

Review 3: Association Rule Mining



한양대학교 ERICA
소프트웨어융합대학
COLLEGE OF COMPUTING

인공지능학과
Department of
Artificial Intelligence

정 우 환 (whjung@hanyang.ac.kr)

Fall 2021

Association Rules





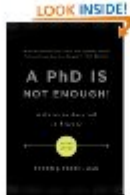
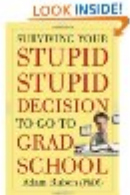

- 데이터 상호간의 연관 규칙
- Example) '{라면, 우유}->{커피}'
 - 라면과 우유를 산 사람은 커피도 같이 산다

사용 사례

- 고객들의 물품 구매 패턴을 분석한 결과에 기반하여
 - 연관 물품 쿠폰이나 할인 행사 제공
 - 온라인 서점에서 다른 구매자들이 구매한 책 정보를 함께 제공

Customers Who Bought This Item Also Bought



			
Graduate Admissions Essays, Fourth ... Donald Asher ★★★★☆ (9) Paperback \$14.95	A PhD Is Not Enough!: A Guide to Survival in ... > Peter J. Feibelman ★★★★☆ (58) Paperback \$10.17	Surviving Your Stupid, Stupid Decision to Go ... > Adam Ruben ★★★★☆ (42) Paperback \$9.60	Get Into Graduate School Kaplan ★★★★★ (2) Paperback \$13.33

<예: 아마존(amazon.com)의 상품 추천>

문제정의) Association Rule Mining

TID	Items
10	a, c, d
20	b, c, e
30	a, b, c, e
40	b, e

- Given:
 - A database of customer transactions D
 - Each transaction is a set of items
 - $MinSupp$: minimum support
 - $MinConf$: minimum confidence

- Find all rules $X \rightarrow Y$ such that

- $Support(X) \geq MinSupp$

$$Support(X) = \frac{\# \text{ transactions that contains } X}{\# \text{ transactions in } D}$$

- $Confidence(X \rightarrow Y) \geq MinConf$

$$Confidence(X \rightarrow Y) = \frac{\# \text{ transactions that contains } X \text{ and } Y}{\# \text{ transactions that contains } X} = \frac{Support(X \cup Y)}{Support(X)}$$

$$Support(\{b\}) = 0.75$$

$$Support(\{b, c\}) = 0.5$$

$$Confidence(\{b\} \rightarrow \{c\}) = 0.66$$

Association Rule Mining

- Consists of 2 steps
 - Step 1: **Find all frequent itemsets** that have minimum support
 - Find all X such that $Support(X) \geq MinSupp$
 - Most expensive phase
 - Lots of research
 - Step 2: **Find all rules** by using the frequent itemsets
 - Find all $X \rightarrow Y$ such that $Confidence(X \rightarrow Y) \geq MinConf$
 - Straightforward

**Step 1: Find all frequent
itemsets**

Naïve Counting of All Itemsets

Itemsets & Counts

Transactions

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



Itemset	Count
A	1
C	1
D	1
A,C	1
A,D	1
C,D	1
A,C,D	1

Naïve Counting of All Itemsets

Itemsets & Counts

Transactions

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



Itemset	Count
A	1
C	2
D	1
A,C	1
A,D	1
C,D	1
A,C,D	1
B	1
E	1
B,C	1
B,E	1
C,E	1
B,C,E	1

Naïve Counting of All Itemsets

Transactions

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



Itemsets & Counts

Itemset	Count	Itemset	Count
A	2	A,B	1
C	3	A,E	1
D	1	A,B,C	1
A,C	2	A,B,E	1
A,D	1	A,B,C,E	1
C,D	1		
A,C,D	1		
B	2		
E	2		
B,C	2		
B,E	2		
C,E	2		
B,C,E	2		

Naïve Counting of All Itemsets

Transactions

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



Itemsets & Counts

Itemset	Count	Itemset	Count
A	2	A,B	1
C	3	A,E	1
D	1	A,B,C	1
A,C	2	A,B,E	1
A,D	1	A,B,C,E	1
C,D	1		
A,C,D	1		
B	3		
E	3		
B,C	2		
B,E	3		
C,E	2		
B,C,E	2		

Naïve Counting of All Itemsets

Transactions

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

Sup_{min}=2

Itemset

Count

A	2
C	3
D	1
A,C	2
A,D	1
C,D	1
A,C,D	1
B	3
E	3
B,C	2
B,E	3
C,E	2
B,C,E	2

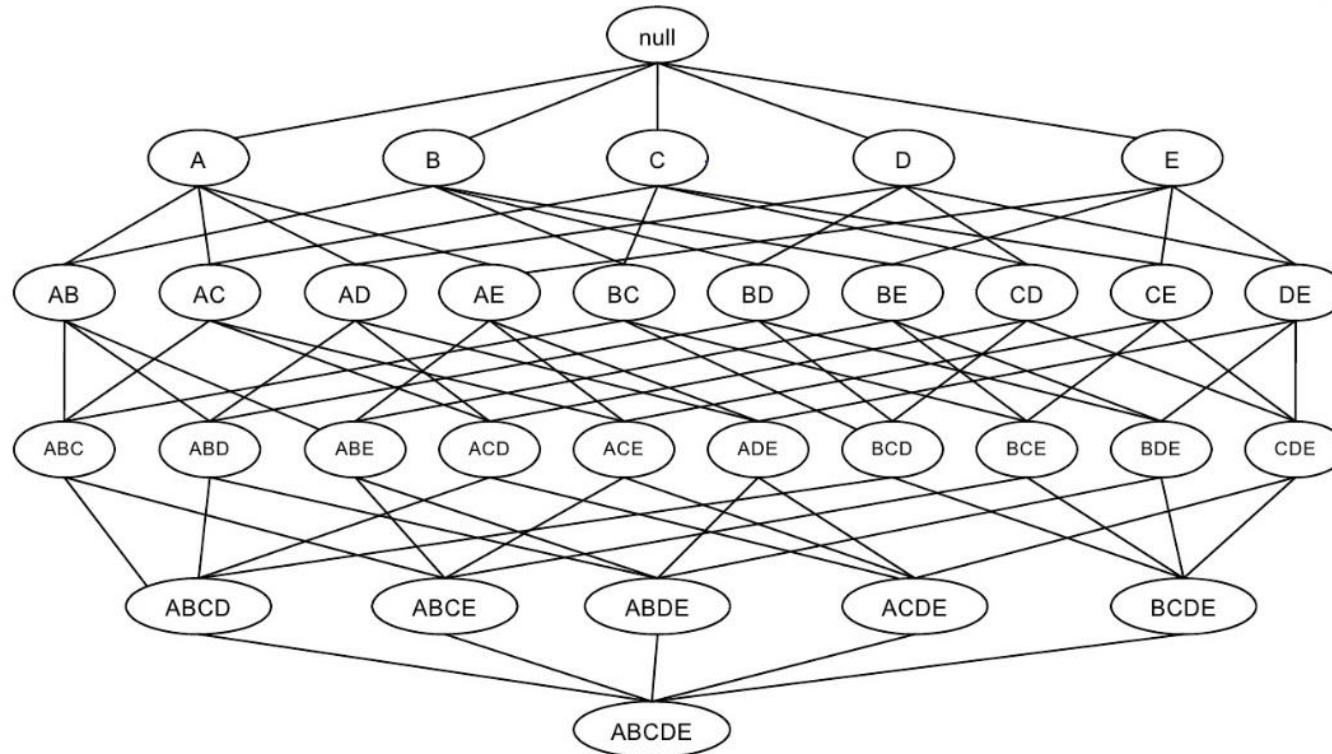
Frequent itemsets

Itemset	Count
A,B	1
A,E	1
A,B,C	1
A,B,E	1
A,B,C,E	1

We may need 2^n itemset entries for counts !

Naïve Counting

- Given d items, there are 2^d itemsets



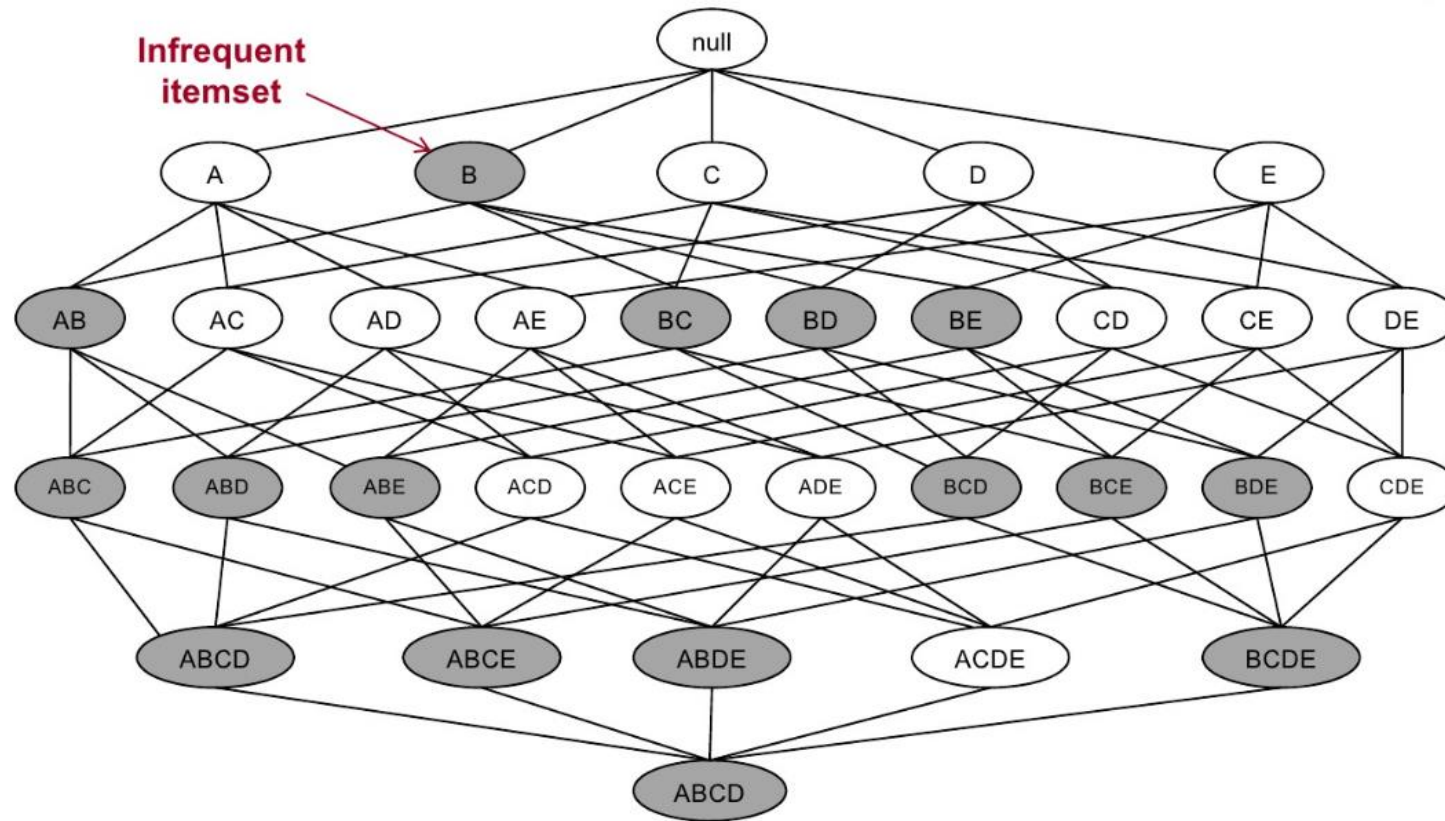
Apriori: A Candidate Generation-and-Test Approach

- Apriori pruning principle: If there is any itemset which is infrequent, its superset should not be generated/tested!

(Agrawal & Srikant @VLDB'94, Mannila, et al. @ KDD' 94)

- Method:
 - Initially, scan DB once to get frequent 1-itemset
 - Generate length $(k+1)$ candidate itemsets from length k frequent itemsets
 - Test the candidates against DB
 - Terminate when no frequent or candidate set can be generated

Candidate Itemset Generation by Apriori



An Apriori Example

