Statistical Learning and Data Mining
CS 363D/ SSC 358

Lecture: Association Rules II

Prof. Pradeep Ravikumar
pradeepr@cs.utexas.edu

Recall: 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

Recall: Apriori Principle for frequent itemsets

Found to be
Infrequent

Pruned
supersets
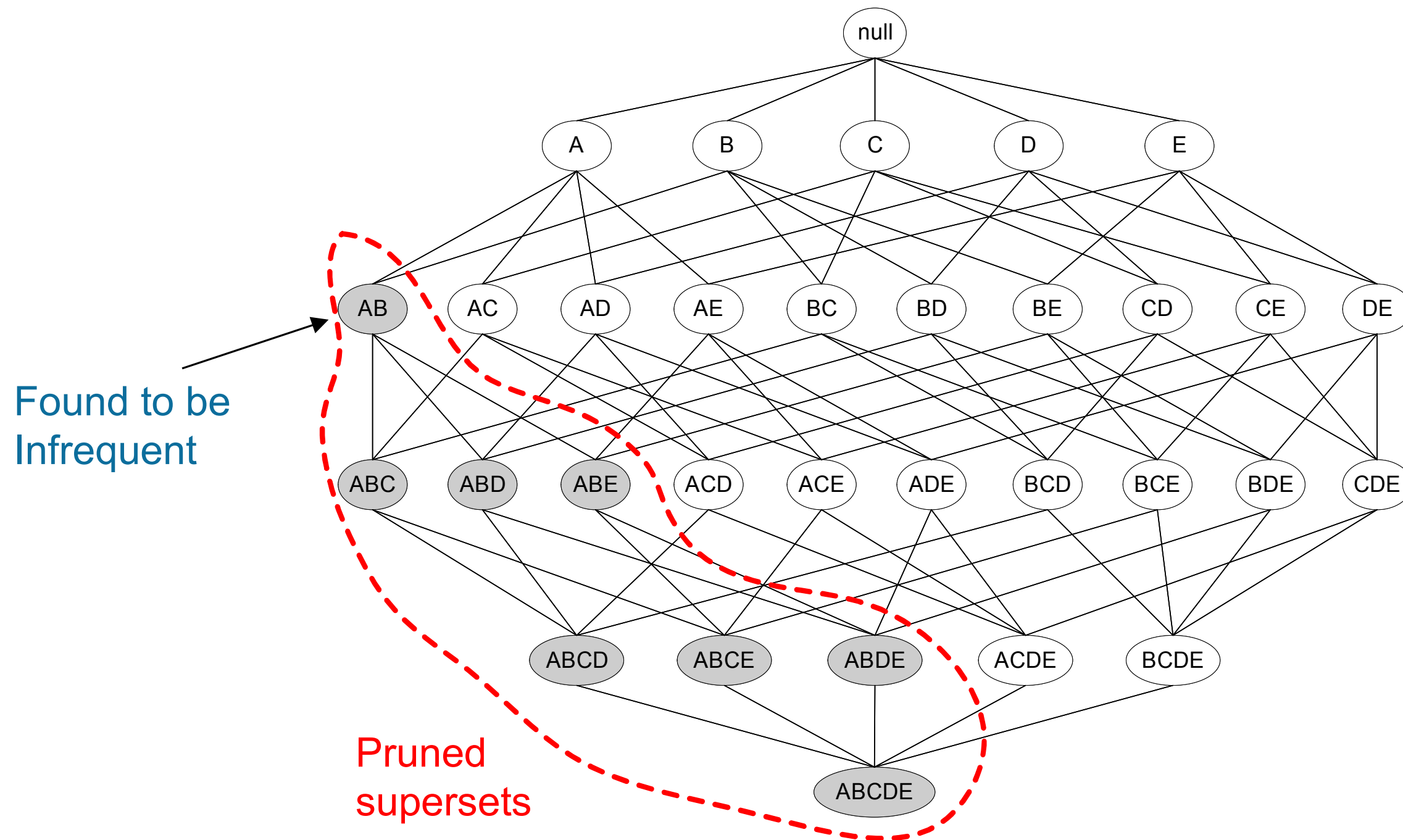
Recall: Apriori Principle for frequent itemsets

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)

Minimum Support = 3

Triplets (3-itemsets)
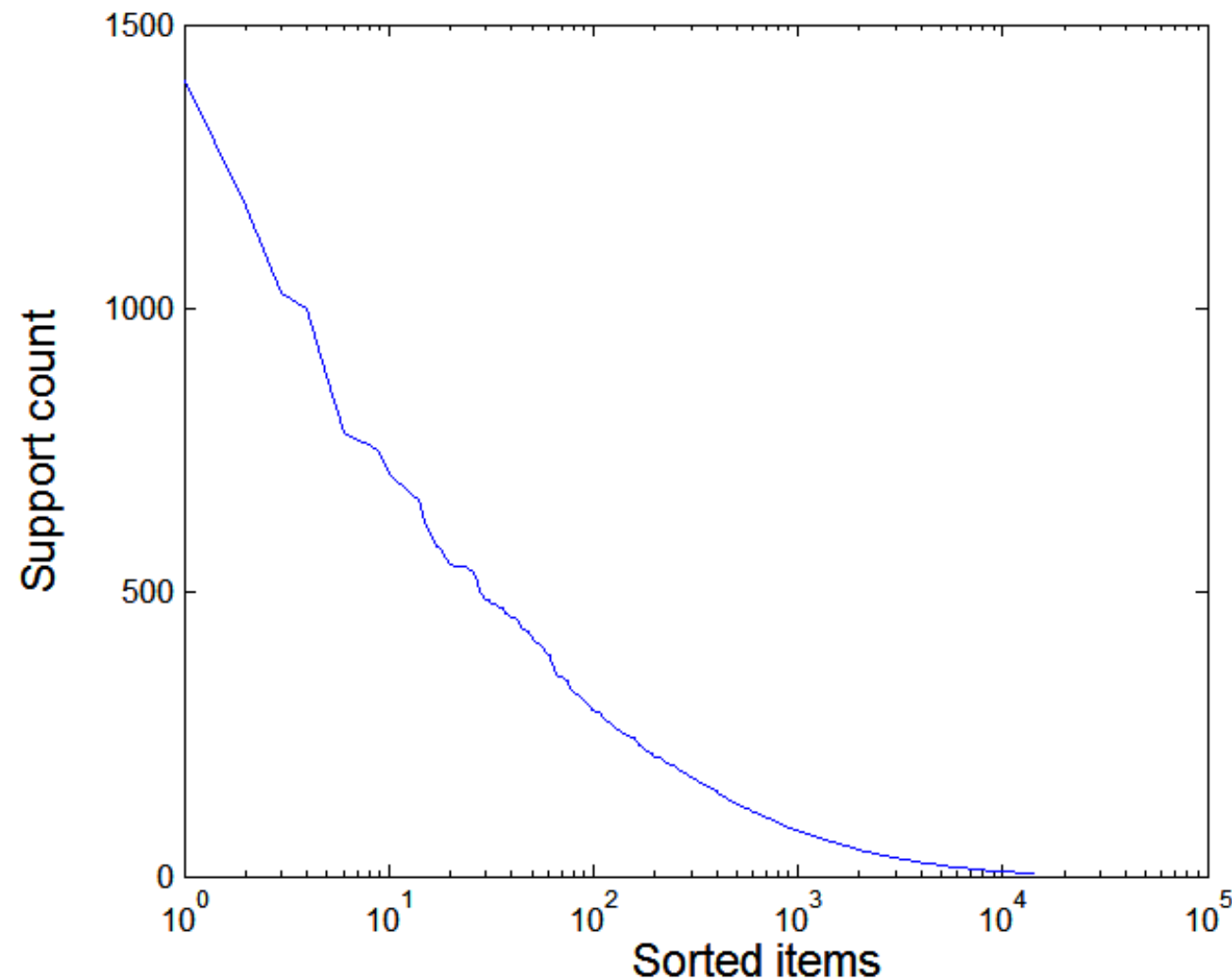
Itemset	Count
{Bread,Milk,Diaper}	3

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

Effect of Support Distribution

- Many real data sets have skewed support distribution

Support distribution of a retail data set

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)

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

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

- Using a single minimum support threshold may not be effective
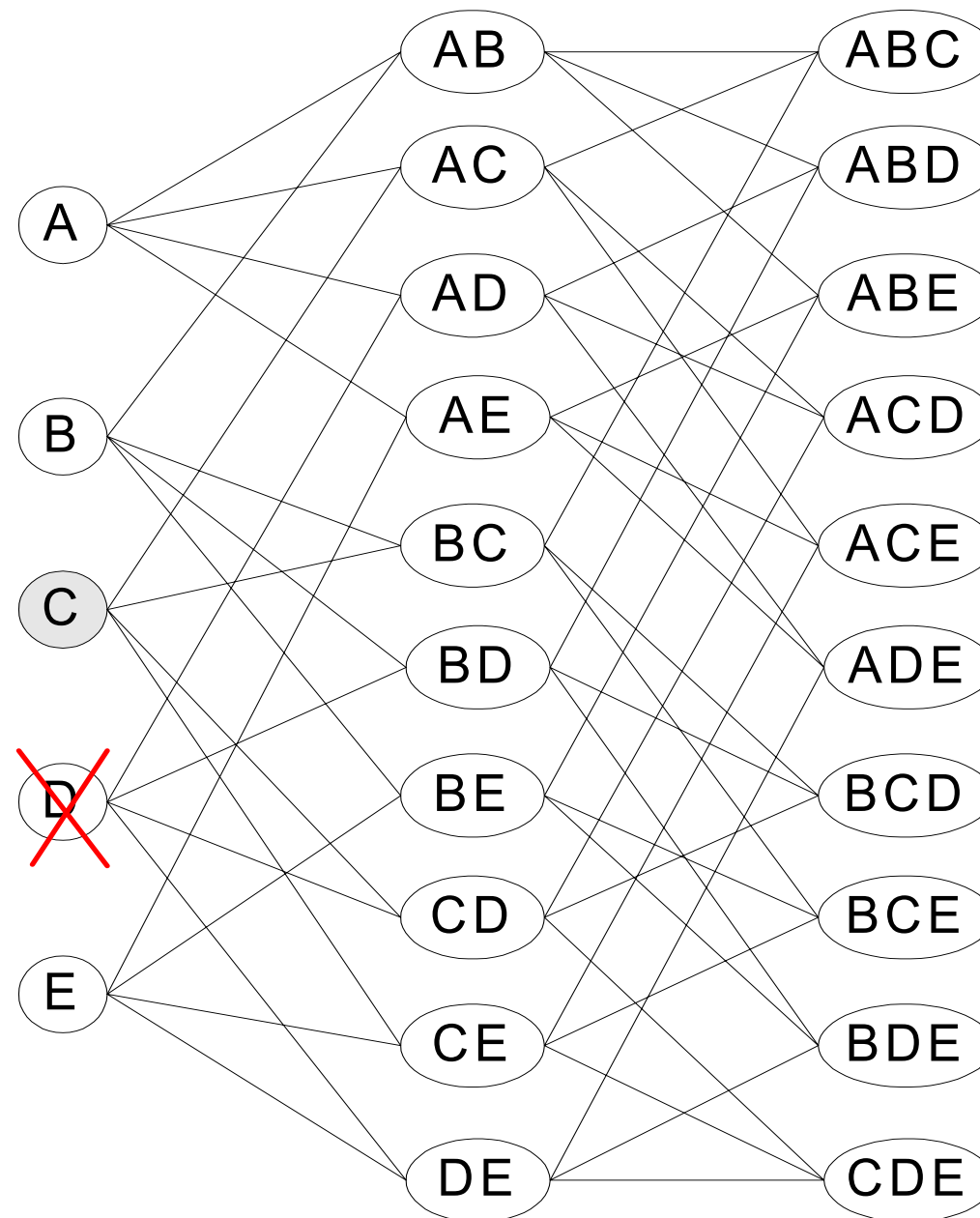
Multiple Minimum Support

- How to apply multiple minimum supports?

Multiple Minimum Support

● How to apply multiple minimum supports?

– MS(i): minimum support for item i

– e.g.:   MS(Milk)=5%,   MS(Coke) = 3%,
            MS(Broccoli)=0.1%,   MS(Salmon)=0.5%

– MS({Milk, Broccoli}) = min (MS(Milk), MS(Broccoli))
                                = 0.1%

Multiple Minimum Support

Item	MS(I)	Sup(I)
A	0.10%	0.25%
B	0.20%	0.26%
C	0.30%	0.30%
D	0.50%	0.05%
E	3%	4.20%

Multiple Minimum Support
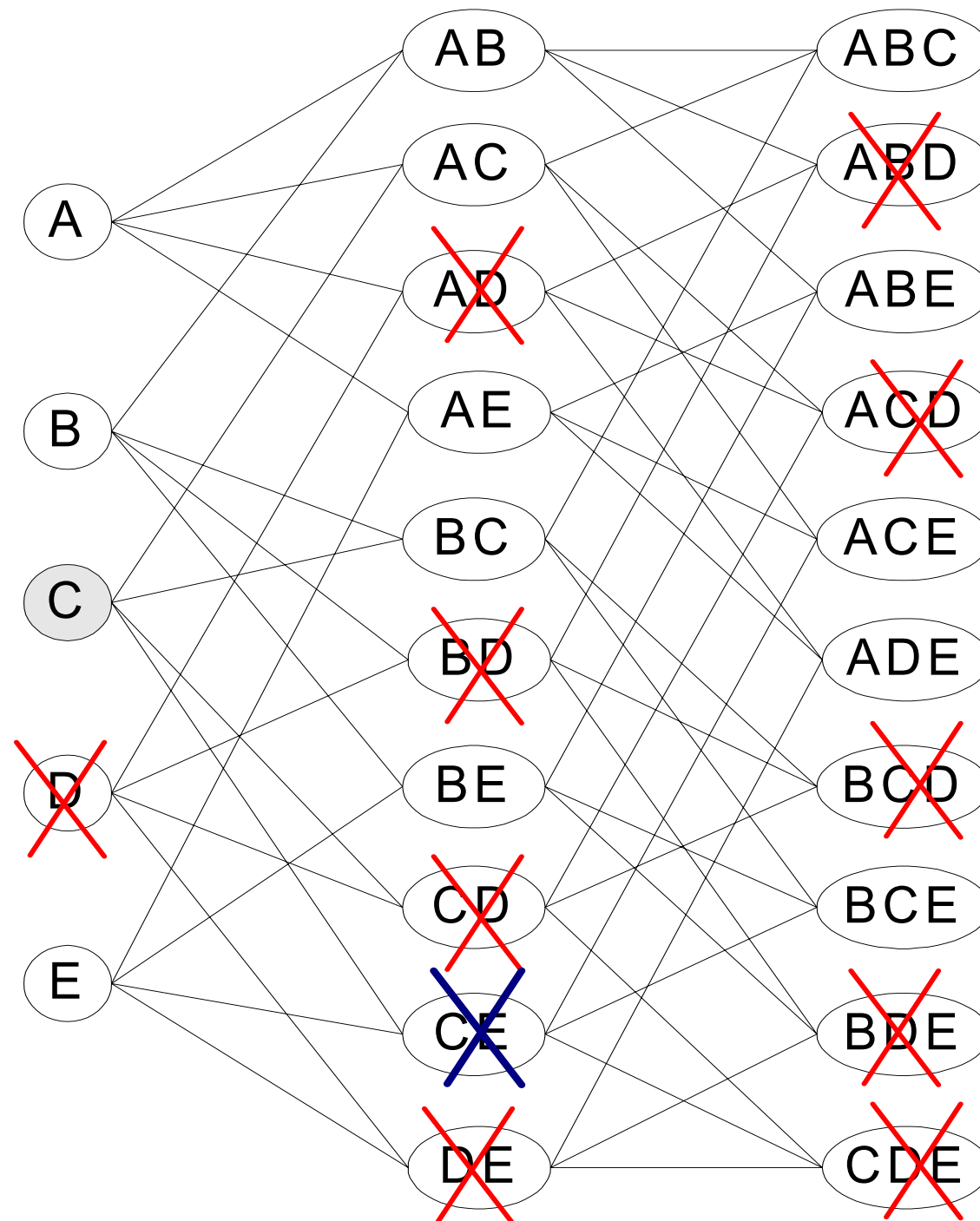
Item	MS(I)	Sup(I)
A	0.10%	0.25%
B	0.20%	0.26%
C	0.30%	0.30%
D	0.50%	0.05%
E	3%	4.20%

Recall: 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

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,		

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 $L \rightarrow \varnothing$ and $\varnothing \rightarrow L$)

Rule Generation

- How to efficiently generate rules from frequent itemsets?

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 $c(AB \rightarrow D)$

Rule Generation

- How to efficiently generate rules from frequent itemsets?
  - In general, confidence does not have an anti-monotone property

    $c(ABC \to D)$ can be larger or smaller than $c(AB \to D)$

  - But confidence of rules generated from the same itemset has an anti-monotone property
  - e.g., L = {A,B,C,D}:

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

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

Rule Generation for Apriori Algorithm

Lattice of rules
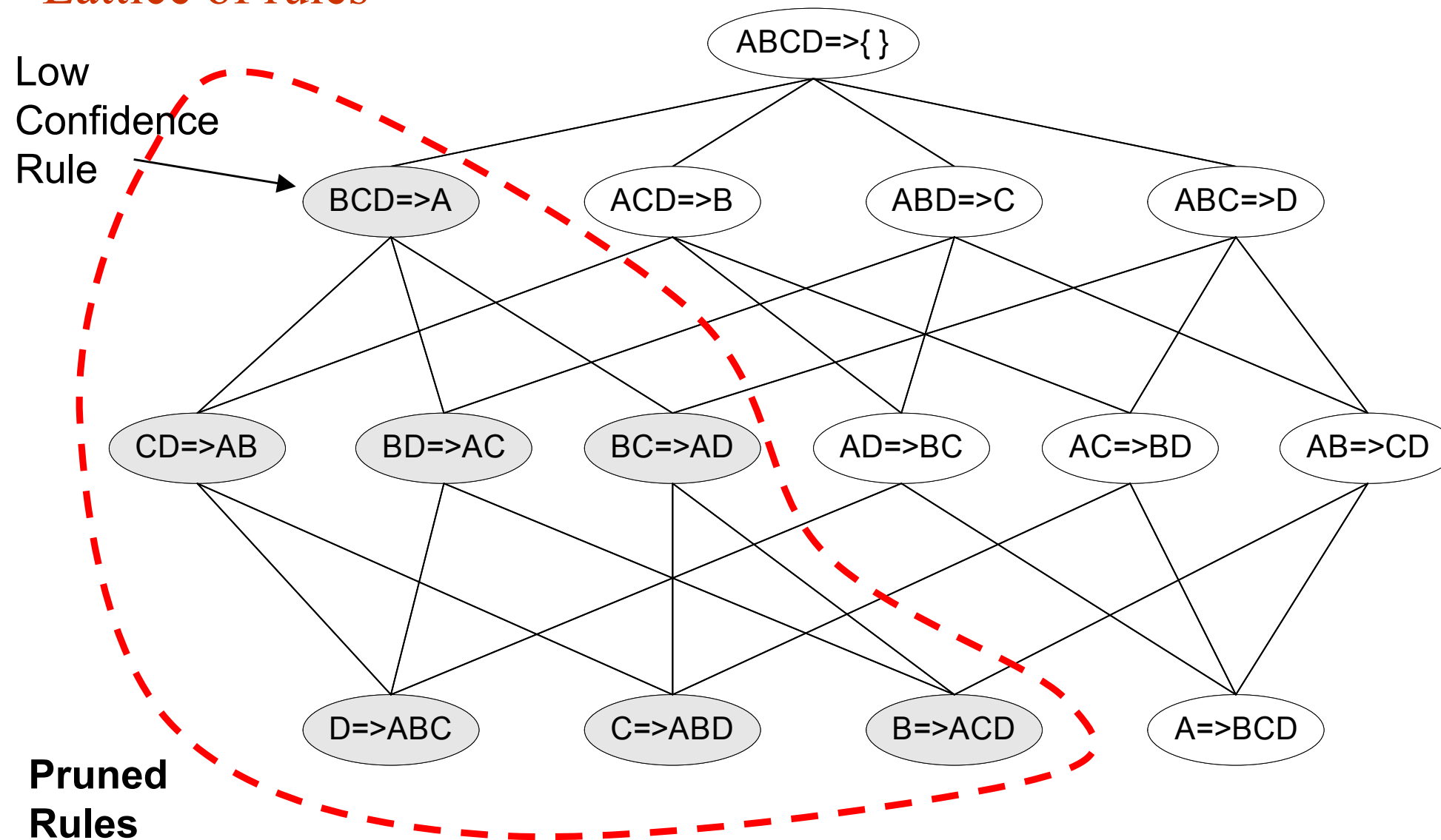
Support-based Pruning

- Most of the association rule mining algorithms use support measure to prune rules and itemsets

Support-based Pruning

- Most of the association rule mining algorithms use support measure to prune rules and itemsets

- Study effect of support pruning on correlation of itemsets
  - Generate 10000 random contingency tables
  - Compute support and pairwise correlation for each table
  - Apply support-based pruning and examine the tables that are removed

Effect of Support Based Pruning

All Itempairs
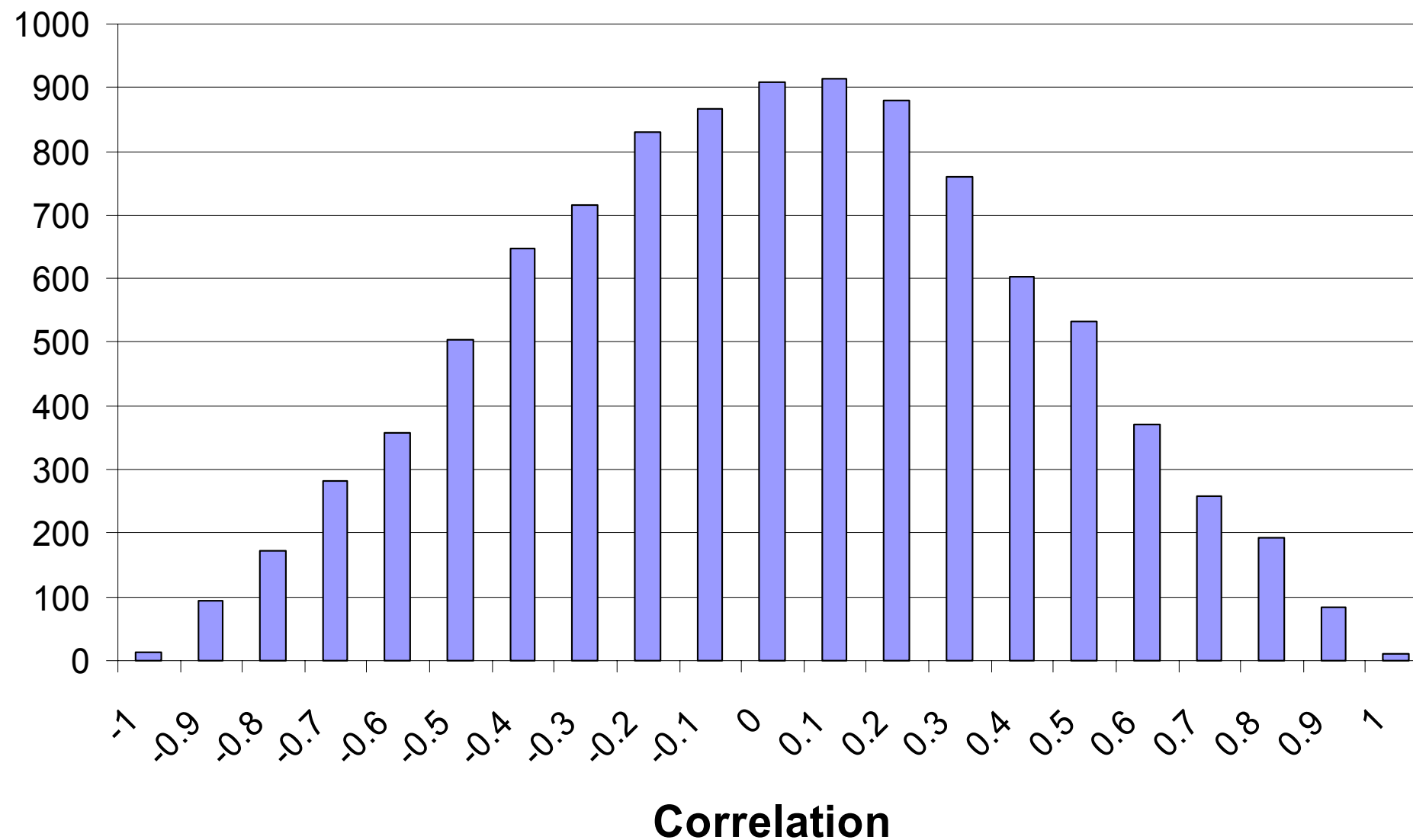
Effect of Support Based Pruning

Support < 0.01

Support < 0.03

Support < 0.05
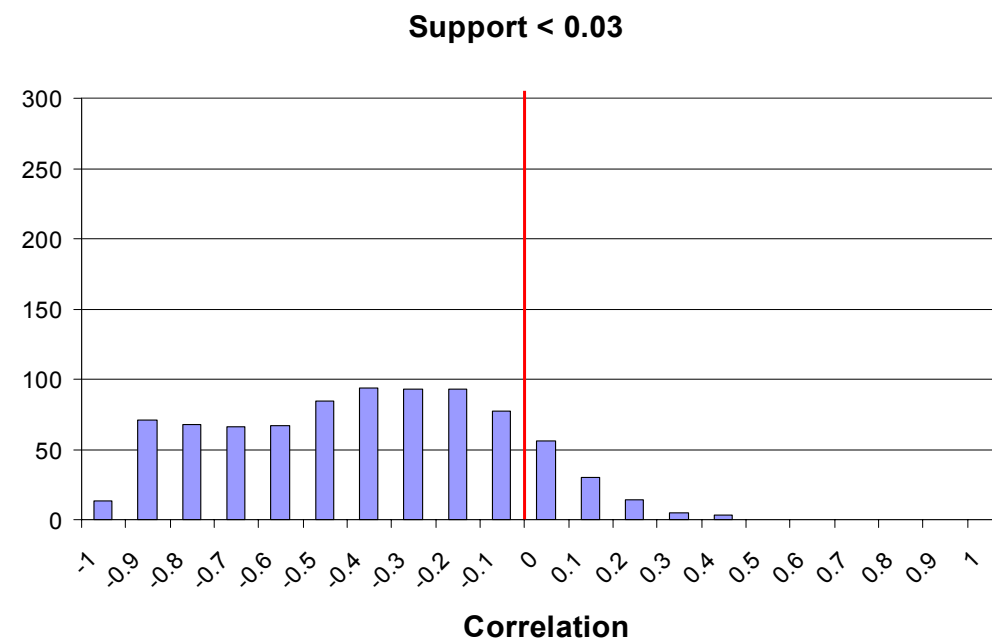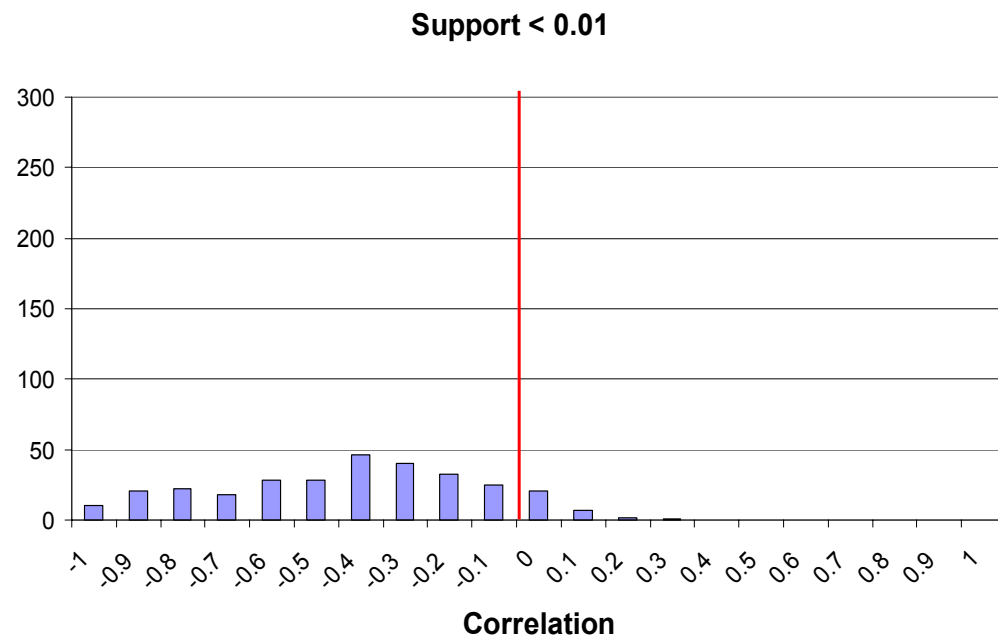
Support-based pruning eliminates mostly negatively correlated itemsets
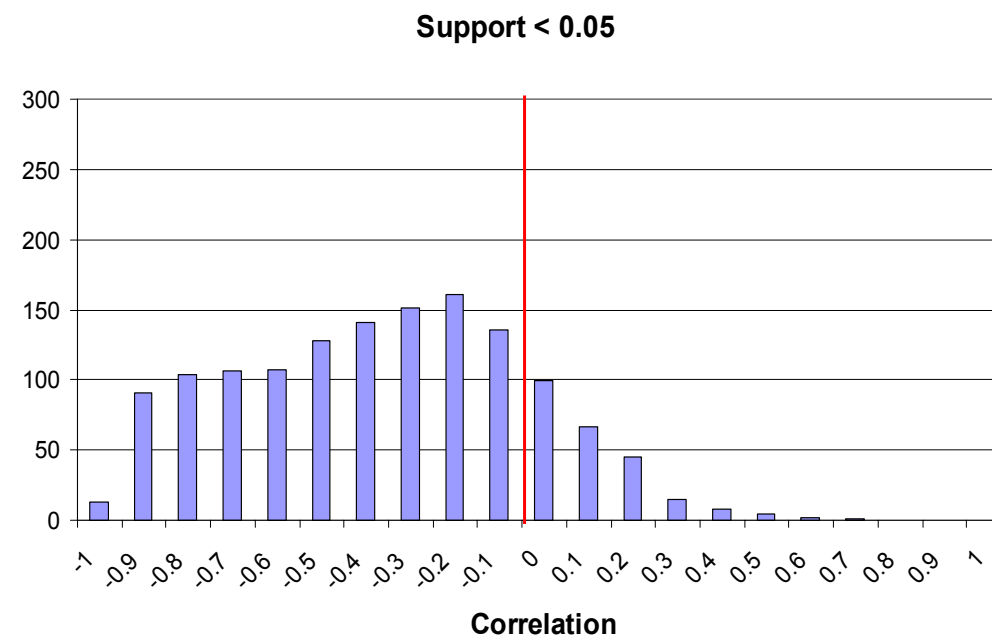
Rule Evaluation

Evaluating Generated Rules

- 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	$\overline{Y}$	
X	f_{11}	f_{10}	f_{1+}
$\overline{X}$	f_{01}	f_{00}	f_{o+}
	f_{+1}	f_{+0}	$|T|$

f_{11}: support of X and Y
f_{10}: support of X and $\overline{Y}$
f_{01}: support of $\overline{X}$ and Y
f_{00}: support of $\overline{X}$ and $\overline{Y}$

Used to define various measures

- support, confidence, lift, Gini, J-measure, etc.

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(Coffee|Tea) =

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(Coffee|Tea) = 0.75

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(Coffee|Tea) = 0.75

but P(Coffee) = 0.9

$\Rightarrow$ Although confidence is high, rule is misleading

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(Coffee|Tea) = 0.75

but P(Coffee) = 0.9

$\Rightarrow$ Although confidence is high, rule is misleading

$\Rightarrow$ P(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)

 – $P(S \wedge B) = 420/1000 = 0.42$

 – $P(S) \times P(B) = 0.6 \times 0.7 = 0.42$

 – $P(S \wedge B) = P(S) \times P(B)$ => Statistical independence

 – $P(S \wedge B) > P(S) \times P(B)$ => Positively correlated

 – $P(S \wedge B) < P(S) \times P(B)$ => Negatively correlated

Statistical Measures

- Measures that take into account statistical dependence

$$Lift = \frac{P(Y \mid X)}{P(Y)} = \frac{\text{conf.}(X \to Y)}{\text{supp.}(Y)}$$

Statistical Measures

- Measures that take into account statistical dependence

$$Lift = \frac{P(Y \mid X)}{P(Y)} = \frac{\text{conf.}(X \rightarrow Y)}{\text{supp.}(Y)}$$

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

Statistical Measures

- Measures that take into account statistical dependence

$$Lift = \frac{P(Y \mid X)}{P(Y)} = \frac{\text{conf.}(X \rightarrow Y)}{\text{supp.}(Y)}$$

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

(Same as "Lift")

Statistical Measures

- Measures that take into account statistical dependence

$$Lift = \frac{P(Y \mid X)}{P(Y)} = \frac{\text{conf.}(X \to Y)}{\text{supp.}(Y)}$$

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

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

Statistical Measures

- Measures that take into account statistical dependence

$$Lift = \frac{P(Y \mid X)}{P(Y)} = \frac{\text{conf.}(X \rightarrow Y)}{\text{supp.}(Y)}$$

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

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

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

Example: Lift/Interest

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

Association Rule: Tea → Coffee

Confidence= P(Coffee|Tea) = 0.75

but P(Coffee) = 0.9

$\Rightarrow$ Lift = 0.75/0.9= 0.8333

Drawback of Lift/Interest

	Y	$\overline{Y}$	
X	10	0	10
$\overline{X}$	0	90	90
	10	90	100

	Y	$\overline{Y}$	
X	90	0	90
$\overline{X}$	0	10	10
	90	10	100

$$Lift = \frac{0.1}{(0.1)(0.1)} = 10$$

$$Lift = \frac{0.9}{(0.9)(0.9)} = 1.11$$

Statistical independence:

If P(X,Y)=P(X)P(Y) => Lift = 1

Comparing Different Measures

10 examples of contingency tables:

Example	f_{11}	f_{10}	f_{01}	f_{00}
E1	8123	83	424	1370
E2	8330	2	622	1046
E3	9481	94	127	298
E4	3954	3080	5	2961
E5	2886	1363	1320	4431
E6	1500	2000	500	6000
E7	4000	2000	1000	3000
E8	4000	2000	2000	2000
E9	1720	7121	5	1154
E10	61	2483	4	7452

Rankings of contingency tables using various measures:

#	ϕ	λ	α	Q	Y	κ	M	J	G	s	c	L	V	I	IS	PS	F	AV	S	ζ	K
E1	1	1	3	3	3	1	2	2	1	3	5	5	4	6	2	2	4	6	1	2	5
E2	2	2	1	1	1	2	1	3	2	2	1	1	1	8	3	5	1	8	2	3	6
E3	3	3	4	4	4	3	3	8	7	1	4	4	6	10	1	8	6	10	3	1	10
E4	4	7	2	2	2	5	4	1	3	6	2	2	2	4	4	1	2	3	4	5	1
E5	5	4	8	8	8	4	7	5	4	7	9	9	9	3	6	3	9	4	5	6	3
E6	6	6	7	7	7	7	6	4	6	9	8	8	7	2	8	6	7	2	7	8	2
E7	7	5	9	9	9	6	8	6	5	4	7	7	8	5	5	4	8	5	6	4	4
E8	8	9	10	10	10	8	10	10	8	4	10	10	10	9	7	7	10	9	8	7	9
E9	9	9	5	5	5	9	9	7	9	8	3	3	3	7	9	9	3	7	9	9	8
E10	10	8	6	6	6	10	5	9	10	10	6	6	5	1	10	10	5	1	10	10	7

Mo measures Mo problems

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?

#	Measure	Formula
1	ϕ-coefficient	$\dfrac{P(A,B)-P(A)P(B)}{\sqrt{P(A)P(B)(1-P(A))(1-P(B))}}$
2	Goodman-Kruskal's (λ)	$\dfrac{\sum_j \max_k P(A_j,B_k)+\sum_k \max_j P(A_j,B_k)-\max_j P(A_j)-\max_k P(B_k)}{2-\max_j P(A_j)-\max_k P(B_k)}$
3	Odds ratio (α)	$\dfrac{P(A,B)P(\overline{A},\overline{B})}{P(A,\overline{B})P(\overline{A},B)}$
4	Yule's Q	$\dfrac{P(A,B)P(\overline{A}B)-P(A,\overline{B})P(\overline{A},B)}{P(A,B)P(\overline{A}B)+P(A,\overline{B})P(\overline{A},B)}=\dfrac{\alpha-1}{\alpha+1}$
5	Yule's Y	$\dfrac{\sqrt{P(A,B)P(\overline{A}B)}-\sqrt{P(A,\overline{B})P(\overline{A},B)}}{\sqrt{P(A,B)P(\overline{A}B)}+\sqrt{P(A,\overline{B})P(\overline{A},B)}}=\dfrac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}$
6	Kappa (κ)	$\dfrac{P(A,B)+P(\overline{A},\overline{B})-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A)P(B)-P(\overline{A})P(\overline{B})}$
7	Mutual Information (M)	$\dfrac{\sum_i \sum_j P(A_i,B_j)\log\frac{P(A_i,B_j)}{P(A_i)P(B_j)}}{\min(-\sum_i P(A_i)\log P(A_i),-\sum_j P(B_j)\log P(B_j))}$
8	J-Measure (J)	$\max\left(P(A,B)\log(\frac{P(B|A)}{P(B)})+P(A\overline{B})\log(\frac{P(\overline{B}|A)}{P(\overline{B})}),\right.$ $\left. P(A,B)\log(\frac{P(A|B)}{P(A)})+P(\overline{A}B)\log(\frac{P(\overline{A}|B)}{P(A)})\right)$
9	Gini index (G)	$\max\left(P(A)[P(B|A)^2+P(\overline{B}|A)^2]+P(\overline{A})[P(B|\overline{A})^2+P(\overline{B}|\overline{A})^2]\right.$ $-P(B)^2-P(\overline{B})^2,$ $P(B)[P(A|B)^2+P(\overline{A}|B)^2]+P(\overline{B})[P(A|\overline{B})^2+P(\overline{A}|\overline{B})^2]$ $\left.-P(A)^2-P(\overline{A})^2\right)$
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)	$\max\left(\frac{P(A)P(\overline{B})}{P(A\overline{B})},\frac{P(B)P(\overline{A})}{P(B\overline{A})}\right)$
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)	$\frac{P(A,B)+P(\overline{A}B)}{P(A)P(B)+P(\overline{A})P(\overline{B})}\times\frac{1-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A,B)-P(\overline{A}B)}$
20	Jaccard (ζ)	$\frac{P(A,B)}{P(A)+P(B)-P(A,B)}$
21	Klosgen (K)	$\sqrt{P(A,B)}\max(P(B|A)-P(B),P(A|B)-P(A))$

Properties of a good measure

- Piatetsky-Shapiro:
  3 properties a good measure M must satisfy:
  - M(A,B) = 0 if A and B are statistically independent

Properties of a good measure

- Piatetsky-Shapiro:
  3 properties a good measure M must satisfy:
  - M(A,B) = 0 if A and B are statistically independent

  - M(A,B) increase monotonically with P(A,B) when P(A) and P(B) remain unchanged

Properties of a good measure

- Piatetsky-Shapiro:
  3 properties a good measure M must satisfy:
  - M(A,B) = 0 if A and B are statistically independent

  - M(A,B) increase monotonically with P(A,B) when P(A) and P(B) remain unchanged

  - M(A,B) decreases monotonically with P(A) [or P(B)] when P(A,B) and P(B) [or P(A)] remain unchanged

Subjective Interestingness Measure

- Objective measure:
  - Rank patterns based on statistics computed from data
  - e.g., 21 measures of association (support, confidence, Laplace, Gini, mutual information, Jaccard, etc).

- Subjective measure:
  - Rank patterns according to user's interpretation
    - A pattern is subjectively interesting if it contradicts the expectation of a user (Silberschatz & Tuzhilin)
    - A pattern is subjectively interesting if it is actionable (Silberschatz & Tuzhilin)