Applied Data Science and Machine Intelligence

Association Rule Mining

Nandan Sudarsanam,
Department of Management Studies,
Robert Bosch Centre for Data Science and AI (RBC-DSAI),
Indian Institute of Technology Madras

Association Rule Mining

TID	Items
1	Bread, Milk,Coffee
2	Bread, Diaper, Baby Food, Eggs
3	Milk, Diaper, Baby Food, Coffee
4	Bread, Milk, Diaper, Baby Food
5	Bread, Milk, Diaper, Coffee

Example of associative rules:

{Diaper} → {Baby Food}
{Bread, Butter} → {Jam}

Definitions:

- Item set
- Antecedent
- Consequent

Association Rule Mining

- The rows of a data set represent the instances, and the columns correspond to a common set of attributes associated with the instances.
- ARM seeks to identify patterns or regularities between the attributes, based on the values they take up with relation to each other.
- A typical example of a pre-processed binary table which feeds into ARM algorithm which represents the presence or absence of an item in dataset.
- This can go well beyond a market-basket scenario
- So how is this analysis different from Clustering?

Customers	Gender	High income	High Expense	Owns Insurance	Lives in X	Age<30
T1	1	1	1	0	0	0
T2	1	0	0	1	1	1
T3	0	1	1	1	1	0
T4	1	1	0	1	1	0
T5	1	1	1	1	0	0

Association Rule Mining: challenges

-

$$R = \sum_{k=1}^{d-1}\left[\binom{d}{k} \times \sum_{j=1}^{d-k}\binom{d-k}{j}\right] = 3^d - 2^{d+1} + 1$$

Candidate Evaluation

-

Backup

Table 5
Objective measures for association patterns

#	Measure	Definition
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)	$\frac{\sum_j \max_k P(A_j,B_k)+\sum_k \max_j P(A_j,B_k)-\max_j P(A_j)-\max_k P(B_k)}{2-\max_j P(A_j)-\max_k P(B_k)}$
3	Odds ratio (α)	$\frac{P(A,B)P(\overline{A},\overline{B})}{P(A,\overline{B})P(\overline{A},B)}$
4	Yule's Q	$\frac{P(A,B)P(\overline{AB})-P(A,\overline{B})P(\overline{A},B)}{P(A,B)P(\overline{AB})+P(A,\overline{B})P(\overline{A},B)} = \frac{\alpha-1}{\alpha+1}$
5	Yule's Y	$\frac{\sqrt{P(A,B)P(\overline{AB})}-\sqrt{P(A,\overline{B})P(\overline{A},B)}}{\sqrt{P(A,B)P(\overline{AB})}+\sqrt{P(A,\overline{B})P(\overline{A},B)}} = \frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}$
6	Kappa (κ)	$\frac{P(A,B)+P(\overline{A},\overline{B})-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A)P(B)-P(\overline{A})P(\overline{B})}$
7	Mutual Information (M)	$\frac{\sum_i \sum_j P(A_i,B_j)\log\frac{P(A_i,B_j)}{P(A_i)P(B_j)}}{\min(-\sum_i P(A_i)\log P(A_i), -\sum_j P(B_j)\log P(B_j))}$
8	J-Measure (J)	$\max\big(P(A,B)\log(\frac{P(B\|A)}{P(B)}) + P(A\overline{B})\log(\frac{P(\overline{B}\|A)}{P(\overline{B})}),$ $P(A,B)\log(\frac{P(A\|B)}{P(A)}) + P(\overline{A}B)\log(\frac{P(\overline{A}\|B)}{P(\overline{A})})\big)$
9	Gini index (G)	$\max\big(P(A)[P(B\|A)^2 + P(\overline{B}\|A)^2] + P(\overline{A})[P(B\|\overline{A})^2 + P(\overline{B}\|\overline{A})^2]$ $-P(B)^2 - P(\overline{B})^2,$ $P(B)[P(A\|B)^2 + P(\overline{A}\|B)^2] + P(\overline{B})[P(A\|\overline{B})^2 + P(\overline{A}\|\overline{B})^2]$ $-P(A)^2 - P(\overline{A})^2\big)$
10	Support (s)	$P(A,B)$
11	Confidence (c)	$\max(P(B\|A), P(A\|B))$
12	Laplace (L)	$\max\big(\frac{NP(A,B)+1}{NP(A)+2}, \frac{NP(A,B)+1}{NP(B)+2}\big)$
13	Conviction (V)	$\max\big(\frac{P(A)P(\overline{B})}{P(A\overline{B})}, \frac{P(B)P(\overline{A})}{P(B\overline{A})}\big)$
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)	$\max\big(\frac{P(B\|A)-P(B)}{1-P(B)}, \frac{P(A\|B)-P(A)}{1-P(A)}\big)$
18	Added Value (AV)	$\max(P(B\|A) - P(B), P(A\|B) - P(A))$
19	Collective strength (S)	$\frac{P(A,B)+P(\overline{AB})}{P(A)P(B)+P(\overline{A})P(\overline{B})} \times \frac{1-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A,B)-P(\overline{AB})}$
20	Jaccard (ζ)	$\frac{P(A,B)}{P(A)+P(B)-P(A,B)}$
21	Klosgen (K)	$\sqrt{P(A,B)}\max(P(B\|A) - P(B), P(A\|B) - P(A))$

Tan, Pang-Ning, Vipin Kumar, and Jaideep Srivastava. "Selecting the right objective measure for association analysis." *Information Systems* 29.4 (2004): 293-313.

Backup

Table 6
Properties of objective measures. Note that none of the measures satisfies all the properties

Symbol	Measure	Range	P1	P2	P3	O1	O2	O3	O3'	O4
ϕ	ϕ-coefficient	$-1 \cdots 0 \cdots 1$	Yes	Yes	Yes	Yes	No	Yes	Yes	No
λ	Goodman-Kruskal's	$0 \cdots 1$	Yes	No	No	Yes	No	No*	Yes	No
α	odds ratio	$0 \cdots 1 \cdots \infty$	Yes*	Yes	Yes	Yes	Yes	Yes*	Yes	No
Q	Yule's Q	$-1 \cdots 0 \cdots 1$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Y	Yule's Y	$-1 \cdots 0 \cdots 1$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
κ	Cohen's	$-1 \cdots 0 \cdots 1$	Yes	Yes	Yes	Yes	No	No	Yes	No
M	Mutual information	$0 \cdots 1$	Yes	Yes	Yes	No**	No	No*	Yes	No
J	J-Measure	$0 \cdots 1$	Yes	No	No	No**	No	No	No	No
G	Gini index	$0 \cdots 1$	Yes	No	No	No**	No	No*	Yes	No
s	Support	$0 \cdots 1$	No	Yes	No	Yes	No	No	No	No
c	Confidence	$0 \cdots 1$	No	Yes	No	No**	No	No	No	Yes
L	Laplace	$0 \cdots 1$	No	Yes	No	No**	No	No	No	No
V	Conviction	$0.5 \cdots 1 \cdots \infty$	No	Yes	No	No**	No	No	Yes	No
I	Interest	$0 \cdots 1 \cdots \infty$	Yes*	Yes	Yes	Yes	No	No	No	No
IS	Cosine	$0 \cdots \sqrt{P(A,B)} \cdots 1$	No	Yes	Yes	Yes	No	No	No	Yes
PS	Piatetsky-Shapiro's	$-0.25 \cdots 0 \cdots 0.25$	Yes	Yes	Yes	Yes	No	Yes	Yes	No
F	Certainty factor	$-1 \cdots 0 \cdots 1$	Yes	Yes	Yes	No**	No	No	Yes	No
AV	Added value	$-0.5 \cdots 0 \cdots 1$	Yes	Yes	Yes	No**	No	No	No	No
S	Collective strength	$0 \cdots 1 \cdots \infty$	No	Yes	Yes	Yes	No	Yes*	Yes	No
ζ	Jaccard	$0 \cdots 1$	No	Yes	Yes	Yes	No	No	No	Yes
K	Klosgen's	$(\frac{2}{\sqrt{3}} - 1)^{1/2}[2 - \sqrt{3} - \frac{1}{\sqrt{3}}] \cdots 0 \cdots \frac{2}{3\sqrt{3}}$	Yes	Yes	Yes	No**	No	No	No	No

where: P1: $O(\mathbf{M}) = 0$ if $det(\mathbf{M}) = 0$, i.e., whenever A and B are statistically independent.
P2: $O(\mathbf{M_2}) > O(\mathbf{M_1})$ if $\mathbf{M_2} = \mathbf{M_1} + [k \ -k; \ -k \ k]$.
P3: $O(\mathbf{M_2}) < O(\mathbf{M_1})$ if $\mathbf{M_2} = \mathbf{M_1} + [0 \ k; \ 0 \ -k]$ or $\mathbf{M_2} = \mathbf{M_1} + [0 \ 0; \ k \ -k]$.
O1: Property 1: Symmetry under variable permutation.
O2: Property 2: Row and Column scaling invariance.
O3: Property 3: Antisymmetry under row or column permutation.
O3': Property 4: Inversion invariance.
O4: Property 5: Null invariance.
Yes*: Yes if measure is normalized.
No*: Symmetry under row or column permutation.
No**: No unless the measure is symmetrized by taking $\max(M(A,B), M(B,A))$.

Tan, Pang-Ning, Vipin Kumar, and Jaideep Srivastava. "Selecting the right objective measure for association analysis." *Information Systems* 29.4 (2004): 293-313.

ARM: Candidate generation

- The *a*PRIORI algorithm. With example:
- Transactions list: {1,2,3,4},{1,2,4},{1,2},{2,3,4},{2,3},{3,4},{2,4}
- Pick some threshold

- The biases from *a*PRIORI. The rare items problem

Apriori Algorithm

- If an itemset is frequent, then all of its subset 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

Apriori Algorithm

1. Find frequent 1-items and put them to L_k (k=1)
2. Use L_k to generate a collection of candidate itemsets C_{k+1} with size (k+1)
3. Scan the database to find which itemsets in C_{k+1} are frequent and put them into $_{Lk+1}$
4. If L_{k+1} is not empty
   - K=k+1
   - GOTO 2

Apriori Pseudocode

- C_k: Candidate itemset of size k
- L_k : frequent itemset of size k
- L_1 = {frequent items};
- for (k = 1; L_k !=∅; k++) do begin

  Ck+1 = candidates generated from L_k;

  // join and prune steps

  for each transaction t in database do

  increment the count of all candidates in C_{k+1} that are contained in t

  L_{k+1} = candidates in C_{k+1} with min_support (frequent)

  End
- Return $\cup_k L_k$;

Sample Problems

- Lift in terms of M:

$$P(AB)/(P(A).P(B)) = \frac{\frac{f11}{f11+f10+f01+f00}}{\frac{f11+f10}{f11+f10+f01+f00} \times \frac{f11+f01}{f11+f10+f01+f00}}$$

- Cosine in terms of M:

$$P(AB)/\sqrt{(P(A).P(B))}$$

What happens to $f00$?

FP-Growth

- Compress a large database into a compact, Frequent-Pattern tree (FP-tree) structure
  - highly condensed, but complete for frequent pattern mining
  - avoid costly database scans
- Develop an efficient, FP-tree and mine from tree
- pattern mining method
  - A divide-and-conquer methodology: decompose mining tasks into smaller ones
  - Avoid candidate generation: sub-database test only!

FP-Tree Steps

1. Scan DB once, find frequent 1-itemsets (single item patterns)
2. Order frequent items in descending order of their frequency
3. Scan DB again, construct FP-tree

Example from Tan, Steinbach and Kumar

- Example taken from: Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.

Transaction Data Set

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

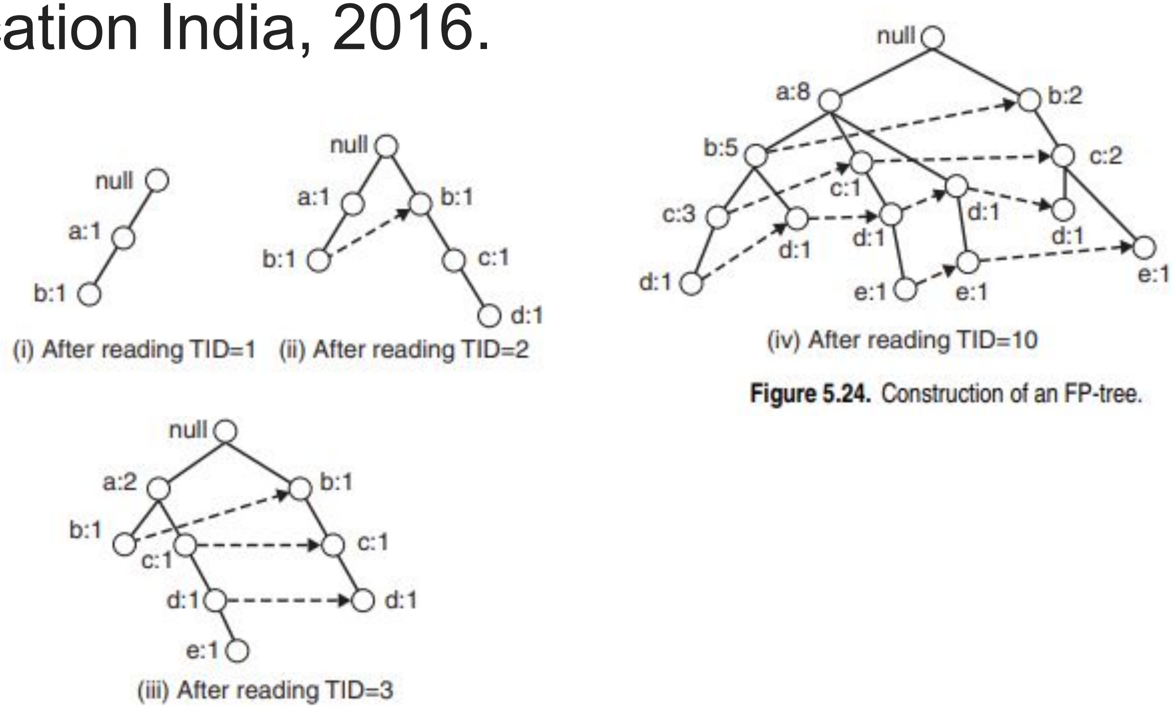


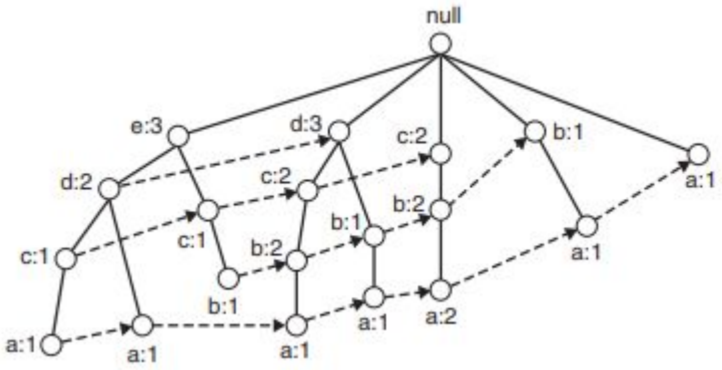
Figure 5.24. Construction of an FP-tree.

Example from Tan, Steinbach and Kumar

- How would the tree look like if reversed the order and transactions

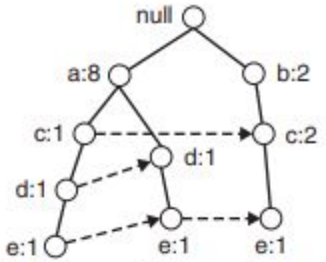
Transaction Data Set

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}



Example from Tan, Steinbach and Kumar

- How do you mine this tree?
- Let us say we start with FPs of e. Then move on ae, be, ce, de. And so on..
- Let us set min support to two
- Construct all paths containing e:

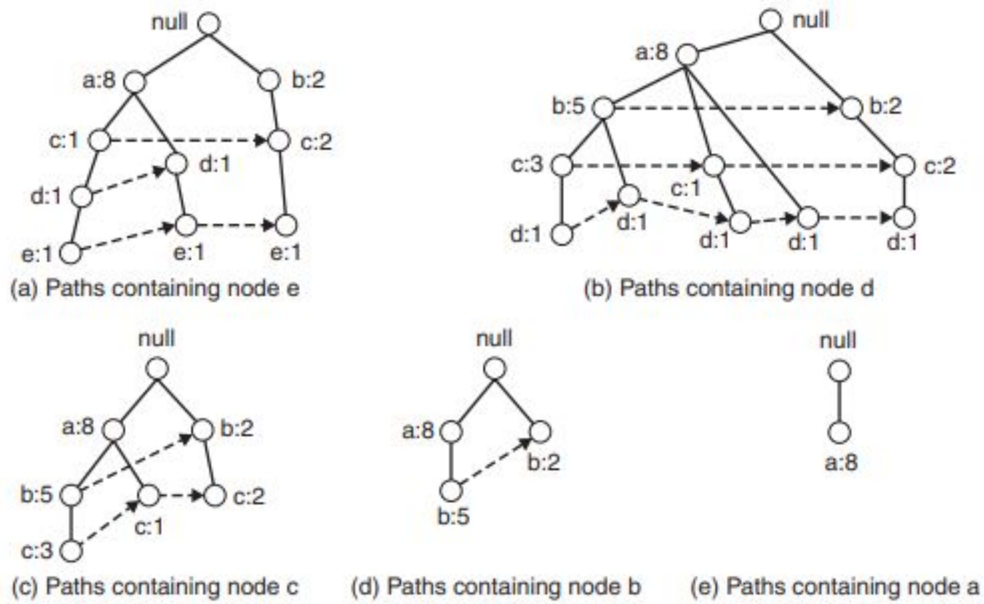


(a) Paths containing node e

Example from Tan, Steinbach and Kumar

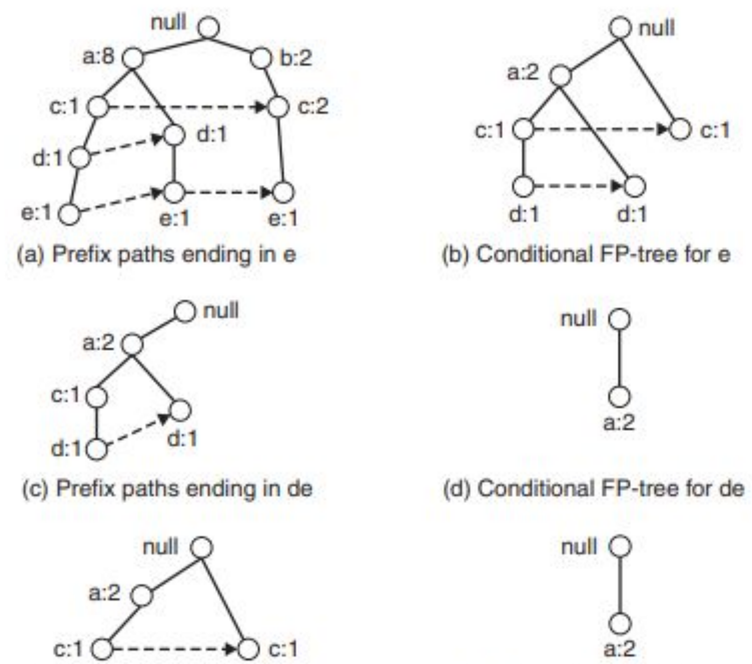
Transaction Data Set

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}



Example from Tan, Steinbach and Kumar

- Create a conditional FP tree for e
  - Remove nodes containing e
  - Remove infrequent items
- Use this to analyze de (c), ce (e), ae (f).
- What do you do after de (repeat the conditional FP tree step) (d)



(a) Prefix paths ending in e
(b) Conditional FP-tree for e
(c) Prefix paths ending in de
(d) Conditional FP-tree for de
(e) Prefix paths ending in ce
(f) Prefix paths ending in ae

Transaction Data Set

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

Properties of metrics

- Objective Measures
  - Objective measures observe some mathematical properties
    - The O matrix:

	B	B'	
A	f11	f10	f1+
A'	f01	f00	f0+
	f+1	f+0	N

    - Properties (note: M refers to the measure):
      - M = 0 if A and B are statistically independent
      - M Monotonically increases with P(A,B) when P(A) and P(B) are kept same
      - M Monotonically decreases with increase in P(A) when P(B) and P(A,B) are unchanged
      - Variable permutation
      - Row and column scaling
      - Row or column permutation and Inversion Invariance
      - Null Invariance

- Tan, Pang-Ning, Vipin Kumar, and Jaideep Srivastava. "Selecting the right objective measure for association analysis." *Information Systems* 29.4 (2004): 293-313.

Property 1

- O(M) = 0 if det(M) = 0, i.e., whenever A and B are statistically independent. Which implies P(A∩B) = P(A)*P(B)
- Hence f11f00 = f10f01 (can we derive this?)
- The measures that have f11f00 – f10f01 as their numerators end up being 0
- Eg: Φ-coefficient, Gooman-Kruskal's, odds ratio, Yule's Q, Yule's Y, Cohen's, Mutual information,J-Measure, Gini-index, etc
- Φ-coefficient (correlation co-efficient) for binary data can be summarized as $$\frac{P(A,B)-P(A)P(B)}{\sqrt{P(A)P(B)(1-P(A))(1-P(B))}}$$

Property 2

- M Monotonically increases with P(A,B) when P(A) and P(B) are kept same
- How do we increase P(A,B) while keeping P(A) and P(B) constant?
- $O(M_2) > O(M_1)$ if $M_2 = M_1 + \{k,-k;-k,k\}$.
- Eg: odds ratio, Yule's Q, Yule's Y, Cohen's, Mutual Information, Piateksy-Shapiro's, Cosine, Laplace etc.

Property 3

- M Monotonically decreases with increase in P(A) when P(B) and P(A,B) are unchanged
- How do we achieve this?
- $O(M_2)<O(M_1)$ if $M_2 = M_1 + \{0\ k;\ 0\ -k\}$ or $M_2 = M_1 + \{0\ 0;\ k\ -k\}$
- Eg: Phi-coefficient, odds ratio, Yule's Q, Yule's Y, Cohen's, Mutual information, Interest etc.

Property 4

- Variable permutation
- A measure O is symmetric under variable permutation (Fig. 1(a)),

  A↔B; if $O(M^T) = O(M)$ for all contingency matrices M: Otherwise, it is called an asymmetric measure.
- In practice, asymmetric measures are used for implication rules, where there is a need to distinguish between the strength of the rule A->B from B->A.
- In some cases people might choose the max of the two possible outcomes if the rule is asymmetric (does not follow this property) if they do not have a specific implication in mind.
- Eg: phi-coefficient, Goodman-Kruskal, odd's ratio, Yule's Q, Yule's Y, Cohen's etc

Property 5

- Row and column scaling
- Let R = C = [k_1 0; 0 k_2] be a 2 X 2 square matrix, where k_1 and k_2 are positive constants. The product R M corresponds to scaling the first row of matrix M by k_1 and the second row by k_2; while the product M C corresponds to scaling the first column of M by k_1 and the second column by k_2.
- A measure O is invariant under row and column scaling if O(RM) = O(M) and O(MC) = O(M) for all contingency matrices, M.
- This property is useful for data sets containing nominal variables such as Mosteller's grade-gender example:

	Male	Female	
High	2	3	5
Low	1	4	5
	3	7	10

	Male	Female	
High	4	30	34
Low	2	40	42
	6	70	76

- Eg: odd's ratio, Yule's Q, Yules Y, Cohen's are invariant.

Property 6

- Antisymmetry under row/column permutation
- Let S = [0 1; 1 0] be a 2 X 2, permutation matrix. A normalized measure O is antisymmetric under the row permutation operation if O(SM)= - O(M); and antisymmetric under the column permutation operation if O(MS)= O(M) for all contingency matrices M.
- The phi-coefficient, PS; Q and Y are examples of antisymmetric measures under the row and column permutation operations while mutual information and Gini index are examples of symmetric measures. Asymmetric measures under this operation include support, confidence, IS and interest factor.
- Measures that are symmetric under the row and column permutation operations do not distinguish between positive and negative correlations of a table. One should be careful when using them to evaluate the interestingness of a pattern

Property 7

- Inversion invariance
- Let S = [0 1; 1 0] be a 2 X 2 permutation matrix. A measure O is invariant under the inversion operation if O(SMS)= O(M) for all contingency matrices M.
- This property helps in highlighting the drawbacks using f-coefficient and other symmetric binary measures for applications that require unequal treatments of the binary values of a variable, such as market basket analysis
- Examples of symmetric binary measures include f; odds ratio, k and collective strength, while the examples for asymmetric binary measures include I; IS; PS and Jaccard measure.

Property 8

- Null Invariance
- A measure is null invariant if O(M + C) = O(M) where C = [0 0; 0 k] and k is a positive constant.
- For binary variables, this operation corresponds to adding more records that do not contain the two variables under consideration
- Some of the null-invariant measures include IS (cosine) and the Jaccard similarity measure, z
- This property is useful for domains having sparse data sets, where co-presence of items is more important than co-absence. Examples include market-basket data and text documents.