NC STATE UNIVERSITY

CSC-422/522: ALDA
M/W. 8.30-9.45am. HL-Auditorium.

Ranga Raju Vatsavai
Chancellors Faculty Excellence Associate Professor in Geospatial Analytics
Department of Computer Science, North Carolina State University (NCSU)
Associate Director, Center for Geospatial Analytics, NCSU
&
Joint Faculty, Oak Ridge National Laboratory (ORNL)

NC STATE UNIVERSITY

Today

- Association Rules

2

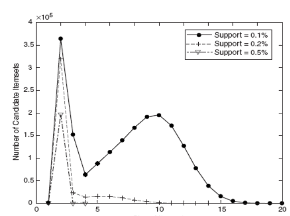
NC STATE UNIVERSITY

Factors Affecting Complexity of Apriori

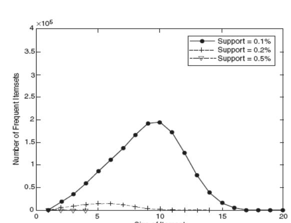
- Choice of minimum support threshold
  - lowering support threshold results in more frequent itemsets
  - this may increase number of candidates and max length of frequent itemsets
- Dimensionality (number of items) of the data set
  - more space is needed to store support count of each item
  - if 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
- Average transaction width
  - transaction width increases with denser data sets
  - This may increase max length of frequent itemsets and traversals of hash tree (number of subsets in a transaction increases with its width)

NC STATE UNIVERSITY

Factors Affecting Complexity of Apriori

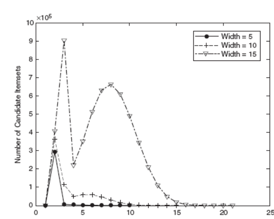


(a) Number of candidate itemsets.

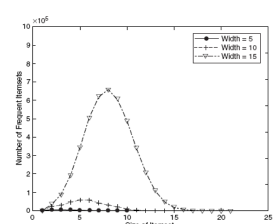


(b) Number of frequent itemsets.

Figure 6.13. Effect of support threshold on the number of candidate and frequent itemsets.



(a) Number of candidate itemsets.



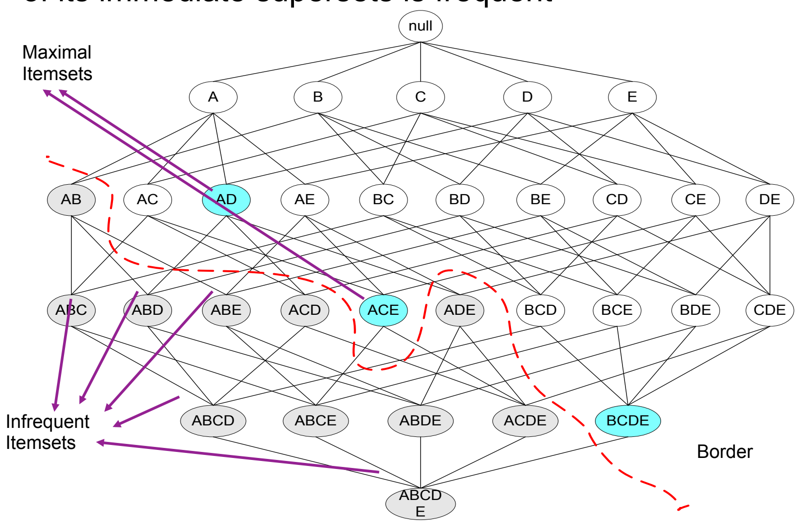
(b) Number of Frequent Itemsets.

Figure 6.14. Effect of average transaction width on the number of candidate and frequent itemsets.

NC STATE UNIVERSITY

Maximal Frequent Itemset

An itemset is maximal frequent if it is frequent and none of its immediate supersets is frequent



NC STATE UNIVERSITY

An illustrative example

Items

Transactions

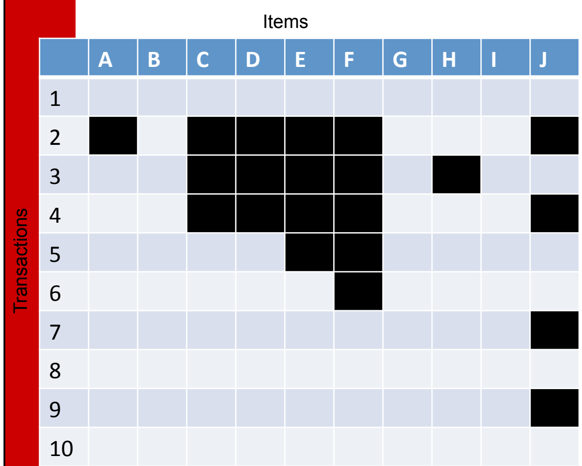
	A	B	C	D	E	F	G	H	I	J
1										
2	■		■	■	■	■				■
3			■	■	■	■		■		
4			■	■	■	■				■
5					■	■				
6						■				
7										■
8										
9										■
10										

Support threshold (by count) : 5

Frequent itemsets: ?

NC STATE UNIVERSITY

An illustrative example

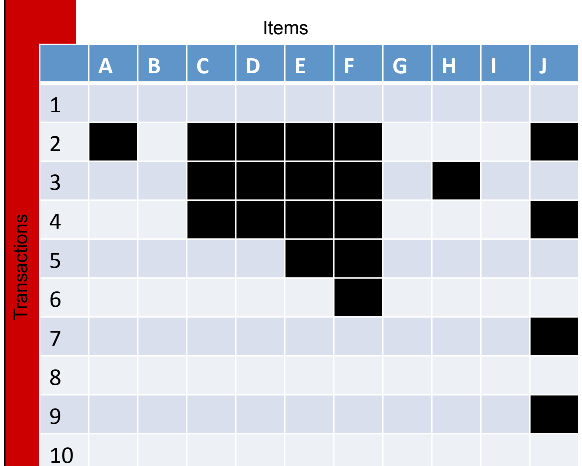


Support threshold (by count) : 5

Frequent itemsets: {F}

NC STATE UNIVERSITY

An illustrative example



Support threshold (by count) : 5

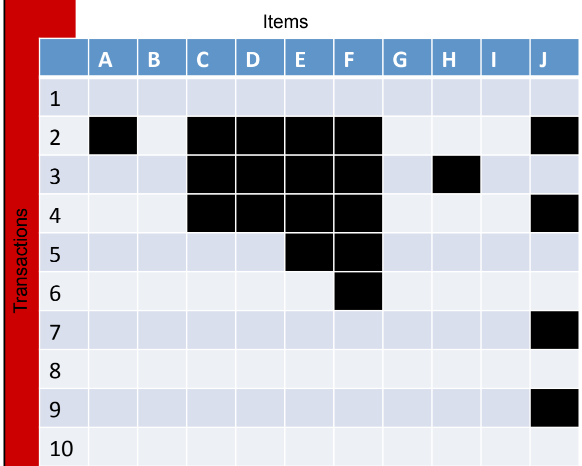
Frequent itemsets: {F}

Support threshold (by count): 4

Frequent itemsets: ?

NC STATE UNIVERSITY

An illustrative example

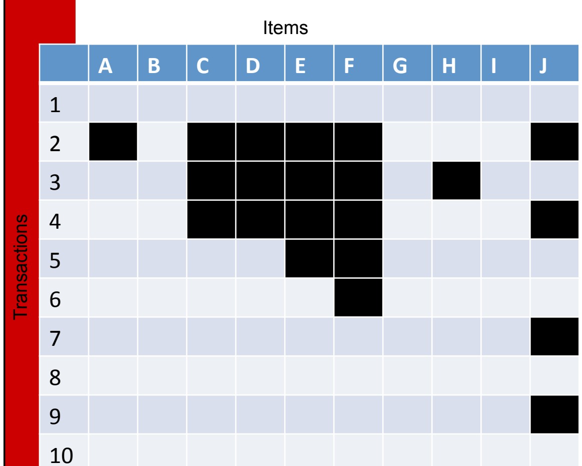


Support threshold (by count) : 5
Frequent itemsets: {F}

Support threshold (by count): 4
Frequent itemsets: {E}, {F}, {E,F}, {J}

NC STATE UNIVERSITY

An illustrative example



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: ?

NC STATE UNIVERSITY

An illustrative example

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}

NC STATE UNIVERSITY

An illustrative example

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: ?

NC STATE UNIVERSITY

An illustrative example

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: ?

NC STATE UNIVERSITY

An illustrative example

Items

Transactions

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

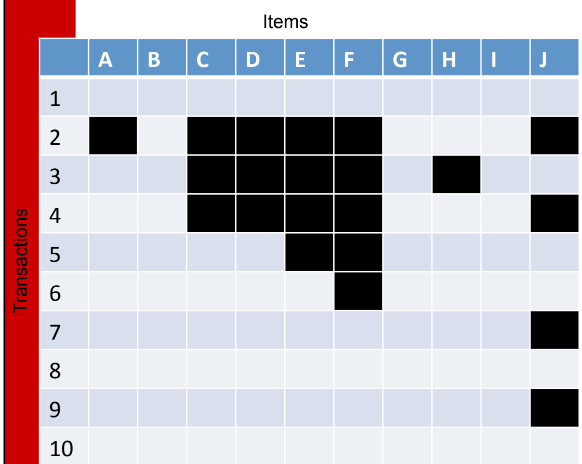
Support threshold (by count) : 5
Frequent itemsets: {F}
Maximal itemsets: {F}

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: ?

NC STATE UNIVERSITY

An illustrative example



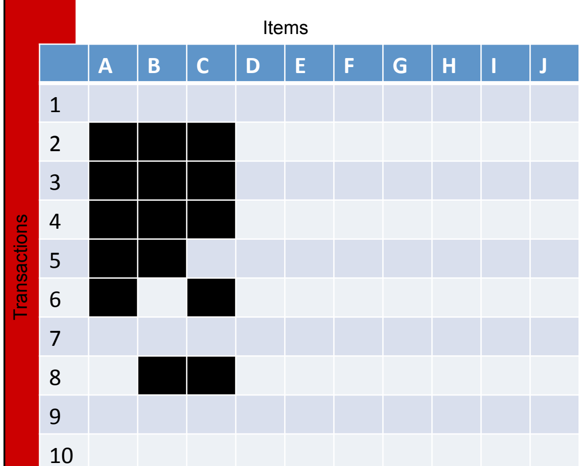
Support threshold (by count) : 5
Frequent itemsets: {F}
Maximal itemsets: {F}

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}

NC STATE UNIVERSITY

Another illustrative example



Support threshold (by count) : 5
Maximal itemsets: {A}, {B}, {C}

Support threshold (by count): 4
Maximal itemsets: {A,B}, {A,C},{B,C}

Support threshold (by count): 3
Maximal itemsets: {A,B,C}

Practice at home

NC STATE UNIVERSITY

Closed Itemset

- An itemset X is closed if none of its immediate supersets has the same support count as the itemset X.
- X is not closed if at least one of its immediate supersets has support count as X.

TID	Items
1	{A,B}
2	{B,C,D}
3	{A,B,C,D}
4	{A,B,D}
5	{A,B,C,D}

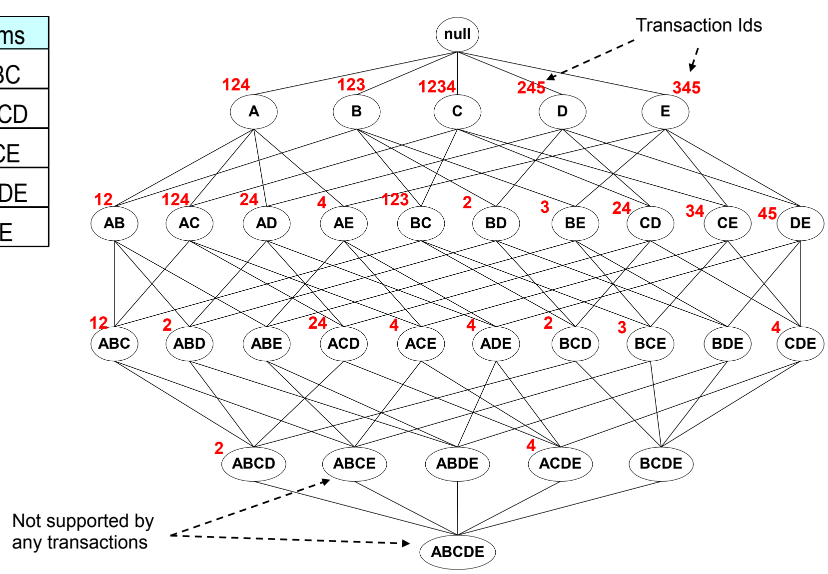
Itemset	Support
{A}	4
{B}	5
{C}	3
{D}	4
{A,B}	4
{A,C}	2
{A,D}	3
{B,C}	3
{B,D}	4
{C,D}	3

Itemset	Support
{A,B,C}	2
{A,B,D}	3
{A,C,D}	2
{B,C,D}	2
{A,B,C,D}	2

NC STATE UNIVERSITY

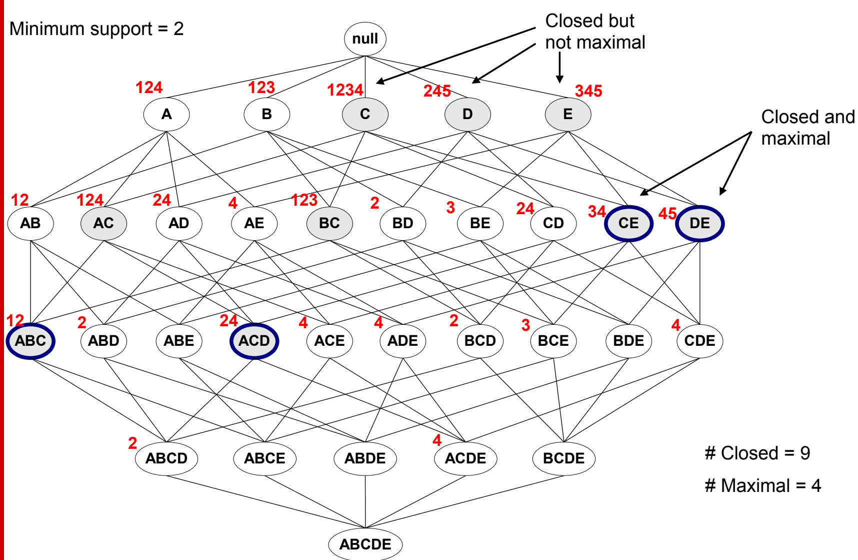
Maximal vs Closed Itemsets

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



NC STATE UNIVERSITY

Maximal vs Closed Itemsets

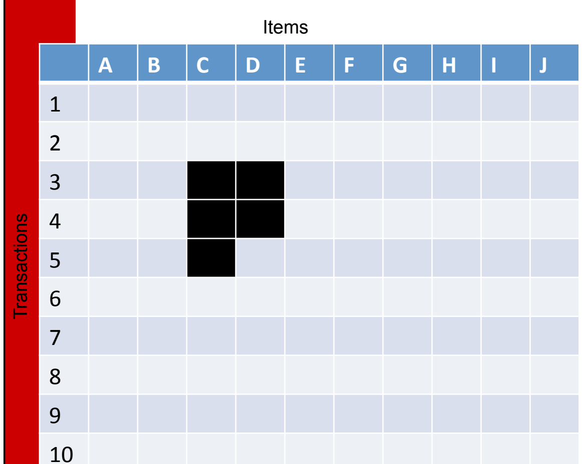


Closed = 9

Maximal = 4

NC STATE UNIVERSITY

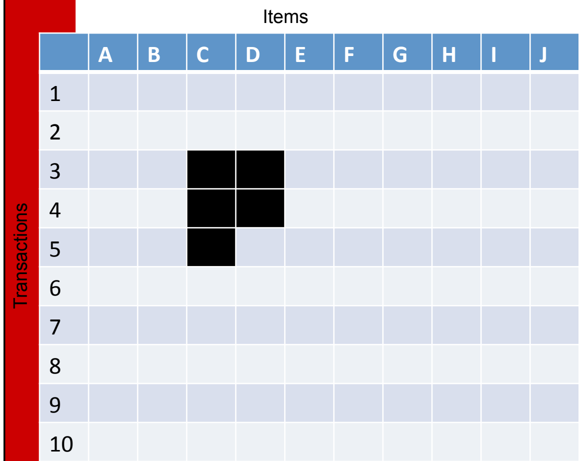
Example 1



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

NC STATE UNIVERSITY

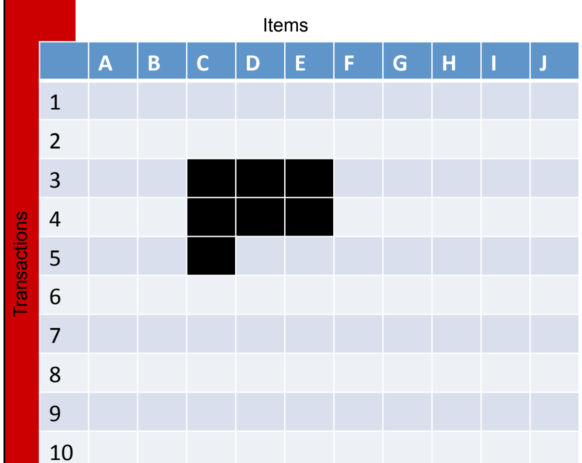
Example 1



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

NC STATE UNIVERSITY

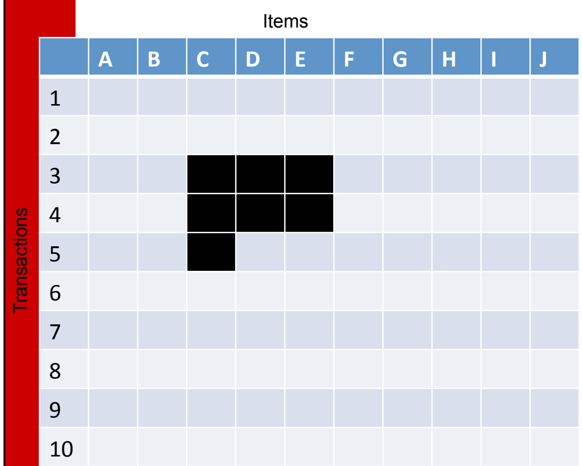
Example 2



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

NC STATE UNIVERSITY

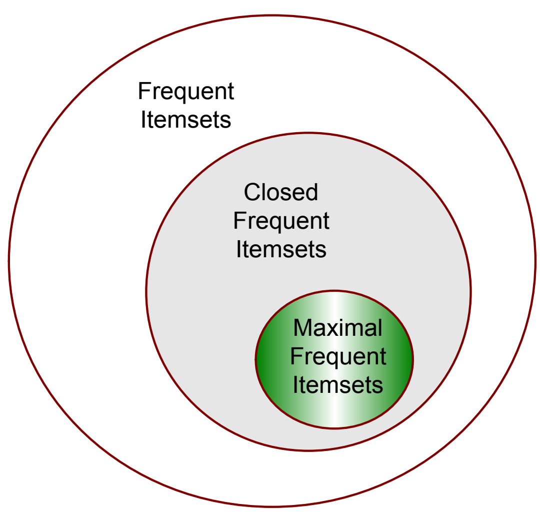
Example 2



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

NC STATE UNIVERSITY

Maximal vs Closed Itemsets

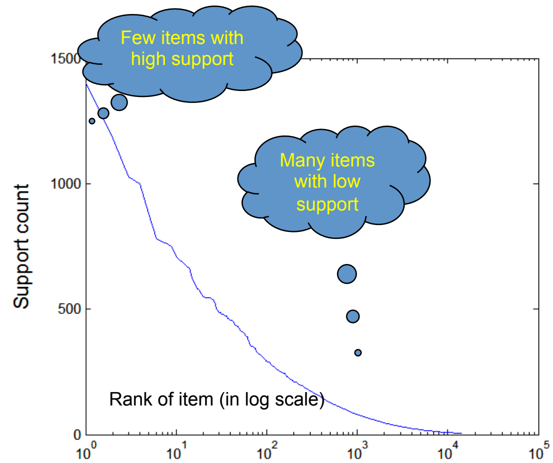


Read Algorithm 6.4. for support counting using closed frequent itemsets

Effect of Support Distribution on Association Mining

- Many real data sets have skewed support distribution

Support distribution of a retail data set

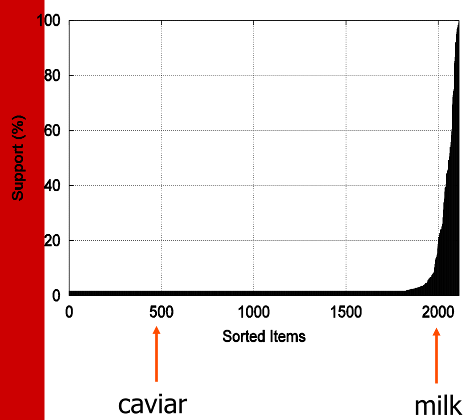


Effect of Support Distribution

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

NC STATE UNIVERSITY

Cross-Support Patterns



A cross-support pattern involves items with varying degree of support

- Example: {caviar, milk}

How to avoid such patterns?

NC STATE UNIVERSITY

A Measure of Cross Support

- Given an itemset, $X = \{i_1, i_2, \ldots, i_k\}$, we can define a measure of cross support, *r*, as

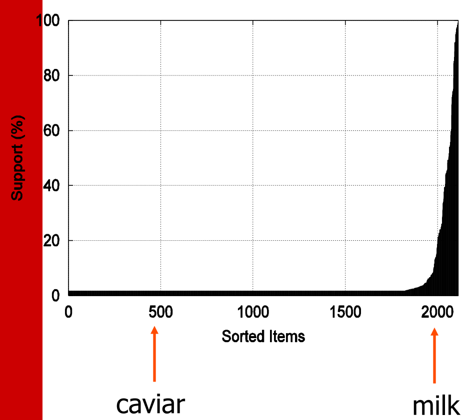
$$r(X) = \frac{\min\left(s(i_1), s(i_2), \ldots, s(i_k)\right)}{\max\left(s(i_1), s(i_2), \ldots, s(i_k)\right)}$$

where (s_i) is the support of item i

- Can be use d to prune cross support patterns, but not to avoid them

NC STATE UNIVERSITY

Confidence and Cross-Support Patterns



Observation:

conf(caviar→milk) is very high

but

conf(milk→caviar) is very low

Therefore,

min(conf(caviar→milk), conf(milk→caviar))

is also very low

NC STATE UNIVERSITY

H-Confidence

- To avoid patterns whose items have very different support, define a new evaluation measure for itemsets
  - Known as h-confidence or all-confidence
- Specifically, given an itemset X={x_1, x_2, …,x_k}
  - h-confidence is the minimum confidence of any association rule formed from itemset

$$h-confidence(X) = \frac{s\left(\left\{x_1, x_2, \ldots, x_k\right\}\right)}{\max\left[s(x_1), s(x_2) \ldots, s(x_k)\right]}$$

Pattern Evaluation

- Association rule algorithms can produce large number of rules

- Interestingness measures can be used to prune/rank the patterns
  - In the original formulation, support & confidence are the only measures used

Computing Interestingness Measure

- Given X → Y or {X,Y}, information needed to compute interestingness can be obtained from a contingency table

Contingency table

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

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, Gini, entropy, etc.

Drawback of Confidence

Customers	Tea	Coffee	…
C1	0	1	…
C2	1	0	…
C3	1	1	…
C4	1	0	…
…			

	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) = 15/20 = 0.75

Confidence > 50%, meaning people who drink tea are more likely to drink coffee than not drink coffee

So rule seems reasonable

Drawback of Confidence

	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) = 15/20 = 0.75

but P(Coffee) = 0.9, which means knowing that a person drinks tea reduces the probability that the person drinks coffee!

⇒ Note that P(Coffee|$\overline{\text{Tea}}$) = 75/80 = 0.9375

Objective Measures

Table 6.11. Examples of symmetric objective measures for the itemset $\{A, B\}$.

Measure (Symbol)	Definition
Correlation (ϕ)	$\frac{Nf_{11} - f_{1+}f_{+1}}{\sqrt{f_{1+}f_{+1}f_{0+}f_{+0}}}$
Odds ratio (α)	$(f_{11}f_{00})/(f_{10}f_{01})$
Kappa (κ)	$\frac{Nf_{11}+Nf_{00}-f_{1+}f_{+1}-f_{0+}f_{+0}}{N^2-f_{1+}f_{+1}-f_{0+}f_{+0}}$
Interest (I)	$(Nf_{11})/(f_{1+}f_{+1})$
Cosine (IS)	$(f_{11})/(\sqrt{f_{1+}f_{+1}})$
Piatetsky-Shapiro (PS)	$\frac{f_{11}}{N} - \frac{f_{1+}f_{+1}}{N^2}$
Collective strength (S)	$\frac{f_{11}+f_{00}}{f_{1+}f_{+1}+f_{0+}f_{+0}} \times \frac{N-f_{1+}f_{+1}-f_{0+}f_{+0}}{N-f_{11}-f_{00}}$
Jaccard (ζ)	$f_{11}/(f_{1+} + f_{+1} - f_{11})$
All-confidence (h)	$\min\left[\frac{f_{11}}{f_{1+}}, \frac{f_{11}}{f_{+1}}\right]$

Objective Measures

Table 6.12. Examples of asymmetric objective measures for the rule $A \longrightarrow B$.

Measure (Symbol)	Definition
Goodman-Kruskal (λ)	$(\sum_j \max_k f_{jk} - max_k f_{+k})/(N - \max_k f_{+k})$
Mutual Information (M)	$(\sum_i \sum_j \frac{f_{ij}}{N} \log \frac{Nf_{ij}}{f_{i+}f_{+j}})/(-\sum_i \frac{f_{i+}}{N} \log \frac{f_{i+}}{N})$
J-Measure (J)	$\frac{f_{11}}{N} \log \frac{Nf_{11}}{f_{1+}f_{+1}} + \frac{f_{10}}{N} \log \frac{Nf_{10}}{f_{1+}f_{+0}}$
Gini index (G)	$\frac{f_{1+}}{N} \times [(\frac{f_{11}}{f_{1+}})^2 + (\frac{f_{10}}{f_{1+}})^2] - (\frac{f_{+1}}{N})^2$ $+ \frac{f_{0+}}{N} \times [(\frac{f_{01}}{f_{0+}})^2 + (\frac{f_{00}}{f_{0+}})^2] - (\frac{f_{+0}}{N})^2$
Laplace (L)	$(f_{11} + 1)/(f_{1+} + 2)$
Conviction (V)	$(f_{1+}f_{+0})/(Nf_{10})$
Certainty factor (F)	$(\frac{f_{11}}{f_{1+}} - \frac{f_{+1}}{N})/(1 - \frac{f_{+1}}{N})$
Added Value (AV)	$\frac{f_{11}}{f_{1+}} - \frac{f_{+1}}{N}$

NC STATE UNIVERSITY

Data Mining

Chapter 7- Association Analysis: Advance Concepts

NC STATE UNIVERSITY

Continuous and Categorical Attributes

How to apply association analysis to non-asymmetric binary variables?

Gender	...	Age	Annual Income	No of hours spent online per week	No of email accounts	Privacy Concern
Female	...	26	90K	20	4	Yes
Male	...	51	135K	10	2	No
Male	...	29	80K	10	3	Yes
Female	...	45	120K	15	3	Yes
Female	...	31	95K	20	5	Yes
Male	...	25	55K	25	5	Yes
Male	...	37	100K	10	1	No
Male	...	41	65K	8	2	No
Female	...	26	85K	12	1	No
...	...	...	...	...	...	...

Example of Association Rule:

{Gender=Male, Age $\in$ [21,30)} $\rightarrow$ {No of hours online $\geq$ 10}

NC STATE UNIVERSITY

Handling Categorical Attributes

- Example: Internet Usage Data

Gender	Level of Education	State	Computer at Home	Online Auction	Chat Online	Online Banking	Privacy Concerns
Female	Graduate	Illinois	Yes	Yes	Daily	Yes	Yes
Male	College	California	No	No	Never	No	No
Male	Graduate	Michigan	Yes	Yes	Monthly	Yes	Yes
Female	College	Virginia	No	Yes	Never	Yes	Yes
Female	Graduate	California	Yes	No	Never	No	Yes
Male	College	Minnesota	Yes	Yes	Weekly	Yes	Yes
Male	College	Alaska	Yes	Yes	Daily	Yes	No
Male	High School	Oregon	Yes	No	Never	No	No
Female	Graduate	Texas	No	No	Monthly	No	No
...	...	...	...	...	...	...	...

{Level of Education=Graduate, Online Banking=Yes}
$\rightarrow$ {Privacy Concerns = Yes}

NC STATE UNIVERSITY

Handling Categorical Attributes

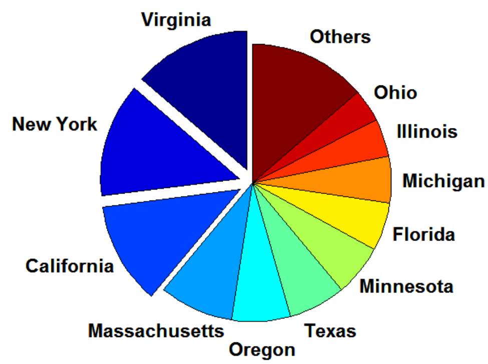
- Introduce a new "item" for each distinct attribute-value pair

Male	Female	Education = Graduate	Education = College	Education = High School	...	Privacy = Yes	Privacy = No
0	1	1	0	0	...	1	0
1	0	0	1	0	...	0	1
1	0	1	0	0	...	1	0
0	1	0	1	0	...	1	0
0	1	1	0	0	...	1	0
1	0	0	1	0	...	1	0
1	0	0	0	0	...	0	1
1	0	0	0	1	...	0	1
0	1	1	0	0	...	0	1
...	...	...	...	...	...	...	...

NC STATE UNIVERSITY

Handling Categorical Attributes

- Some attributes can have many possible values
  - Many of their attribute values have very low support
    - Potential solution: values



NC STATE UNIVERSITY

Handling Categorical Attributes

- Distribution of attribute values can be highly skewed
  - Example: 85% of survey participants own a computer at home
    - Most records have Computer at home = Yes
    - Computation becomes expensive; many frequent itemsets involving the binary item (Computer at home = Yes)
    - Potential solution:
      - discard the highly frequent items
      - Use alternative measures such as h-confidence
- Computational Complexity
  - Binarizing the data increases the number of items
  - But the width of the “transactions” remain the same as the number of original (non-binarized) attributes
  - Produce more frequent itemsets but maximum size of frequent itemset is limited to the number of original attributes

NC STATE UNIVERSITY

Handling Continuous Attributes

- Different methods:
  - Discretization-based
  - Statistics-based
  - Non-discretization based
    - minApriori
- Different kinds of rules can be produced:
  - {Age∈[21,30), No of hours online∈[10,20)} → {Chat Online =Yes}
  - {Age∈[21,30), Chat Online = Yes} → No of hours online: μ=14, σ=4

NC STATE UNIVERSITY

Discretization-based Methods

Gender	···	Age	Annual Income	No of hours spent online per week	No of email accounts	Privacy Concern
Female	···	26	90K	20	4	Yes
Male	···	51	135K	10	2	No
Male	···	29	80K	10	3	Yes
Female	···	45	120K	15	3	Yes
Female	···	31	95K	20	5	Yes
Male	···	25	55K	25	5	Yes
Male	···	37	100K	10	1	No
Male	···	41	65K	8	2	No
Female	···	26	85K	12	1	No
···	···	···	···	···	···	···

Male	Female	··· ···	Age < 13	Age $\in [13,21)$	Age $\in [21,30)$	··· ···	Privacy = Yes	Privacy = No
0	1	···	0	0	1	···	1	0
1	0	···	0	0	0	···	0	1
1	0	···	0	0	1	···	1	0
0	1	···	0	0	0	···	1	0
0	1	···	0	0	0	···	1	0
1	0	···	0	0	1	···	1	0
1	0	···	0	0	0	···	0	1
1	0	···	0	0	0	···	0	1
0	1	···	0	0	1	···	0	1
···	···	···	···	···	···	···	···	···

NC STATE UNIVERSITY

Discretization-based Methods

- Unsupervised:
  - Equal-width binning <1 2 3> <4 5 6> <7 8 9>
  - Equal-depth binning <1 2 > <3 4 5 6 7 > < 8 9>
  - Cluster-based
- Supervised discretization

Continuous attribute, v

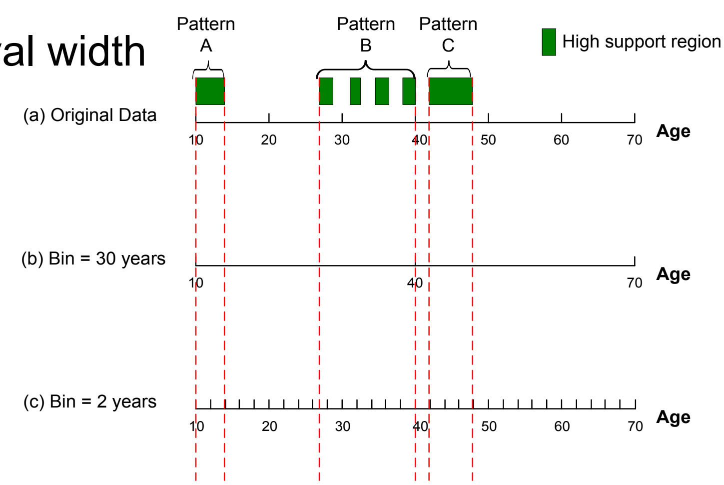
	1	2	3	4	5	6	7	8	9
Chat Online = Yes	0	0	20	10	20	0	0	0	0
Chat Online = No	150	100	0	0	0	100	100	150	100

bin_1 (1–2) bin_2 (3–5) bin_3 (6–9)

NC STATE UNIVERSITY

Discretization Issues

- Interval width



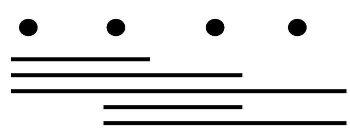
Pattern A: Age $\in [10, 15) \longrightarrow$ Chat Online = Never
Pattern B: Age $\in [26, 41) \longrightarrow$ Chat Online = Never
Pattern C: Age $\in [42, 48) \longrightarrow$ Online Banking = Yes

Discretization Issues

- Interval too wide (e.g., Bin size= 30)
  - May merge several disparate patterns
    - Patterns A and B are merged together
  - May lose some of the interesting patterns
    - Pattern C may not have enough confidence
- Interval too narrow (e.g., Bin size = 2)
  - Pattern A is broken up into two smaller patterns
    - Can recover the pattern by merging adjacent subpatterns
  - Pattern B is broken up into smaller patterns
    - Cannot recover the pattern by merging adjacent subpatterns
  - Some windows may not meet support threshold

Discretization: all possible intervals

Number of intervals = k
Total number of Adjacent intervals = k(k-1)/2



- Execution time
  - If the range is partitioned into k intervals, there are $O(k^2)$ new items
  - If an interval [a,b) is frequent, then all intervals that subsume [a,b) must also be frequent
    - E.g.: if {Age ∈[21,25), Chat Online=Yes} is frequent, then {Age ∈[10,50), Chat Online=Yes} is also frequent
  - Improve efficiency:
    - Use maximum support to avoid intervals that are too wide

Discretization Issues

- Redundant rules

R1: {Age $\in$[18,20), Age $\in$[10,12)} $\rightarrow$ {Chat Online=Yes}

R2: {Age $\in$[18,23), Age $\in$[10,20)} $\rightarrow$ {Chat Online=Yes}

  – If both rules have the same support and confidence, prune the more specific rule (R1)

Statistics-based Methods

- Example:

  {Income > 100K, Online Banking=Yes} $\rightarrow$ Age: μ=34
- Rule consequent consists of a continuous variable, characterized by their statistics
  – mean, median, standard deviation, etc.
- Approach:
  – Withhold the target attribute from the rest of the data
  – Extract frequent itemsets from the rest of the attributes
    - Binarized the continuous attributes (except for the target attribute)
  – For each frequent itemset, compute the corresponding descriptive statistics of the target attribute
    - Frequent itemset becomes a rule by introducing the target variable as rule consequent
  – Apply statistical test to determine interestingness of the rule

NC STATE UNIVERSITY

Statistics-based Methods

Gender	...	Age	Annual Income	No of hours spent online per week	No of email accounts	Privacy Concern
Female	...	26	90K	20	4	Yes
Male	...	51	135K	10	2	No
Male	...	29	80K	10	3	Yes
Female	...	45	120K	15	3	Yes
Female	...	31	95K	20	5	Yes
Male	...	25	55K	25	5	Yes
Male	...	37	100K	10	1	No
Male	...	41	65K	8	2	No
Female	...	26	85K	12	1	No
...	...	...	...	...	...	...

Frequent Itemsets:

{Male, Income > 100K}
{Income < 30K, No hours ∈[10,15)}
{Income > 100K, Online Banking = Yes}
....

Association Rules:

{Male, Income > 100K} → Age: μ = 30
{Income < 40K, No hours ∈[10,15)} → Age: μ = 24
{Income > 100K,Online Banking = Yes}
→ Age: μ = 34
....

NC STATE UNIVERSITY

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:

    $A \Rightarrow B: \mu$ versus $\cancel{A} \Rightarrow B: \mu'$

  - Statistical hypothesis testing:
    - Null hypothesis: H0: $\mu' = \mu + \Delta$
    - Alternative hypothesis: H1: $\mu' > \mu + \Delta$
    - Z has zero mean and variance 1 under null hypothesis

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

NC STATE UNIVERSITY

Statistics-based Methods

- Example:

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

  – Rule is interesting if difference between μ and μ' is more than 5 years (i.e., Δ = 5)
  – For r, suppose n1 = 50, s1 = 3.5
  – For r' (complement): n2 = 250, s2 = 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