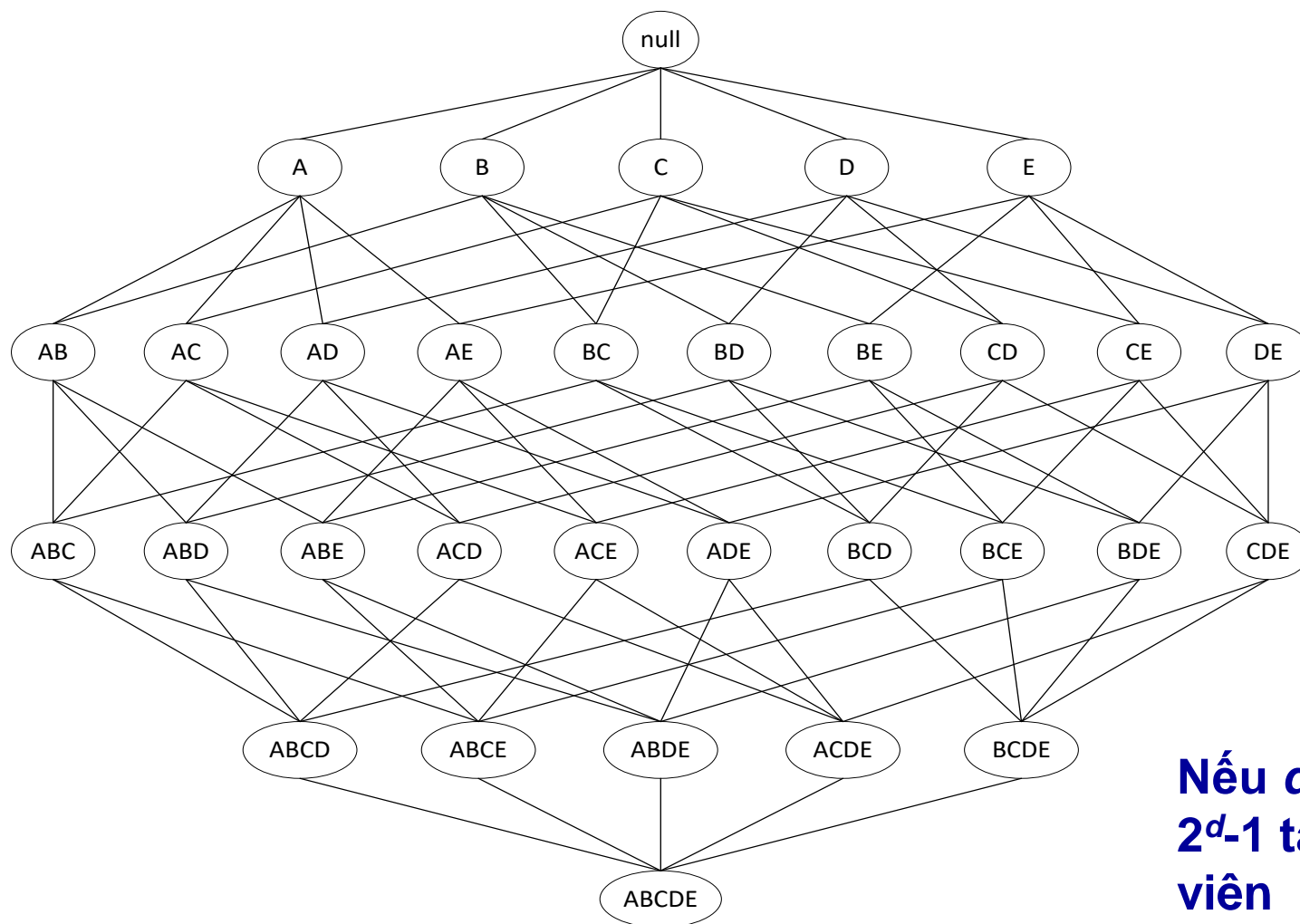


I. Giới thiệu

Bài toán khai thác luật kết hợp được chia làm hai giai đoạn:

1. Khai thác tập phổ biến (FIs – Frequent Itemsets)
2. Khai thác luật từ các tập phổ biến (ARs – Association Rules)

I. Giới thiệu



Nếu d phần tử, thì có $2^d - 1$ tập phần tử ứng viên



II. Các TT khai thác tập phổ biến

Bài toán khai thác tập phổ biến (frequent itemset) là bài toán rất quan trọng trong lĩnh vực data mining.

Bài toán khai thác tập phổ biến là bài toán tìm tất cả tập các hạng mục (itemset) S có độ phổ biến (support) thỏa mãn độ phổ biến tối thiểu minsupp :
 $\text{supp}(S) \geq \text{minsupp}$.



II. Các TT khai thác tập phổ biến

Một số thuật toán khai thác tập phổ biến:

- Thuật toán Apriori
- Thuật toán Eclat
- Thuật toán FP-Tree
- Thuật toán IT-tree
- Các thuật toán khác như LCM, DCI, PrePost, ...



II.1 Thuật toán Apriori

Được đề xuất bởi Agrawal et al năm 1993. Mục đích là tìm mối liên hệ giữa các mặt hàng được bán trong siêu thị.

Thuật toán Apriori là một thuật toán điển hình áp dụng trong khai phá luật kết hợp. Thuật toán dựa trên nguyên lý Apriori “*tập con bất kỳ của một tập phổ biến cũng là một tập phổ biến và mọi tập cha của tập không phổ biến là không phổ biến*”.



II.1 Thuật toán Apriori

Mục đích của thuật toán Apriori là tìm ra được tất cả các tập phổ biến có thể có trong cơ sở dữ liệu giao dịch D.



II.1 Thuật toán Apriori

Ý nghĩa nguyên lý Apriori:

1. Mọi tập con của tập phổ biến đều phổ biến, nghĩa là $\forall X \subseteq Y$, nếu $\text{supp}(Y) \geq \text{minsupp}$ thì $\text{supp}(X) \geq \text{minsupp}$.
2. Mọi tập cha của tập không phổ biến đều không phổ biến. nghĩa là $\forall X \subseteq Y$, nếu $\text{supp}(X) \leq \text{minsupp}$ thì $\text{supp}(Y) \leq \text{minsupp}$.



II.1 Thuật toán Apriori

Thuật toán Apriori

- **Input:** CSDL giao dịch T và ngưỡng phổ biến minsupp.
- **Output:** FIs chứa tất cả các tập phổ biến của T.



II.1 Thuật toán Apriori

Ý tưởng thuật toán Apriori

1. Tìm tất cả các tập phổ biến 1 hạng mục (L_1)
2. Tạo các tập ứng viên kích thước k hạng mục C_k từ các tập phổ biến L_{k-1} hạng mục (*nguyên lý Apriori*). Ví dụ, tạo ứng viên C_2 từ tập phổ biến L_1 .
3. Kiểm tra độ phổ biến của các ứng viên C_k trên CSDL và loại các ứng viên không phổ biến ta được $L_k (i=1,2,..., k)$.
4. Dừng khi không tạo được tập phổ biến hay tập ứng viên.



II.1 Thuật toán Apriori

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Beer, Bread, Diaper, Eggs
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Bread, Coke, Diaper, Milk



Items (1-itemsets)

Item	Count
Bread	4
Coke	2
Milk	4
Beer	3
Diaper	4
Eggs	1

Minimum Support = 3



II.1 Thuật toán Apriori

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Beer, Bread, Diaper, Eggs
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Bread, Coke, Diaper, Milk



Items (1-itemsets)

Item	Count
Bread	4
Coke	2
Milk	4
Beer	3
Diaper	4
Eggs	1

Minimum Support = 3



II.1 Thuật toán Apriori

Item	Count
Bread	4
Coke	2
Milk	4
Beer	3
Diaper	4
Eggs	1

Items (2-itemsets)



Itemset
{Bread, Milk}
{Bread, Beer }
{Bread, Diaper}
{Beer, Milk}
{Diaper, Milk}
{Beer, Diaper}

Pairs (2-itemsets)

**Không cần tạo ứng viên
chứa Coke hoặc Eggs**

Minimum Support = 3



II.1 Thuật toán Apriori

Items (2-itemsets)

Item	Count
Bread	4
Coke	2
Milk	4
Beer	3
Diaper	4
Eggs	1



Itemset	Count
{Bread, Milk}	3
{Beer, Bread}	2
{Bread, Diaper}	3
{Beer, Milk}	2
{Diaper, Milk}	3
{Beer, Diaper}	3

Minimum Support = 3



II.1 Thuật toán Apriori

Item	Count
Bread	4
Coke	2
Milk	4
Beer	3
Diaper	4
Eggs	1

Items (3-itemsets)



Itemset	Count
{Bread,Milk}	3
{Bread,Beer}	2
{Bread,Diaper}	3
{Milk,Beer}	2
{Milk,Diaper}	3
{Beer,Diaper}	3



Triplets (3-itemsets)

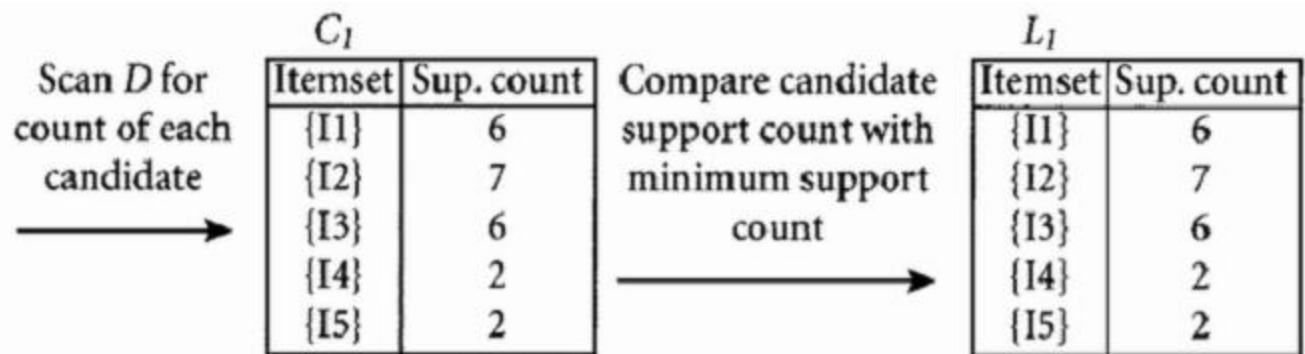
Minimum Support = 3

Itemset
{Bread, Diaper, Milk}

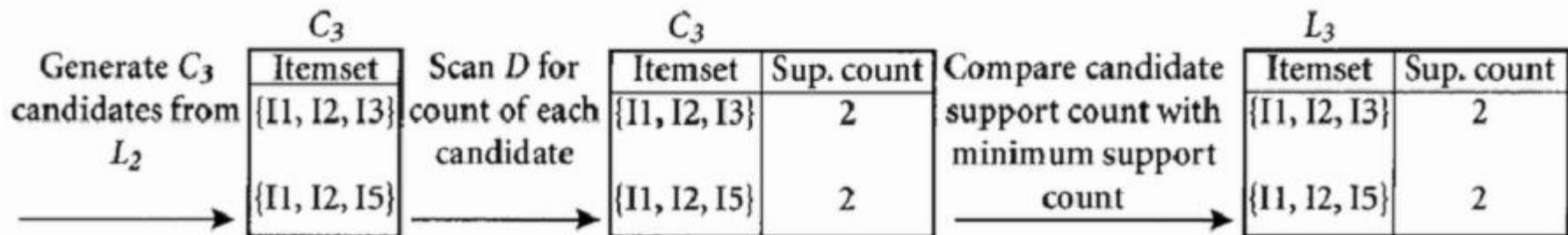
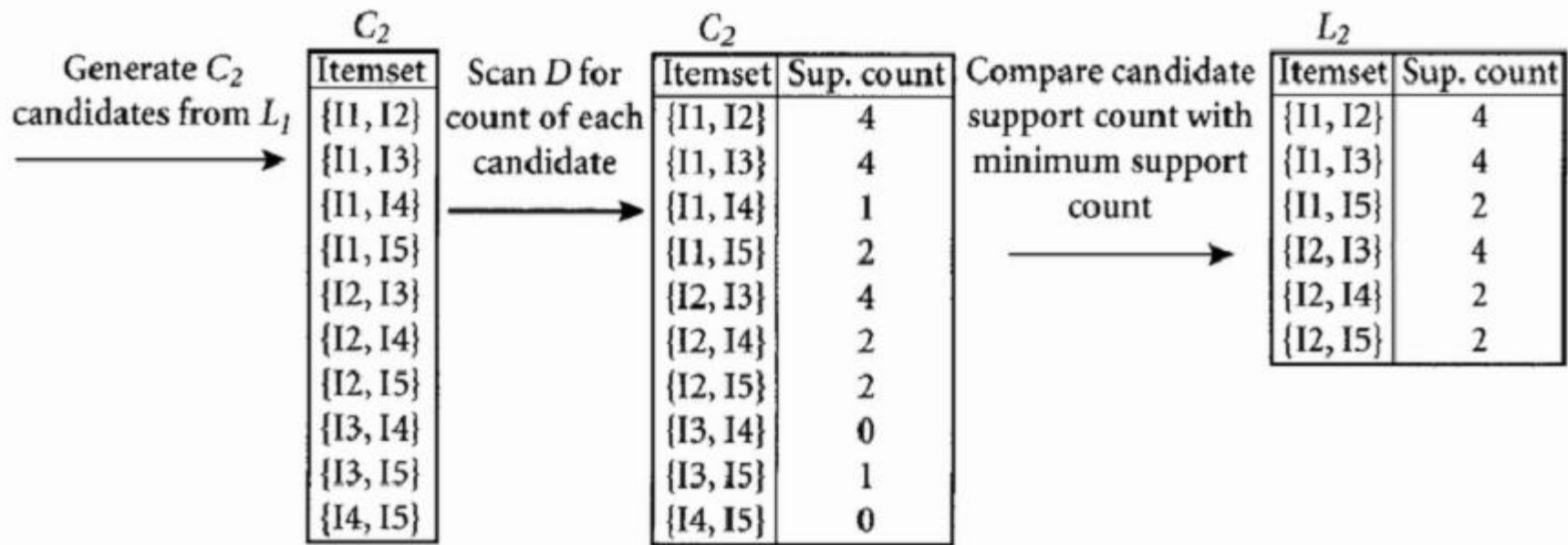
THUẬT TOÁN Apriori – Ví dụ

Cho *minimum support* là 20%, count = 1

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3



THUẬT TOÁN Apriori – Ví dụ





II.1 Thuật toán Apriori

- **Mã giả:**

Gọi C_k là tập các ứng viên có kích thước k

L_k các tập phổ biến có kích thước k

$L_1 = \{i \in I : \text{supp}(i) \geq \text{minsupp}\}$ // các phần tử phổ biến

for ($k = 1; L_k \neq \emptyset; k++$) do

$C_{k+1} = \text{apriori_gen}(L_k)$ // tập ứng viên được tạo từ L_k

for each $t \in T$ do // mỗi giao dịch t trong CSDL

for each $c \in C_{k+1}$ do

if $c \subseteq t$ then $c.\text{count}++$ // đếm ứng viên trong tập C_{k+1} mà chứa trong t

$L_{k+1} = \{c \in C_{k+1} \mid c.\text{count} \geq \text{minSup}\}$ // các ứng viên trong C_{k+1} có min_support

$FIs = \bigcup_k L_{k+1}$;



II.1 Thuật toán Apriori

Hàm apriori_gen:

- Input: L_k tập phổ biến kích thước k
- Output: C_{k+1} tập ứng viên kích thước $k+1$

THUẬT TOÁN Apriori

- Cách tạo các ứng viên: Tạo tập ứng viên C_{k+1} hạng mục từ tập L_k hạng mục:
 - Bước 1: tự kết hợp tập L_k
 - Bước 2: cắt bỏ
- Tính độ hỗ trợ của các ứng viên
- Ví dụ về tạo ứng viên
 - $L_3 = \{abc, abd, acd, ace, bcd\}$
 - Tự kết hợp: $L_3 * L_3$
 - $abcd$ từ abc và abd
 - $acde$ từ acd và ace
 - Cắt bỏ:
 - $acde$ được loại bỏ vì ade không có trong tập L_3

II.1 Thuật toán Apriori

- **Ví dụ:** Cho CSDL sau và minsupp = 50% (count = 2)

Mã giao dịch (TID)	Nội dung giao dịch
1	A, C, D
2	B, C, E
3	A, B, C, E
4	B, E



C_1	
Itemset	Supp
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3



L_1	
Itemset	Supp
{A}	2
{B}	2
{C}	3
{E}	3

II.1 Thuật toán Apriori

Itemset	Supp
{A}	2
{B}	2
{C}	3
{E}	3



C_2

Itemset	Supp
{A, B}	1
{A, C}	2
{A, E}	1
{B, C}	2
{B, E}	3
{C, E}	2



L_2

Itemset	Supp
{A, C}	2
{B, C}	2
{B, E}	3
{C, E}	2

II.1 Thuật toán Apriori

Itemset	Supp
{A, C}	2
{B, C}	2
{B, E}	3
{C, E}	2



C_3

Itemset	Supp
{B, C, E}	2



L_3

Itemset	Supp
{B, C, E}	2

- Chú ý: {A, C, B} không có trong C_3 vì {A, B} không có trong L_2 .



II.1 Thuật toán Apriori

- Các tập phổ biến tìm được là:

$FIs = \{\{A\}, \{B\}, \{C\}, \{E\}, \{A, C\}, \{B, C\}, \{B, E\}, \{C, E\}, \{B, C, E\}\}.$



II.1 Thuật toán Apriori

- Thuật toán Apriori là thuật toán dễ hiểu, dễ cài đặt. Tuy nhiên có các nhược điểm sau:
 - Phải duyệt CSDL nhiều lần.
 - Số lượng tập ứng viên rất lớn ($=2^{|T|} - 1$).
 - Thực hiện việc tính độ phổ biến nhiều, đơn điệu.



II.1 Thuật toán Apriori

- Cải tiến Apriori: Ý tưởng chung:
 - Giảm số lần duyệt CSDL.
 - Giảm số lượng (không sinh) tập ứng viên.
 - Tìm qui trình tính độ phổ biến hiệu quả hơn.



Bài tập

Bài 1: Từ CSDL giao dịch, Cho minsupp = 50% (minsup = 3)
hãy tìm:

- a) Tập phổ biến 1 phần tử
- b) Tất cả tập phổ biến bằng Apriori
- c) Tìm tập phổ biến tối đại.
- d) Tìm tập phổ biến đóng.

Mã giao dịch	Nội dung giao dịch
1	A, C, T, W
2	C, D, W
3	A, C, T, W
4	A, C, D, W
5	A, C, D, T, W
6	C, D, T



Bài tập

Bài 2. Cho cơ sở dữ liệu giao dịch như sau:

Sử dụng các giá trị ngưỡng $\text{minsupport} = 30\%$ ($\text{minsup} = 2,4$)

- a) Hãy liệt kê tất cả các tập phổ biến 1 phần tử
- b) Chạy từng bước thuật toán Apriori tìm tất cả tập phổ biến.
- c) Tìm tất cả các tập MFIs, FCIs trong cơ sở dữ liệu
- d) Xây dựng cây FP-Tree cho CSDL giao dịch trên

TID	Items
T01	A, B, C, D
T02	A, C, D, F
T03	C, D, E, G, A
T04	A, D, F, B
T05	B, C, G
T06	D, F, G
T07	A, B, G
T08	C, D, F, G



Bài tập

Bài 3. Cho cơ sở dữ liệu giao dịch như sau:

Sử dụng các ngưỡng support = 30%

- a) Hãy liệt kê tất cả các tập phổ biến 1 phần tử
- b) Chạy từng bước thuật toán Apriori tìm tất cả tập phổ biến.
- c) Tìm các tập phổ biến đóng (FCIs)
- d) Xây dựng cây FP-Tree cho CSDL giao dịch trên

TID	Items
T01	A1, B1, C2
T02	A2, C1, D1
T03	B2, C2, E2
T04	B1, C1, E1
T05	A3, C3, E2
T06	C1, D2, E2