

Knowledge Discovery and Data Mining

Unit # 15

Acknowledgement

- Most of the slides in this presentation are taken from course slides provided by
 - Han and Kimber (Data Mining Concepts and Techniques) and
 - Tan, Steinbach and Kumar (Introduction to Data Mining)

Association Rule Mining

- Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction

Market-Basket transactions

<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

Example of Association Rules

{Diaper} → {Beer},
 {Milk, Bread} → {Eggs,Coke},
 {Beer, Bread} → {Milk},

Implication means co-occurrence, not causality!

Definition: Frequent Itemset

- Itemset**
 - A collection of one or more items
 - Example: {Milk, Bread, Diaper}
 - k-itemset
 - An itemset that contains k items
- Support count (σ)**
 - Frequency of occurrence of an itemset
 - E.g. $\sigma(\{\text{Milk, Bread, Diaper}\}) = 2$
- Support**
 - Fraction of transactions that contain an itemset
 - E.g. $s(\{\text{Milk, Bread, Diaper}\}) = 2/5$
- Frequent Itemset**
 - An itemset whose support is greater than or equal to a *minsup* threshold

<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

Definition: Frequent Itemset

Definition: Association Rule

- Association Rule

- An implication expression of the form $X \rightarrow Y$, where X and Y are itemsets
- Example:
 $\{\text{Milk, Diaper}\} \rightarrow \{\text{Beer}\}$

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

- Rule Evaluation Metrics

- Support (s)
 - ◆ Fraction of transactions that contain both X and Y
- Confidence (c)
 - ◆ Measures how often items in Y appear in transactions that contain X

Example:

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

$$s = \frac{\sigma(\text{Milk, Diaper, Beer})}{|T|} = \frac{2}{5} = 0.4$$

$$c = \frac{\sigma(\text{Milk, Diaper, Beer})}{\sigma(\text{Milk, Diaper})} = \frac{2}{3} = 0.67$$

Association Rule Mining Task

- Given a set of transactions T , the goal of association rule mining is to find all rules having
 - support $\geq \textit{minsup}$ threshold
 - confidence $\geq \textit{minconf}$ threshold
- Brute-force approach:
 - List all possible association rules
 - Compute the support and confidence for each rule
 - Prune rules that fail the *minsup* and *minconf* thresholds

\Rightarrow **Computationally prohibitive!**

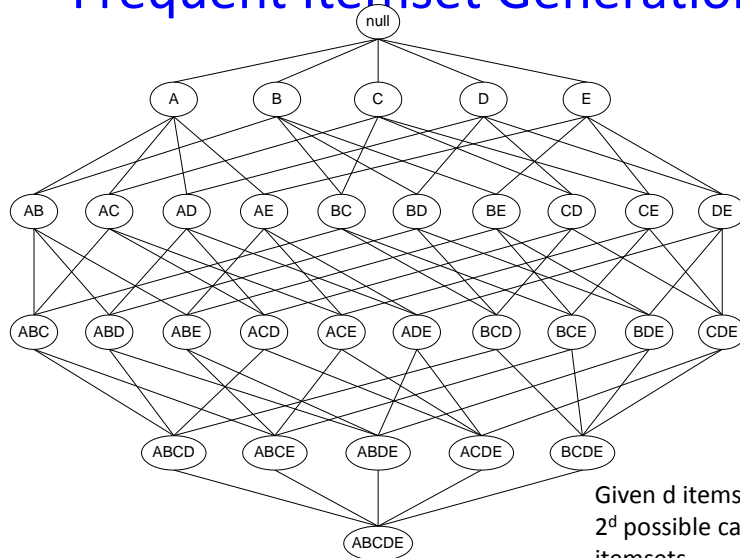
Apriori

- Apriori is a seminal algorithm proposed by R. Agrawal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules.
- The name of the algorithm is based on the fact that the algorithm uses *prior knowledge of frequent itemset properties*.
- Two steps process
 - Join
 - Prune

Mining Association Rules

- Two-step approach:
 1. **Frequent Itemset Generation**
 - Generate all itemsets whose support \geq minsup
 2. **Rule Generation**
 - Generate high confidence rules from each frequent itemset, where each rule is a binary partitioning of a frequent itemset
- Frequent itemset generation is still computationally expensive

Frequent Itemset Generation



Sajjad Haider

Fall 2011

9

Reducing Number of Candidates

- **Apriori principle:**
 - If an itemset is frequent, then all of its subsets must also be frequent
- Apriori principle holds due to the following property of the support measure:

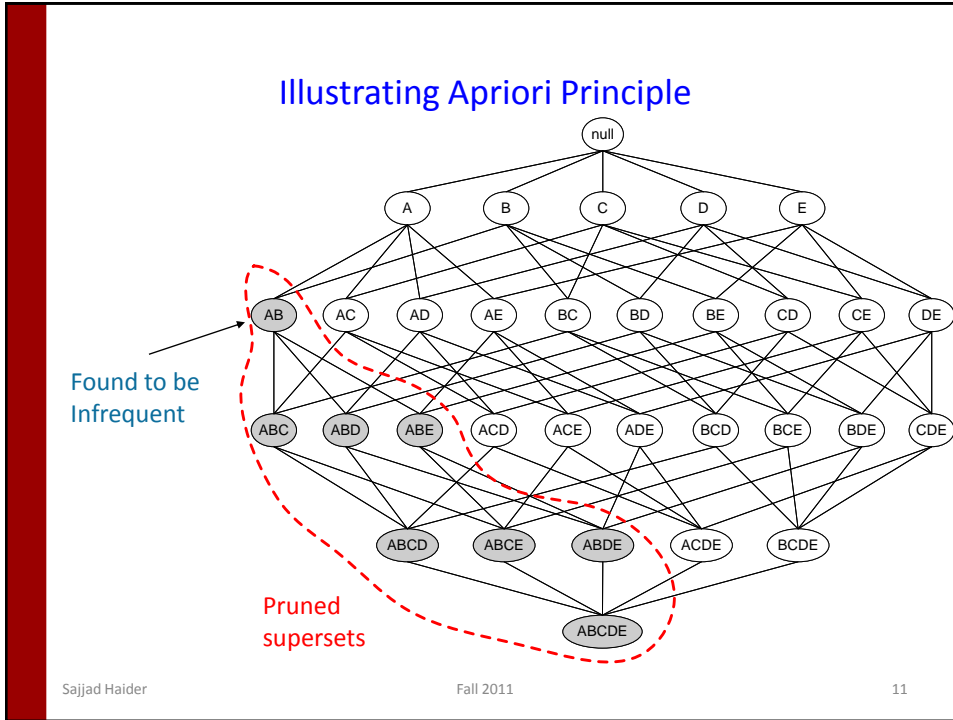
$$\forall X, Y : (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$$

- Support of an itemset never exceeds the support of its subsets
- This is known as the **anti-monotone** property of support

Sajjad Haider

Fall 2011

10



Illustrating Apriori Principle

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

Items (1-itemsets)

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

Pairs (2-itemsets)
(No need to generate candidates involving Coke or Eggs)

Itemset	Count
{Bread,Milk,Diaper}	3

Triplets (3-itemsets)

Minimum Support = 3

If every subset is considered,
 ${}^6C_1 + {}^6C_2 + {}^6C_3 = 41$
 With support-based pruning,
 $6 + 6 + 1 = 13$

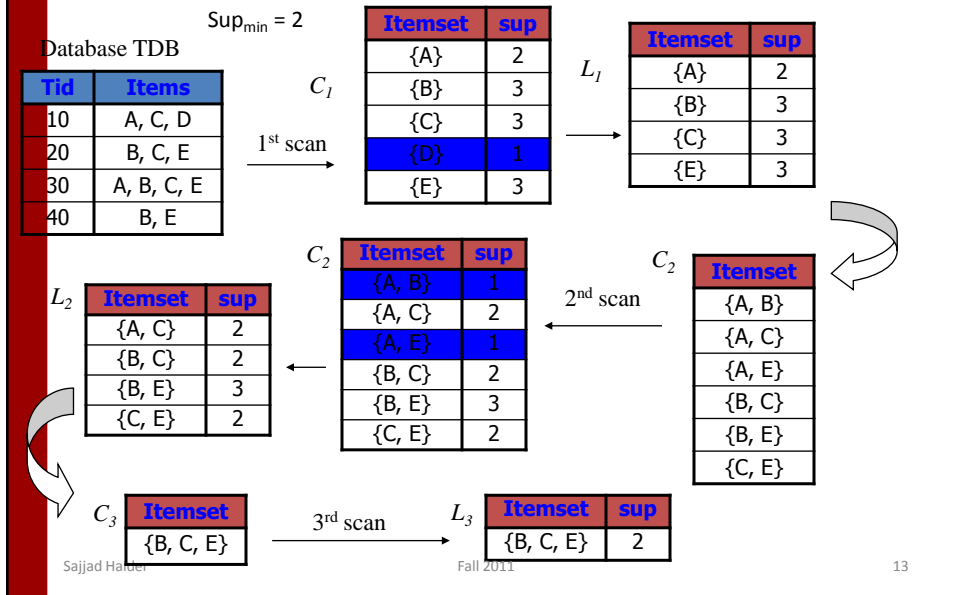
...

Sajjad Haider

Fall 2011

12

The Apriori Algorithm—An Example



Apriori Algorithm

- Method:
 - Let $k=1$
 - Generate frequent itemsets of length 1
 - Repeat until no new frequent itemsets are identified
 - Generate length $(k+1)$ candidate itemsets from length k frequent itemsets
 - Prune candidate itemsets containing subsets of length k that are infrequent
 - Count the support of each candidate by scanning the DB
 - Eliminate candidates that are infrequent, leaving only those that are frequent

Factors Affecting Complexity

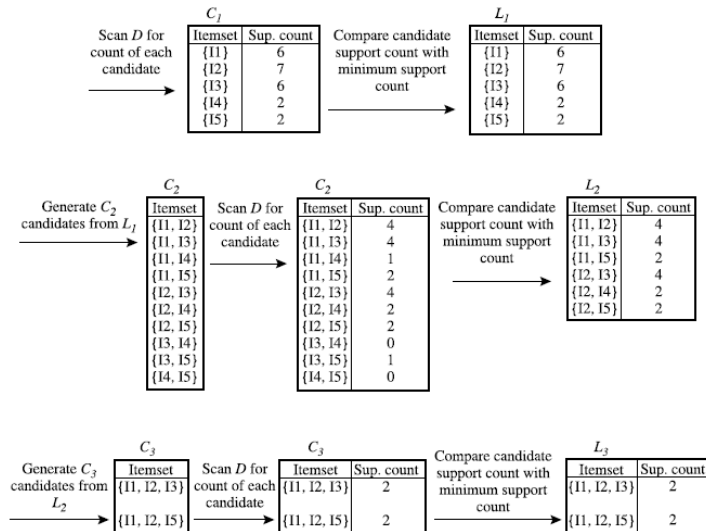
- Choice of minimum support threshold
 - lowering support threshold results in more frequent itemsets
 - this may increase number of candidates and max length of frequent itemsets
- Dimensionality (number of items) of the data set
 - more space is needed to store support count of each item
 - if number of frequent items also increases, both computation and I/O costs may also increase
- Size of database
 - since Apriori makes multiple passes, run time of algorithm may increase with number of transactions

Example

<i>TID</i>	<i>List of item IDs</i>
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

Minimum Support = 2

Example (Cont'd)



Sajjad Hai

17

Rule Generation

- Given a frequent itemset L , find all non-empty subsets $f \subset L$ such that $f \rightarrow L - f$ satisfies the minimum confidence requirement
 - If $\{A, B, C, D\}$ is a frequent itemset, candidate rules:

$ABC \rightarrow D,$	$ABD \rightarrow C,$	$ACD \rightarrow B,$	$BCD \rightarrow A,$
$A \rightarrow BCD,$	$B \rightarrow ACD,$	$C \rightarrow ABD,$	$D \rightarrow ABC$
$AB \rightarrow CD,$	$AC \rightarrow BD,$	$AD \rightarrow BC,$	$BC \rightarrow AD,$
$BD \rightarrow AC,$	$CD \rightarrow AB,$		
- If $|L| = k$, then there are $2^k - 2$ candidate association rules (ignoring $L \rightarrow \emptyset$ and $\emptyset \rightarrow L$)

Sajjad Haider

Fall 2011

18

Rule Generation

- How to efficiently generate rules from frequent itemsets?

- In general, confidence does not have an anti-monotone property
 $c(ABC \rightarrow D)$ can be larger or smaller than $c(AB \rightarrow D)$

- But confidence of rules generated from the same itemset has an anti-monotone property

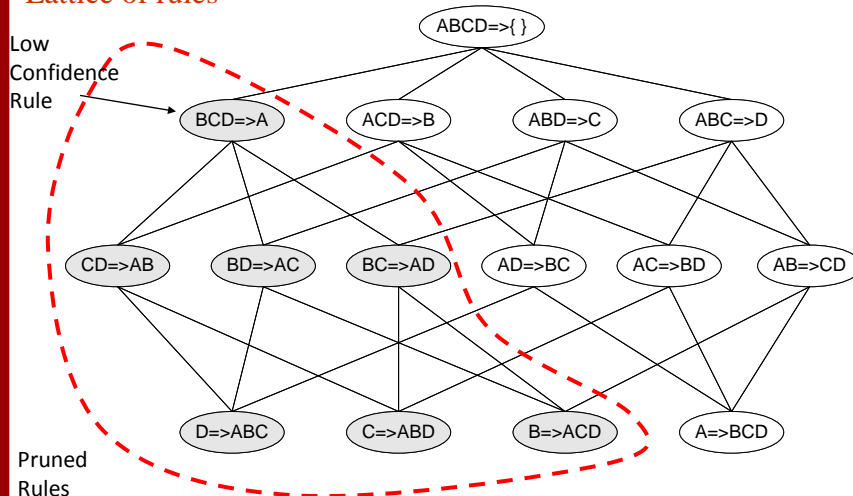
- e.g., $L = \{A,B,C,D\}$:

$$c(ABC \rightarrow D) \geq c(AB \rightarrow CD) \geq c(A \rightarrow BCD)$$

- Confidence is anti-monotone w.r.t. number of items on the RHS of the rule

Rule Generation for Apriori Algorithm

Lattice of rules



Rule Generation Example

- Suppose the data contains the frequent itemset $I = \{I1, I2, I5\}$. What are the association rules that can be generated from I ?
- The nonempty subsets of I are $\{I1, I2\}$, $\{I1, I5\}$, $\{I2, I5\}$, $\{I1\}$, $\{I2\}$, and $\{I5\}$.
- If the minimum confidence threshold is, say, 70%, then only the second, third, and last rules below are output

$I1 \wedge I2 \Rightarrow I5,$	$confidence = 2/4 = 50\%$
$I1 \wedge I5 \Rightarrow I2,$	$confidence = 2/2 = 100\%$
$I2 \wedge I5 \Rightarrow I1,$	$confidence = 2/2 = 100\%$
$I1 \Rightarrow I2 \wedge I5,$	$confidence = 2/6 = 33\%$
$I2 \Rightarrow I1 \wedge I5,$	$confidence = 2/7 = 29\%$
$I5 \Rightarrow I1 \wedge I2,$	$confidence = 2/2 = 100\%$

Effect of Support Distribution

- How to set the appropriate *minsup* threshold?
 - If *minsup* is set too high, we could miss itemsets involving interesting rare items (e.g., expensive products)
 - If *minsup* is set too low, it is computationally expensive and the number of itemsets is very large

Pattern Evaluation

- Association rule algorithms tend to produce too many rules
 - many of them are uninteresting or redundant
 - Redundant if $\{A,B,C\} \rightarrow \{D\}$ and $\{A,B\} \rightarrow \{D\}$ have same support & confidence
- Interestingness measures can be used to prune/rank the derived patterns
- In the original formulation of association rules, support & confidence are the only measures used

Computing Interestingness Measure

- Given a rule $X \rightarrow Y$, information needed to compute rule interestingness can be obtained from a contingency table

Contingency table for $X \rightarrow Y$

	Y	\bar{Y}	
X	f_{11}	f_{10}	f_{1+}
\bar{X}	f_{01}	f_{00}	f_{0+}
	f_{+1}	f_{+0}	T

f_{11} : support of X and Y

f_{10} : support of X and \bar{Y}

f_{01} : support of \bar{X} and Y

f_{00} : support of \bar{X} and \bar{Y}

Drawback of Confidence

	Coffee	$\overline{\text{Coffee}}$	
Tea	15	5	20
$\overline{\text{Tea}}$	75	5	80
	90	10	100

Association Rule: Tea \rightarrow Coffee

Confidence = $P(\text{Coffee}|\text{Tea}) = 0.75$

but $P(\text{Coffee}) = 0.9$

\Rightarrow Although confidence is high, rule is misleading

$\Rightarrow P(\text{Coffee}|\overline{\text{Tea}}) = 0.9375$

Statistical Independence

- Population of 1000 students
 - 600 students know how to swim (S)
 - 700 students know how to bike (B)
 - 420 students know how to swim and bike (S,B)
 - $P(S \wedge B) = 420/1000 = 0.42$
 - $P(S) \times P(B) = 0.6 \times 0.7 = 0.42$
 - $P(S \wedge B) = P(S) \times P(B) \Rightarrow$ Statistical independence
 - $P(S \wedge B) > P(S) \times P(B) \Rightarrow$ Positively correlated
 - $P(S \wedge B) < P(S) \times P(B) \Rightarrow$ Negatively correlated

Statistical-based Measures

- Measures that take into account statistical dependence

$$Lift = \frac{P(Y | X)}{P(Y)}$$

$$Interest = \frac{P(X, Y)}{P(X)P(Y)}$$

$$PS = P(X, Y) - P(X)P(Y)$$

$$\phi - \text{coefficient } t = \frac{P(X, Y) - P(X)P(Y)}{\sqrt{P(X)[1 - P(X)]P(Y)[1 - P(Y)]}}$$

Example: Lift/Interest

	Coffee	<u>Coffee</u>	
Tea	15	5	20
<u>Tea</u>	75	5	80
	90	10	100

Association Rule: Tea → Coffee

Confidence = $P(\text{Coffee} | \text{Tea}) = 0.75$

but $P(\text{Coffee}) = 0.9$

⇒ Lift = $0.75/0.9 = 0.8333$ (< 1 , therefore is negatively associated)

#	Measure	Formula
1	ϕ -coefficient	$\frac{P(A,B) - P(A)P(B)}{\sqrt{P(A)P(B)(1-P(A))(1-P(B))}}$
2	Goodman-Kruskal's (λ)	$\frac{\sum_{j_1} \max_k P(A_{j_1}, B_k) - \max_k P(A_{j_1}) - \max_k P(B_k)}{2 - \max_k P(A_{j_1}) - \max_k P(B_k)}$
3	Odds ratio (α)	$\frac{P(A,B)P(\bar{A},\bar{B})}{P(\bar{A},B)P(A,\bar{B})}$
4	Yule's Q	$\frac{P(A,B)P(\bar{A},\bar{B}) - P(A,\bar{B})P(\bar{A},B)}{P(A,B)P(\bar{A},\bar{B}) + P(A,\bar{B})P(\bar{A},B)} = \frac{\alpha-1}{\alpha+1}$
5	Yule's Y	$\frac{\sqrt{P(A,B)P(\bar{A},\bar{B})} - \sqrt{P(A,\bar{B})P(\bar{A},B)}}{\sqrt{P(A,B)P(\bar{A},\bar{B})} + \sqrt{P(A,\bar{B})P(\bar{A},B)}} = \frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}$
6	Kappa (κ)	$\frac{P(A,B) + P(\bar{A},\bar{B}) - P(A)P(B) - P(\bar{A})P(\bar{B})}{1 - P(A)P(B) - P(\bar{A})P(\bar{B})}$
7	Mutual Information (M)	$\frac{\sum_i \sum_j P(A_i, B_j) \log \frac{P(A_i, B_j)}{P(A_i)P(B_j)}}{\min(-\sum_i P(A_i) \log P(A_i), -\sum_j P(B_j) \log P(B_j))}$
8	J-Measure (J)	$\max \left(P(A, B) \log \left(\frac{P(B A)}{P(B)} \right) + P(\bar{A}\bar{B}) \log \left(\frac{P(\bar{B} \bar{A})}{P(\bar{B})} \right), P(A, B) \log \left(\frac{P(A B)}{P(A)} \right) + P(\bar{A}\bar{B}) \log \left(\frac{P(\bar{A} \bar{B})}{P(\bar{A})} \right) \right)$
9	Gini index (G)	$\max \left(P(A)[P(B A)]^2 + P(\bar{B} A)]^2 + P(\bar{A})[P(B \bar{A})]^2 + P(\bar{B} \bar{A})]^2 - P(B)^2 - P(\bar{B})^2, P(\bar{B})[P(A \bar{B})]^2 + P(\bar{A} \bar{B})]^2 + P(A)[P(A B)]^2 + P(\bar{A} B)]^2 - P(A)^2 - P(\bar{A})^2 \right)$
10	Support (s)	$P(A, B)$
11	Confidence (c)	$\max(P(B A), P(A B))$
12	Laplace (L)	$\max \left(\frac{NP(A,B)+1}{NP(A)+2}, \frac{NP(A,B)+1}{NP(B)+2} \right)$
13	Conviction (V)	$\max \left(\frac{P(A)P(\bar{B})}{P(A\bar{B})}, \frac{P(\bar{B})P(A)}{P(\bar{B}A)} \right)$
14	Interest (I)	$\frac{P(A,B)}{P(A)P(B)}$
15	cosine (IS)	$\frac{P(A,B)}{\sqrt{P(A)P(B)}}$
16	Piatetsky-Shapiro's (PS)	$P(A, B) - P(A)P(B)$
17	Certainty factor (F)	$\max \left(\frac{P(B A) - P(B)}{1 - P(B)}, \frac{P(A B) - P(A)}{1 - P(A)} \right)$
18	Added Value (AV)	$\max(P(B A) - P(B), P(A B) - P(A))$
19	Collective strength (S)	$\frac{P(A,B) + P(\bar{A}\bar{B})}{P(A,B) + P(\bar{A}\bar{B})} \times \frac{1 - P(A)P(B) - P(\bar{A})P(\bar{B})}{1 - P(A,B) - P(\bar{A}\bar{B})}$
20	Jaccard (ζ)	$\frac{P(A) + P(B) - P(A,B)}{P(A,B)}$
21	Klogsen (K)	$\sqrt{P(\bar{A}, \bar{B})} \max(P(B A) - P(B), P(A B) - P(A))$

There are lots of measures proposed in the literature

Some measures are good for certain applications, but not for others

What criteria should we use to determine whether a measure is good or bad?

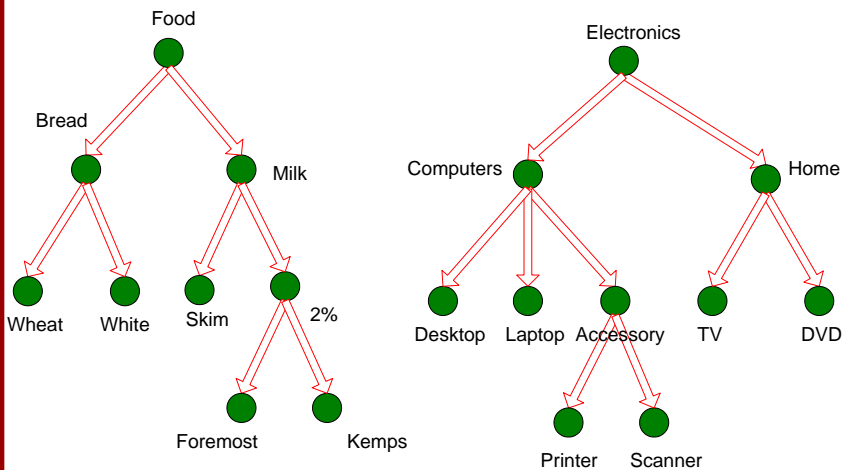
What about Apriori-style support based pruning? How does it affect these measures?

Sajjad Haider

Fall 2011

29

Multi-level Association Rules



Sajjad Haider

Fall 2011

30

Multi-level Association Rules

- Why should we incorporate concept hierarchy?
 - Rules at lower levels may not have enough support to appear in any frequent itemsets
 - Rules at lower levels of the hierarchy are overly specific
 - e.g., skim milk → white bread, 2% milk → wheat bread, skim milk → wheat bread, etc.
- are indicative of association between milk and bread

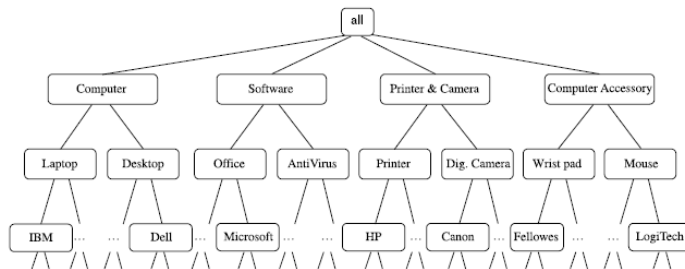
Sajjad Haider

Fall 2011

31

Example

<i>TID</i>	<i>Items Purchased</i>
T100	IBM-ThinkPad-T40/2373, HP-Photosmart-7660
T200	Microsoft-Office-Professional-2003, Microsoft-Plus!-Digital-Media
T300	Logitech-MX700-Cordless-Mouse, Fellowes-Wrist-Rest
T400	Dell-Dimension-XPS, Canon-PowerShot-S400
T500	IBM-ThinkPad-R40/P4M, Symantec-Norton-Antivirus-2003
...	...

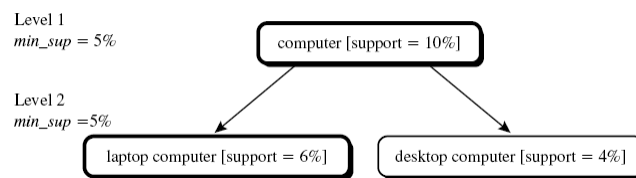


Concept Hierarchies

- Multilevel association rules can be mined efficiently using concept hierarchies under a support-confidence framework.
- In general, a top-down strategy is employed, where counts are accumulated for the calculation of frequent itemsets at each concept level, starting at the concept level 1 and working downward in the hierarchy toward the more specific concept levels, until no more frequent itemsets can be found.
- For each level, any algorithm for discovering frequent itemsets may be used, such as Apriori or its variations.

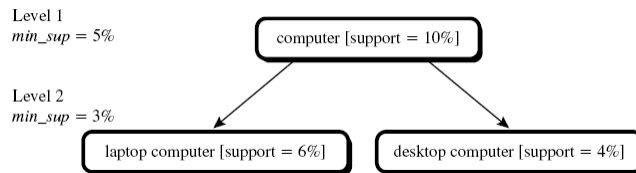
Uniform Minimum Support for All Levels

- The same minimum support threshold is used when mining at each level of abstraction.
- For example, a minimum support threshold of 5% is used throughout (e.g., for mining from “computer” down to “laptop computer”).
- Both “computer” and “laptop computer” are found to be frequent, while “desktop computer” is not.



Reduced Minimum Support at Lower Levels

- Each level of abstraction has its own minimum support threshold. The deeper the level of abstraction, the smaller the corresponding threshold is.
- For example, the minimum support thresholds for levels 1 and 2 are 5% and 3%, respectively.
- In this way, “computer,” “laptop computer,” and “desktop computer” are all considered frequent.



Sajjad Haider

Fall 2011

35

Continuous and Categorical Attributes

How to apply association analysis formulation to non-symmetric binary variables?

Session Id	Country	Session Length (sec)	Number of Web Pages viewed	Gender	Browser Type	Buy
1	USA	982	8	Male	IE	No
2	China	811	10	Female	Netscape	No
3	USA	2125	45	Female	Mozilla	Yes
4	Germany	596	4	Male	IE	Yes
5	Australia	123	9	Male	Mozilla	No
...

Example of Association Rule:

$$\{\text{Number of Pages} \in [5,10) \wedge (\text{Browser}=\text{Mozilla})\} \rightarrow \{\text{Buy} = \text{No}\}$$

Sajjad Haider

Fall 2011

36

Handling Categorical Attributes

- Transform categorical attribute into asymmetric binary variables.
- Introduce a new “item” for each distinct attribute-value pair
 - Example: replace Browser Type attribute with
 - Browser Type = Internet Explorer
 - Browser Type = Mozilla
 - Browser Type = Netscape
- What if attribute has many possible values
 - Example: attribute country has more than 200 possible values
 - Many of the attribute values may have very low support
 - Potential solution: Aggregate the low-support attribute values

Handling Continuous Attributes

- Different kinds of rules:
 - $\text{Age} \in [21,35) \wedge \text{Salary} \in [70\text{k},120\text{k}) \rightarrow \text{Buy}$
 - $\text{Salary} \in [70\text{k},120\text{k}) \wedge \text{Buy} \rightarrow \text{Age: } \mu=28, \sigma=4$
- Different methods:
 - Discretization-based
 - Statistics-based
 - Non-discretization based
 - minApriori

Discretization

- Discretization is the most common approach for handling continuous attributes.
- This approach groups the adjacent values of a continuous attribute into a finite number of intervals.
- The discrete intervals are then mapped into asymmetric binary attributes so that existing association analysis algorithms can be applied.

Discretization Issues

- Size of the discretized intervals affect support & confidence

{Refund = No, (Income = \$51,250)} → {Cheat = No}

{Refund = No, (60K ≤ Income ≤ 80K)} → {Cheat = No}

{Refund = No, (0K ≤ Income ≤ 1B)} → {Cheat = No}

- If intervals too small
 - may not have enough support
- If intervals too large
 - may not have enough confidence

Statistics-based Methods

- Example:
Browser=Mozilla \wedge Buy=Yes \rightarrow Age: $\mu=23$
- Rule consequent consists of a continuous variable, characterized by their statistics
 - mean, median, standard deviation, etc.
- Approach:
 - Withhold the target variable from the rest of the data
 - Apply existing frequent itemset generation on the rest of the data
 - For each frequent itemset, compute the descriptive statistics for the corresponding target variable
 - Frequent itemset becomes a rule by introducing the target variable as rule consequent
 - Apply statistical test to determine interestingness of the rule

Statistics-based Methods

- How to determine whether an association rule interesting?
 - Compare the statistics for segment of population covered by the rule vs segment of population not covered by the rule:

$$A \Rightarrow B: \mu \quad \text{versus} \quad A \Rightarrow B: \mu'$$
 - Statistical hypothesis testing:
 - Null hypothesis: $H_0: \mu' = \mu + \Delta$
 - Alternative hypothesis: $H_1: \mu' > \mu + \Delta$
 - Z has zero mean and variance 1 under null hypothesis

$$Z = \frac{\mu' - \mu - \Delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Statistics-based Methods

- Example:

r: Browser=Mozilla \wedge Buy=Yes \rightarrow Age: $\mu=23$

- Rule is interesting if difference between μ and μ' is greater than 5 years (i.e., $\Delta = 5$)
- For r, suppose $n_1 = 50, s_1 = 3.5$
- For r' (complement): $n_2 = 250, s_2 = 6.5$

$$Z = \frac{\mu' - \mu - \Delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{30 - 23 - 5}{\sqrt{\frac{3.5^2}{50} + \frac{6.5^2}{250}}} = 3.11$$

- For 1-sided test at 95% confidence level, critical Z-value for rejecting null hypothesis is 1.64.
- Since Z is greater than 1.64, r is an interesting rule

Non-discretization Methods

- There are certain applications in which analysts are more interested in finding associations among the continuous attributes, rather than associations among discrete intervals of the continuous attributes.

Min-Apriori (Han et al)

Document-term matrix:

TID	W1	W2	W3	W4	W5
D1	2	2	0	0	1
D2	0	0	1	2	2
D3	2	3	0	0	0
D4	0	0	1	0	1
D5	1	1	1	0	2

Example:

W1 and W2 tends to appear together in the same document

Min-Apriori

- Data contains only continuous attributes of the same “type”
 - e.g., frequency of words in a document

TID	W1	W2	W3	W4	W5
D1	2	2	0	0	1
D2	0	0	1	2	2
D3	2	3	0	0	0
D4	0	0	1	0	1
D5	1	1	1	0	2

- Potential solution:
 - Convert into 0/1 matrix and then apply existing algorithms
 - lose word frequency information
 - Discretization does not apply as users want association among words not ranges of words

Min-Apriori

- How to determine the support of a word?
 - If we simply sum up its frequency, support count will be greater than total number of documents!
 - Normalize the word vectors – e.g., using L_1 norm
 - Each word has a support equals to 1.0

TID	W1	W2	W3	W4	W5
D1	2	2	0	0	1
D2	0	0	1	2	2
D3	2	3	0	0	0
D4	0	0	1	0	1
D5	1	1	1	0	2

→ Normalize

TID	W1	W2	W3	W4	W5
D1	0.40	0.33	0.00	0.00	0.17
D2	0.00	0.00	0.33	1.00	0.33
D3	0.40	0.50	0.00	0.00	0.00
D4	0.00	0.00	0.33	0.00	0.17
D5	0.20	0.17	0.33	0.00	0.33

Sajjad Haider

Fall 2011

47

Min-Apriori

- New definition of support:

$$\text{sup}(C) = \sum_{i \in T} \min_{j \in C} D(i, j)$$

TID	W1	W2	W3	W4	W5
D1	0.40	0.33	0.00	0.00	0.17
D2	0.00	0.00	0.33	1.00	0.33
D3	0.40	0.50	0.00	0.00	0.00
D4	0.00	0.00	0.33	0.00	0.17
D5	0.20	0.17	0.33	0.00	0.33

Example:

$$\begin{aligned} \text{Sup}(W1, W2, W3) \\ &= 0 + 0 + 0 + 0 + 0.17 \\ &= 0.17 \end{aligned}$$

Sajjad Haider

Fall 2011

48

Anti-monotone property of Support

TID	W1	W2	W3	W4	W5
D1	0.40	0.33	0.00	0.00	0.17
D2	0.00	0.00	0.33	1.00	0.33
D3	0.40	0.50	0.00	0.00	0.00
D4	0.00	0.00	0.33	0.00	0.17
D5	0.20	0.17	0.33	0.00	0.33

Example:

$$\text{Sup}(W1) = 0.4 + 0 + 0.4 + 0 + 0.2 = 1$$

$$\text{Sup}(W1, W2) = 0.33 + 0 + 0.4 + 0 + 0.17 = 0.9$$

$$\text{Sup}(W1, W2, W3) = 0 + 0 + 0 + 0 + 0.17 = 0.17$$