

Regras de Associação

ACH5504 – Mineração de Dados

Notas de aulas baseadas no livro

“Introduction to Data Mining”

Tan, Steinbach, Karpatne, Kumar

Resumo

- Definição de associação
- Abordagem de força bruta

O que é Análise de Padrões Frequentes?

- **Padrões frequentes**: um padrão (um conjunto de itens, subsequências, subestruturas, etc.) que ocorrem frequentemente em um conjunto de dados
- Proposto inicialmente por Agrawal, Imieliński e Swamu (1993) no contexto de **conjunto de itens frequentes** e **mineração de regras de associação**
- Motivação: Encontrar regularidades em dados
 - Que produtos são usualmente comprados em conjunto?
 - Quais são as compras subsequentes a compra de um PC?
 - Que tipos de DNA são sensíveis a uma droga?
 - Podemos classificar automaticamente os documentos da Web?
- Aplicações
 - Análise de dados de cesta, cross-marketing, design de catálogos, análise de campanha de venda, análise de log de Web (fluxo de clique) e análise de sequência de DNA.

Mineração de Regras de Associação

- Dado um conjunto de transações, encontre as regras que irão prever a ocorrência de um item baseando nas ocorrências de outros itens em transação

Transações de cesta de mercado

<i>TID</i>	<i>Itens</i>
1	Pão, Leite
2	Pão, Fralda, Cerveja, Ovos
3	Leite, Fralda, Cerveja, Coca
4	Pão, Leite, Fralda, Cerveja
5	Pão, Leite, Fralda, Coca

Exemplo de regras de associação

$\{\text{Fralda}\} \rightarrow \{\text{Cerveja}\},$
 $\{\text{Leite, Pão}\} \rightarrow \{\text{Ovos, Coca}\},$
 $\{\text{Cerveja, Pão}\} \rightarrow \{\text{Leite}\},$

Implicação significa **co-ocorrência**,
não causalidade!

Definição: Conjunto de Itens Frequentes

- **Conjunto de itens**

- Uma coleção de um ou mais itens
 - Exemplo: {Leite, Pão, Fralda}
- Conjunto de k itens
 - Um conjunto que contem k itens

- **Contagem de suporte (σ)**

- A frequência de ocorrência de um conjunto de itens
- E.g. $\sigma(\{\text{Leite, Pão, Fralda}\}) = 2$

- **Suporte**

- A fração de transações que contem um conjunto de itens
- E.g. $s(\{\text{Leite, Pão, Fralda}\}) = 2/5$

- **Conjunto de itens frequente**

- Um conjunto de itens com suporte maior ou igual do que o limiar *minsup*

<i>TID</i>	<i>Itens</i>
1	Pão, Leite
2	Pão, Fralda, Cerveja, Ovos
3	Leite, Fralda, Cerveja, Coca
4	Pão, Leite, Fralda, Cerveja
5	Pão, Leite, Fralda, Coca

Definição: Regra de Associação

- **Regra de associação**

- Uma expressão de implicação de forma $X \rightarrow Y$, onde X e Y são conjuntos de itens
- Exemplo:
 $\{\text{Leite, Fralda}\} \rightarrow \{\text{Cerveja}\}$

<i>TID</i>	<i>Itens</i>
1	Pão, Leite
2	Pão, Fralda, Cerveja, Ovos
3	Leite, Fralda, Cerveja, Coca
4	Pão, Leite, Fralda, Cerveja
5	Pão, Leite, Fralda, Coca

- **Métricas de avaliação de regras**

- Suporte (s)
 - ◆ Fração (probabilidade) de transações que contem tanto X como Y
- Confiança (c)
 - ◆ Mede a frequência (probabilidade condicional) com que os itens em Y aparecem também em transações que contem X

Exemplo:

$\{\text{Leite, Fralda}\} \rightarrow \{\text{Cerveja}\}$

$$s = \frac{\sigma(\text{Leite, Fralda, Cerveja})}{|T|} = \frac{2}{5} = 0.4$$

$$c = \frac{\sigma(\text{Leite, Fralda, Cerveja})}{\sigma(\text{Leite, Fralda})} = \frac{2}{3} = 0.67$$

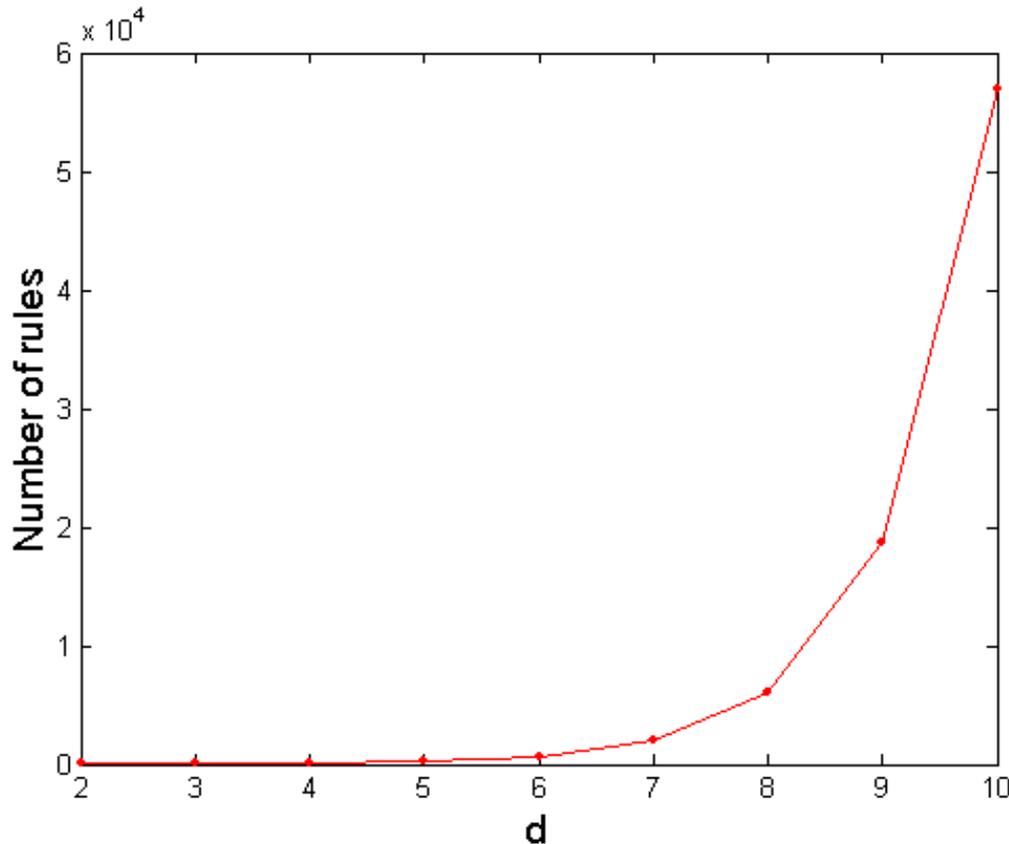
Tarefa de Mineração de Regras de Associação

- Dado um conjunto de transações T , o objetivo da mineração de regras de associação é encontrar todas as regras que têm
 - suporte $\geq \textit{minsup}$
 - confiança $\geq \textit{minconf}$
- Abordagem de força bruta:
 - Liste todas possíveis regras de associação
 - Compute o suporte e confiança de cada regra
 - Elimine regras que falham em \textit{minsup} e $\textit{minconf}$

⇒ **Computacionalmente proibitivo!**

Complexidade computacional

- Dado d itens únicos:
 - O número total de conjuntos de itens = 2^d
 - O número total de possíveis regras de associação:



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

**Se $d = 6$, termos
 $R = 602$ regras**

Minerando as Regras de Associação

<i>TID</i>	<i>Itens</i>
1	Pão, Leite
2	Pão, Fralda, Cerveja, Ovos
3	Leite, Fralda, Cerveja, Coca
4	Pão, Leite, Fralda, Cerveja
5	Pão, Leite, Fralda, Coca

Exemplos de regras:

{Leite, Fralda} → {Cerveja} (s=0.4, c=0.67)

{Leite, Cerveja} → {Fralda} (s=0.4, c=1.0)

{Fralda, Cerveja} → {Leite} (s=0.4, c=0.67)

{Cerveja} → {Leite, Fralda} (s=0.4, c=0.67)

{Fralda} → {Leite, Cerveja} (s=0.4, c=0.5)

{Leite} → {Fralda, Cerveja} (s=0.4, c=0.5)

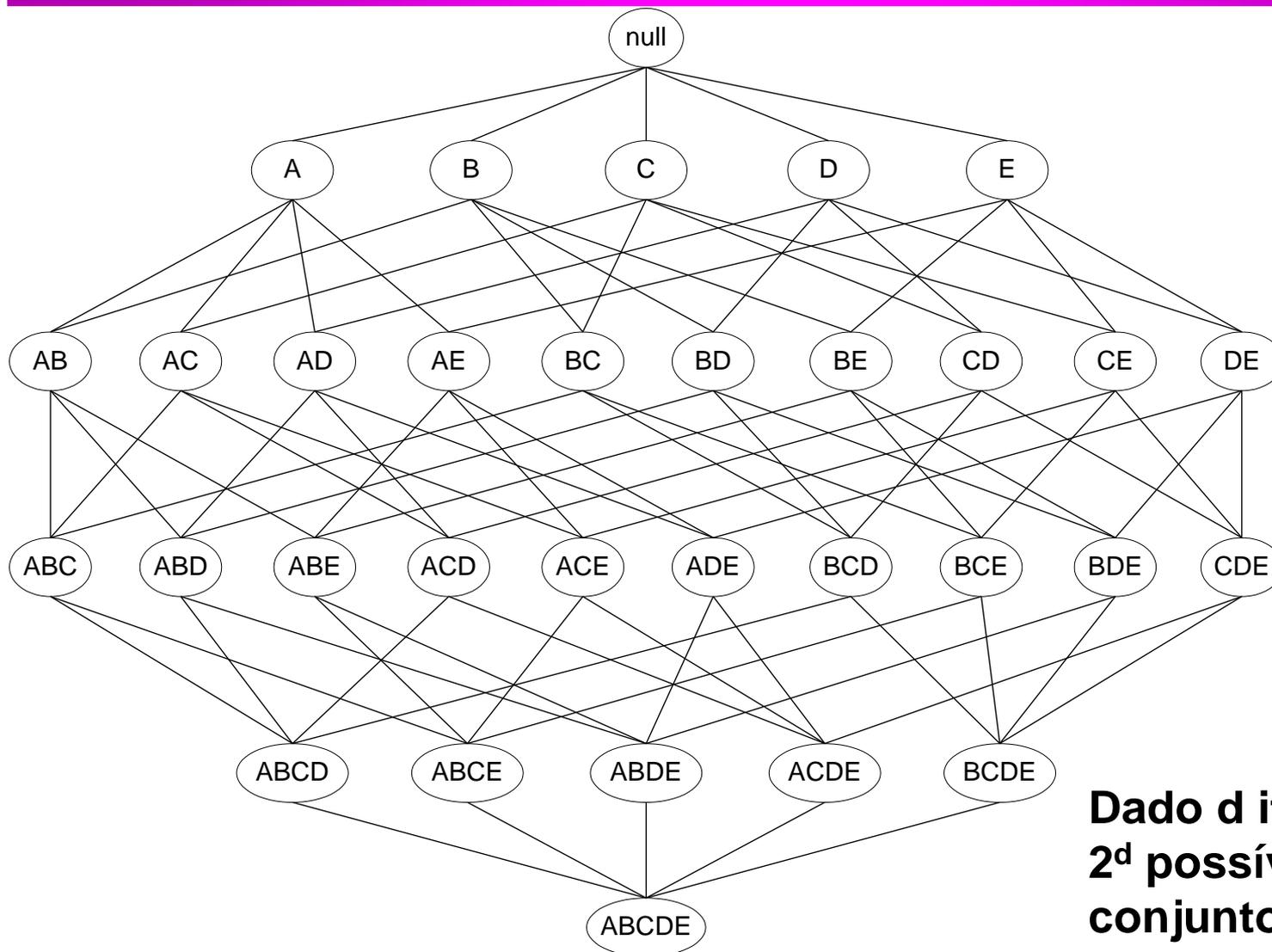
Observações:

- Todas as regras acima são partições do mesmo conjunto de itens:
{Leite, Fralda, Cerveja}
- O suporte é igual para regras que tem origem do mesmo conjunto de itens, a confiança pode ser diferente
- Assim, poderemos desacoplar os requisitos de suporte e confiança

Minerando as Regras de Associação

- Abordagem de dois passos:
 1. **Geração de conjuntos frequentes de itens**
 - Gere todos conjuntos de itens com suporte \geq minsup
 2. **Geração de regras**
 - Gere regras de alta confiança para cada conjunto frequente de itens, em que cada regra é uma partição binária do conjunto frequente de itens
- Geração de conjuntos frequentes de itens ainda é computacionalmente cara

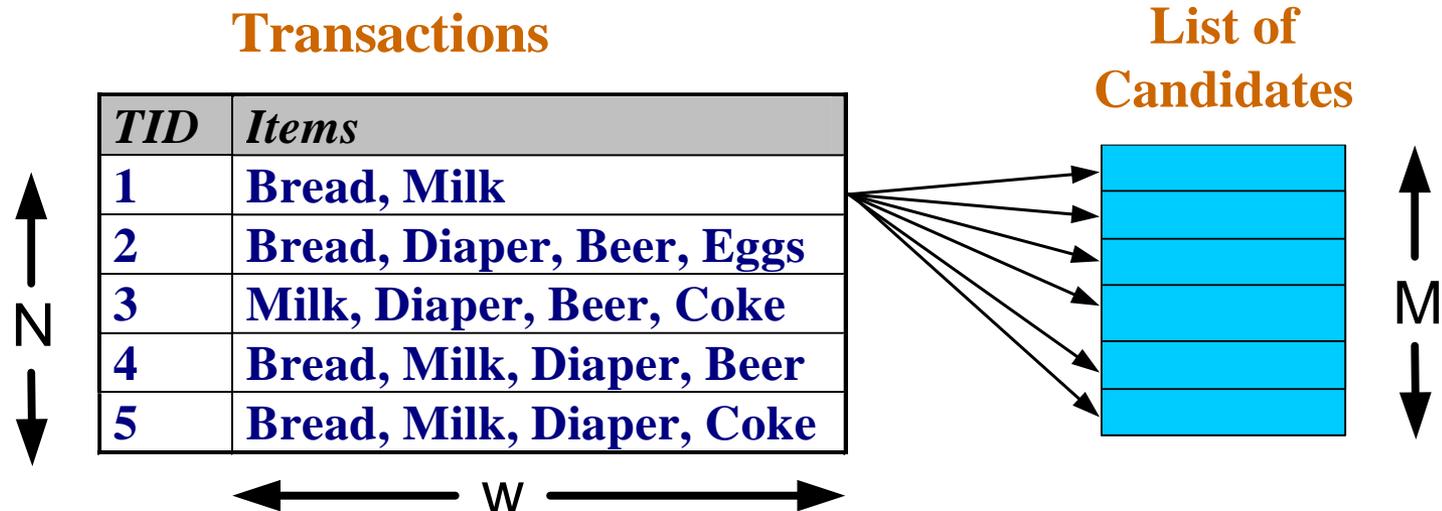
Geração de Conjuntos Frequentes



**Dado d itens, existem
2^d possíveis
conjuntos de itens
candidatos**

Geração de Conjuntos Frequentes

- Abordagem de força bruta:
 - Cada conjunto de itens na estrutura é um **candidato** para conjunto frequente
 - Conte o suporte de cada candidato



- Combine cada transação com cada candidato
- Complexidade $\sim O(N \times M \times w) \Rightarrow$ **Caro, desde $M = 2^d$!!!**

Estratégias de Geração de Conjuntos Frequentes

- Reduza o **número de candidatos** (M)
 - Pesquisa completa: $M=2^d$
 - Use técnicas de eliminação para reduzir M
- Reduza o **número de transações** (N)
 - Reduza N conforme o tamanho de conjuntos aumenta
- Reduza o **número de comparações** (NM)
 - Use estruturas de dados eficientes para armazenar os candidatos ou transações
 - Não há necessidade de comparar todos os candidatos contra cada transação

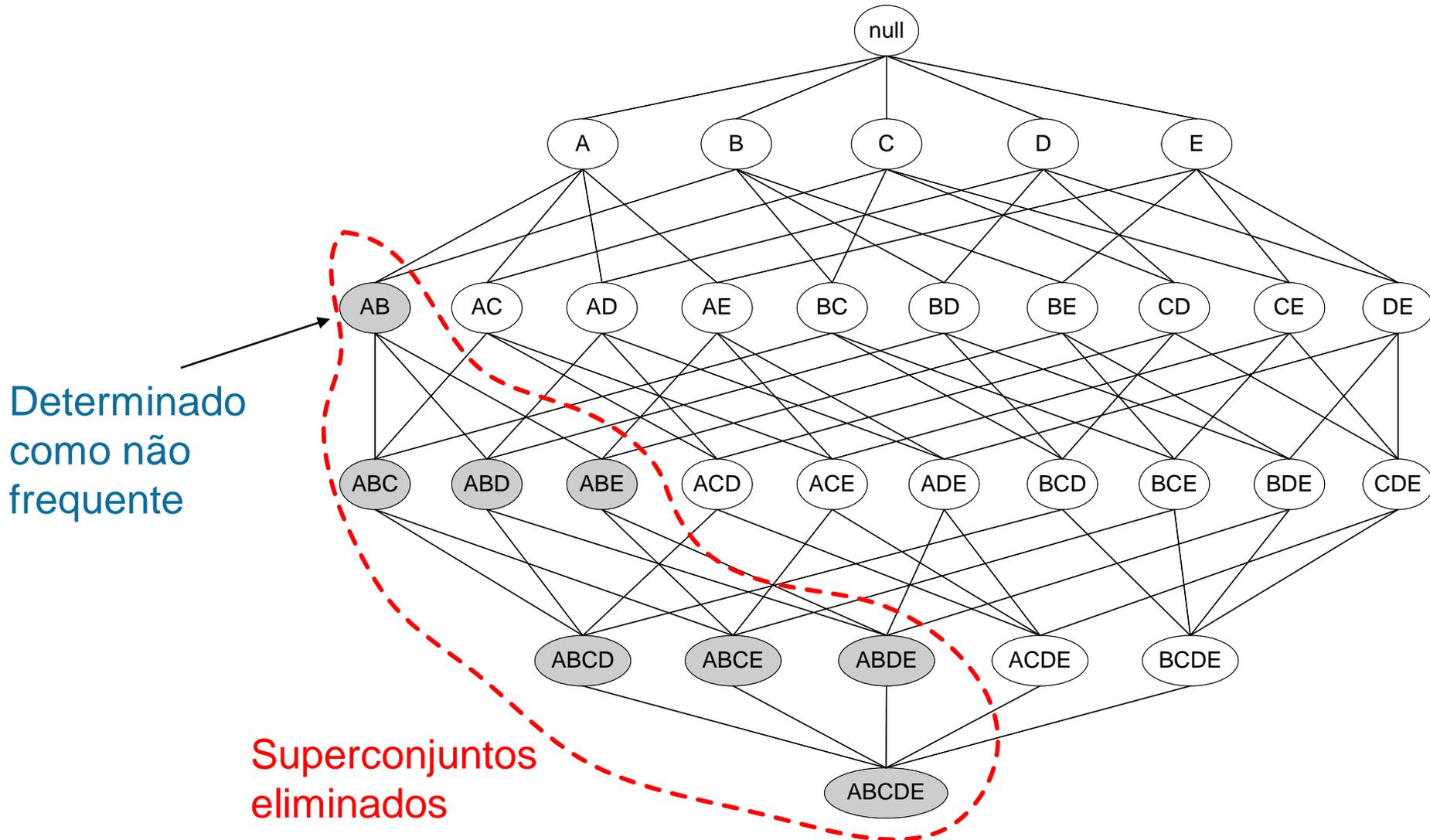
Reduzindo o Número de Candidatos

- **Princípio Apriori:**
 - Se um conjunto de itens é frequente, então todos os seus subconjuntos também devem ser frequentes
- O princípio Apriori se aplica por causa da seguinte propriedade de medida de suporte:

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

- O suporte de um conjunto de itens nunca excede o suporte de seus subconjuntos
- É conhecida como a propriedade **anti-monótona** do suporte

Demonstrando o princípio Apriori



Mostrando o princípio Apriori

<i>TID</i>	<i>Itens</i>
1	Pão, Leite
2	Pão, Fralda, Cerveja, Ovos
3	Leite, Fralda, Cerveja, Coca
4	Pão, Leite, Fralda, Cerveja
5	Pão, Leite, Fralda, Coca



Itens (1-conjunto)

Item	Contagem
Pão	4
Coca	2
Leite	4
Cerveja	3
Fralda	4
Ovos	1

Suporte Mínimo = 3

Se todos os subconjuntos seriam considerados,

$${}^6C_1 + {}^6C_2 + {}^6C_3$$

$$6 + 15 + 20 = \mathbf{41}$$

Com eliminação baseada no suporte,

$$6 + 6 + 4 = \mathbf{16}$$

Mostrando o princípio Apriori

<i>TID</i>	<i>Itens</i>
1	Pão, Leite
2	Pão, Fralda, Cerveja, Ovos
3	Leite, Fralda, Cerveja, Coca
4	Pão, Leite, Fralda, Cerveja
5	Pão, Leite, Fralda, Coca



Itens (1-conjunto)

Item	Contagem
Pão	4
Coca	2
Leite	4
Cerveja	3
Fralda	4
Ovos	1

Suporte Mínimo = 3

Se todos os subconjuntos seriam considerados,

$${}^6C_1 + {}^6C_2 + {}^6C_3$$

$$6 + 15 + 20 = \mathbf{41}$$

Com eliminação baseada no suporte,

$$6 + 6 + 4 = \mathbf{16}$$

Mostrando o princípio Apriori

Item	Contagem
Pão	4
Coca	2
Leite	4
Cerveja	3
Fralda	4
Ovos	1

Itens (1-conjunto)



Conjunto
{Pão, Leite}
{Pão, Cerveja}
{Pão, Fralda}
{Cerveja, Leite}
{Fralda, Leite}
{Cerveja, Fralda}

Pares (2-conjuntos)

(Não há necessidade de gerar candidatos com Coca ou Ovos)

Suporte Mínimo = 3

Se todos os subconjuntos seriam considerados,

$${}^6C_1 + {}^6C_2 + {}^6C_3$$

$$6 + 15 + 20 = \mathbf{41}$$

Com eliminação baseada no suporte,

$$6 + 6 + 4 = \mathbf{16}$$

Mostrando o princípio Apriori

Item	Contagem
Pão	4
Coca	2
Leite	4
Cerveja	3
Fralda	4
Ovos	1

Itens (1-conjunto)



Conjunto	Contagem
{Pão, Leite}	3
{Pão, Cerveja}	2
{Pão, Fralda}	3
{Leite, Cerveja}	2
{Leite, Fralda}	3
{Cerveja, Fralda}	3

Pares (2-conjuntos)

(Não há necessidade de gerar candidatos com Coca ou Ovos)

Suporte Mínimo = 3

Se todos os subconjuntos seriam considerados,

$${}^6C_1 + {}^6C_2 + {}^6C_3$$

$$6 + 15 + 20 = \mathbf{41}$$

Com eliminação baseada no suporte,

$$6 + 6 + 4 = \mathbf{16}$$

Mostrando o princípio Apriori

Item	Contagem
Pão	4
Coca	2
Leite	4
Cerveja	3
Fralda	4
Ovos	1

Itens (1-conjunto)



Conjunto	Contagem
{Pão, Leite}	3
{Pão, Cerveja}	2
{Pão, Fralda}	3
{Leite, Cerveja}	2
{Leite, Fralda}	3
{Cerveja, Fralda}	3

Pares (2-conjuntos)

(Não há necessidade de gerar candidatos com Coca ou Ovos)

Suporte Mínimo = 3

Se todos os subconjuntos seriam considerados,

$${}^6C_1 + {}^6C_2 + {}^6C_3$$

$$6 + 15 + 20 = \mathbf{41}$$

Com eliminação baseada no suporte,

$$6 + 6 + 4 = \mathbf{16}$$



Trios (3-conjuntos)

Conjuntos
{Cerveja, Fralda, Leite}
{Cerveja, Pão, Fralda}
{Pão, Fralda, Leite}
{Cerveja, Pão, Leite}

Mostrando o princípio Apriori

Item	Contagem
Pão	4
Coca	2
Leite	4
Cerveja	3
Fralda	4
Ovos	1

Itens (1-conjunto)



Conjunto	Contagem
{Pão, Leite}	3
{Pão, Cerveja}	2
{Pão, Fralda}	3
{Leite, Cerveja}	2
{Leite, Fralda}	3
{Cerveja, Fralda}	3

Pares (2-conjuntos)

(Não há necessidade de gerar candidatos com Coca ou Ovos)

Suporte Mínimo = 3

Se todos os subconjuntos seriam considerados,

$${}^6C_1 + {}^6C_2 + {}^6C_3$$

$$6 + 15 + 20 = \mathbf{41}$$

Com eliminação baseada no suporte,

$$6 + 6 + 4 = \mathbf{16}$$



Trios (3-conjuntos)

Conjunto	Contagem
{Cerveja, Fralda, Leite}	2
{Cerveja, Pão, Fralda}	2
{Pão, Fralda, Leite}	2
{Cerveja, Pão, Leite}	1

Algoritmo Apriori

- F_k : conjunto frequente de k itens
- L_k : candidato para conjunto de k itens
- Algoritmo
 - Deixe $k=1$
 - Gere $F_1 = \{\text{conjunto frequente de 1 item}\}$
 - Repete até F_k está vazio
 - **Geração de Candidato:** Gere L_{k+1} usando F_k
 - **Eliminação de Candidato:** Elimine o candidato em L_{k+1} que contem subconjuntos de tamanho k que são infrequentes
 - **Contagem de Suporte:** Conte o suporte para cada candidato em L_{k+1} examinando o banco de dados
 - **Eliminação de Candidato:** Elimine os candidatos infrequentes em L_{k+1} , deixando somente os frequentes $\Rightarrow F_{k+1}$

Algoritmo Apriori - Exemplo

Banco de dados $Sup_{min} = 2$

Tid	Itens
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

1º scan

C_1

Conjunto	sup
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3

L_1

Conjunto	sup
{A}	2
{B}	3
{C}	3
{E}	3

L_2

Conjunto	sup
{A, C}	2
{B, C}	2
{B, E}	3
{C, E}	2

C_2

Conjunto	sup
{A, B}	1
{A, C}	2
{A, E}	1
{B, C}	2
{B, E}	3
{C, E}	2

2º scan

C_2

Conjunto
{A, B}
{A, C}
{A, E}
{B, C}
{B, E}
{C, E}

C_3

Conjunto
{B, C, E}

3º scan

L_3

Conjunto	sup
{B, C, E}	2

Geração de Candidatos

- Combine dois conjuntos frequentes de (k-1) itens, se os primeiros (k-2) itens são idênticos
- $F_3 = \{ABC, ABD, ABE, ACD, BCD, BDE, CDE\}$
 - Combine(ABC, ABD) = ABCD
 - Combine(ABC, ABE) = ABCE
 - Combine(ABD, ABE) = ABDE
 - Não combine (ABD, ACD) porque eles compartilham apenas o prefixo de comprimento 1 em vez de comprimento 2

Eliminação de Candidato

- Seja $F_3 = \{ABC, ABD, ABE, ACD, BCD, BDE, CDE\}$ o conjunto frequente de 3 itens
- $L_4 = \{ABCD, ABCE, ABDE\}$ é o candidato de conjunto de 4 itens gerado
- Elimine candidato
 - Elimine ABCE porque ACE e BCE são infrequentes
 - Elimine ABDE porque ADE é infrequente
- Depois da eliminação temos: $L_4 = \{ABCD\}$

Contagem de Suporte de Conjuntos de Candidato

- Varra o bando de dados de transações para determinar o suporte de cada candidato
 - Deve corresponder a cada candidato em cada transação, o que é uma operação cara

<i>TID</i>	<i>Itens</i>
1	Pão, Leite
2	Cerveja, Pão, Fralda, Ovos
3	Cerveja, Coca, Fralda, Leite
4	Cerveja, Pão, Fralda, Leite
5	Pão, Coca, Fralda, Leite

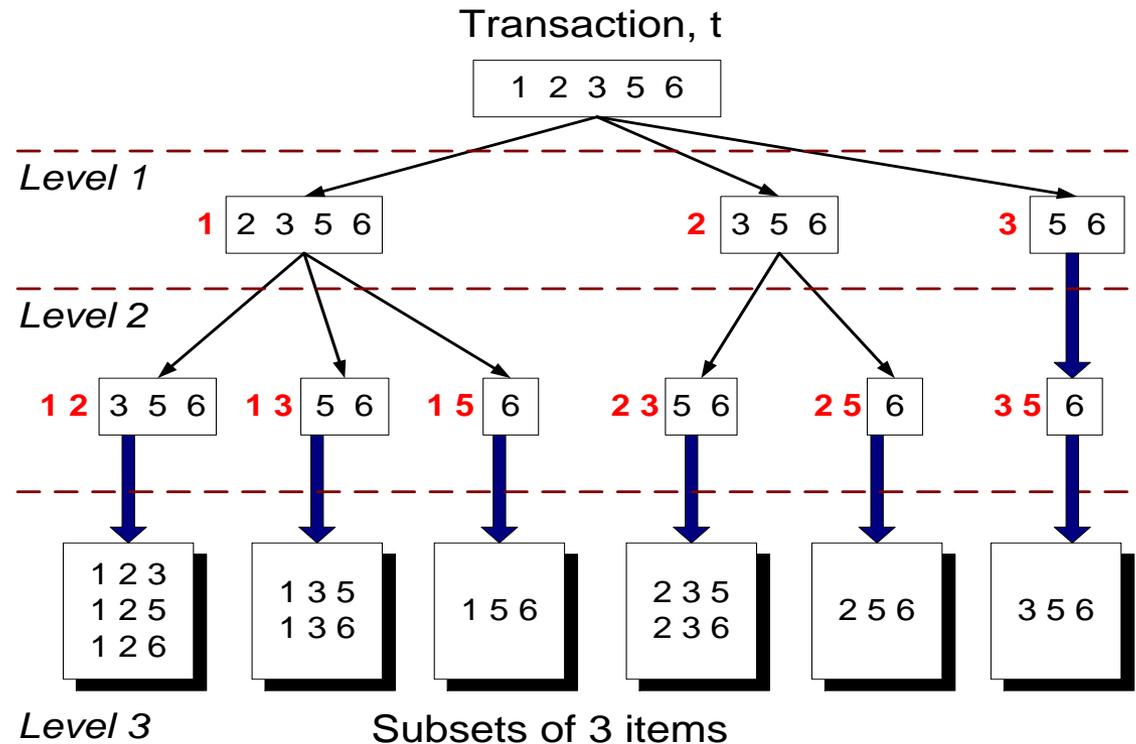
Conjunto
{ Cerveja, Fralda, Leite }
{ Cerveja, Pão, Fralda }
{ Pão, Fralda, Leite }
{ Cerveja, Pão, Leite }

Contagem de Suporte: Exemplo

Suponha que você tem 15 candidatos de conjuntos de itens de comprimento 3:

{1 4 5}, {1 2 4}, {4 5 7}, {1 2 5}, {4 5 8}, {1 5 9}, {1 3 6}, {2 3 4}, {5 6 7}, {3 4 5}, {3 5 6}, {3 5 7}, {6 8 9}, {3 6 7}, {3 6 8}

Quantos conjuntos são suportados pela transação (1,2,3,5,6)?



Geração de Regras

- Dado um conjunto frequente L , encontre todos subconjuntos não vazios $f \subset L$ tais que $f \rightarrow L - f$ satisfaz o requisito de confiança mínima
 - Se $\{A, B, C, D\}$ for um conjunto frequente de itens, as regras do candidato:

$ABC \rightarrow D,$	$ABD \rightarrow C,$	$ACD \rightarrow B,$	$BCD \rightarrow A,$
$A \rightarrow BCD,$	$B \rightarrow ACD,$	$C \rightarrow ABD,$	$D \rightarrow ABC$
$AB \rightarrow CD,$	$AC \rightarrow BD,$	$AD \rightarrow BC,$	$BC \rightarrow AD,$
$BD \rightarrow AC,$	$CD \rightarrow AB,$		

- Se $|L| = k$, então há $2^k - 2$ candidatos de regras de associação (ignorando $L \rightarrow \emptyset$ e $\emptyset \rightarrow L$)

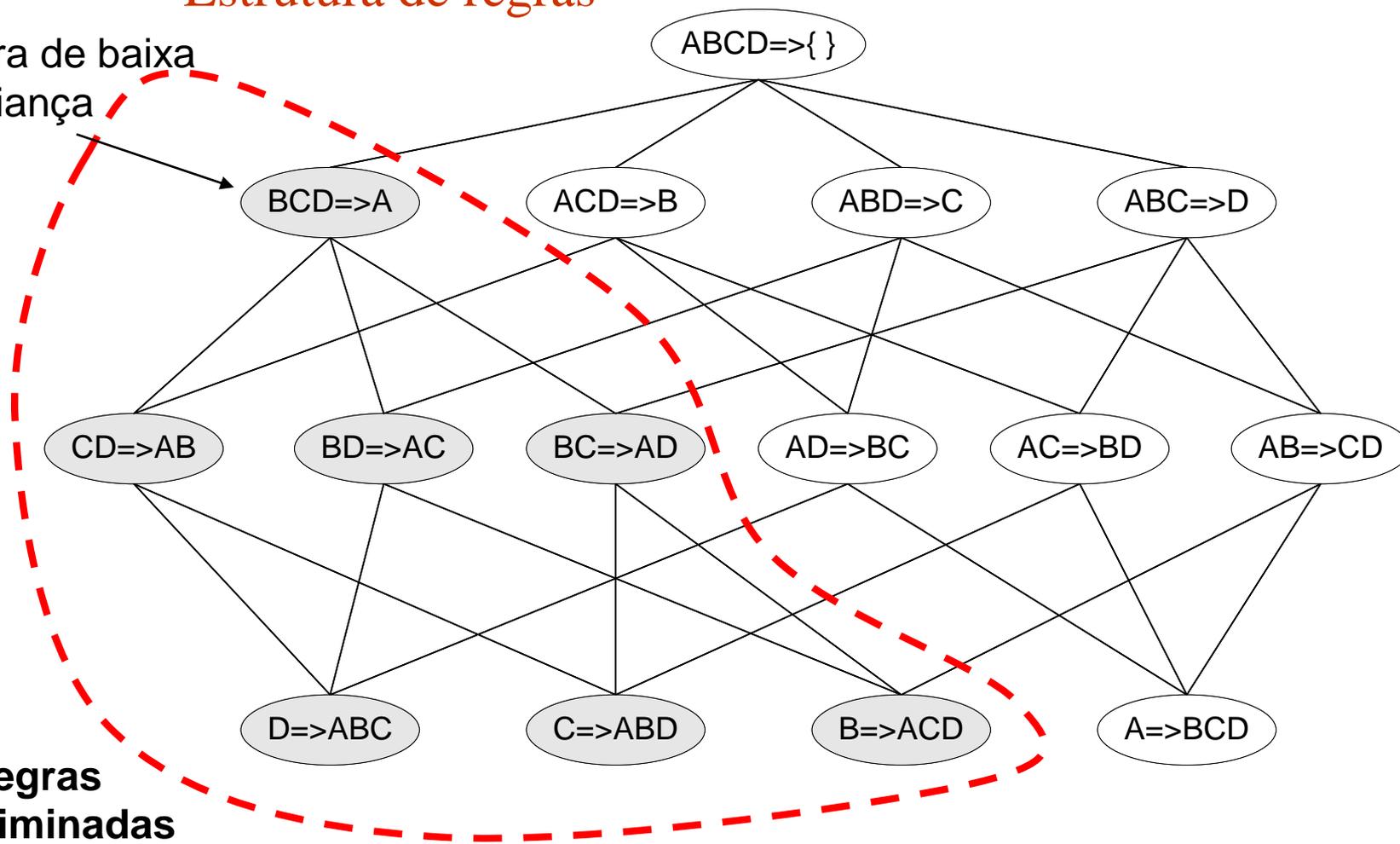
Geração de Regras

- Em geral, a confiança não tem uma propriedade anti-monótona
 - $c(ABC \rightarrow D)$ pode ser maior ou menor do que $c(AB \rightarrow D)$
- Mas a confiança das regras geradas a partir do mesmo conjunto de itens tem uma propriedade anti-monótona
 - E.g., Suponha que $\{A,B,C,D\}$ é um conjunto frequente de 4-itens:
 - $$c(ABC \rightarrow D) \geq c(AB \rightarrow CD) \geq c(A \rightarrow BCD)$$
 - A confiança é anti-monótona em relação ao número de itens do lado direito da regra

Geração de Regras para Algoritmo Apriori

Estrutura de regras

Regra de baixa
confiança



Problemas da Abordagem Apriori

- Principais desafios computacionais
 - Múltiplas varreduras da base de dados
 - Enorme quantidade de candidatos
 - Carga de trabalho de contagem de suporte
- Possíveis melhoras
 - Reduzir passagens na base de dados
 - Diminuir quantidade de candidatos
 - Facilitar contagem de suporte
 - Amostragem da base de dados

Fatores que afetam a complexidade de Apriori

- Escolha do limiar de suporte mínimo
 - diminuir o suporte resulta em mais conjuntos frequentes de itens
 - isso pode aumentar o número de candidatos e tamanho máximo dos conjuntos frequentes de itens
- Dimensionalidade (número de itens)
 - mais espaço é necessário para guardar a contagem de suporte de cada item
 - se o número de itens frequentes também aumenta, tanto custos de computação como I/O aumentam
- Tamanho da base de dados
 - como Apriori faz múltiplas varreduras, o tempo de execução do algoritmo pode aumentar com o número de transações
- Tamanho médio da transação
 - tamanho da transação aumenta com conjuntos de dados mais densos
 - isso pode aumentar comprimento máximo de conjuntos frequentes de itens e varreduras por subconjuntos em uma transação (maior quantidade de subconjuntos)

Avaliação de padrões

- Métodos de regras de associação tendem a produzir muitas regras
 - Muitas não são interessantes ou são redundantes
 - Redundante: se $\{A,B,C\} \rightarrow \{D\}$ e $\{A,B\} \rightarrow \{D\}$ tem mesmo suporte e confiança
- Medidas de interesse podem ser usadas para eliminar ou ranquear os padrões obtidos
 - Na formulação original de regras de associação, suporte e confiança são as únicas medidas usadas

Calculando a Medida de Interesse

- Dada uma regra $X \rightarrow Y$ ou $\{X, Y\}$, a informação para analisar interesse pode ser obtida da tabela de contigência

Tabela de contigência

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

f_{11} : suporte de X e Y

f_{10} : suporte de X e \bar{Y}

f_{01} : suporte de \bar{X} e Y

f_{00} : suporte de \bar{X} e \bar{Y}

Usada para definir várias medidas

- suporte, confiança, Gini, entropia, etc.

Limitações de confiança

Cientes	Chá	Café	...
C1	0	1	...
C2	1	0	...
C3	1	1	...
C4	1	0	...
...			

	Café	<u>Café</u>	
Chá	15	5	20
<u>Chá</u>	75	5	80
	90	10	100

Regra de associação: Chá → Café

Confiança $\cong P(\text{Café}|\text{Chá}) = 15/20 = 0.75$

Confiança > 50%, significa que as pessoas que bebem chá são mais propensos a beber café do que não beber café

Então a regra parece razoável

Limitações de confiança

	Café	<u>Café</u>	
Chá	15	5	20
<u>Chá</u>	75	5	80
	90	10	100

Regra de Associação: Chá → Café

Confiança = $P(\text{Café}|\text{Chá}) = 15/20 = 0.75$

mas $P(\text{Café}) = 0.9$, que significa que a pessoa que está bebendo chá reduz a probabilidade de beber café!

⇒ Anote que $P(\text{Café}|\overline{\text{Chá}}) = 75/80 = 0.9375$

Medidas para Regras de Associação

- Então, que tipo de regras nós realmente queremos?
 - Confiança($X \rightarrow Y$) tem que ser alta suficiente
 - Para garantir que as pessoas que compram X também tendem a comprar Y
 - Confiança($X \rightarrow Y$) > suporte(Y)
 - Caso contrário, a regra será enganosa porque possuir o item X realmente reduz a chance de possuir o item Y na mesma transação
 - Existe alguma medida que captura essa restrição?
 - Resposta: **Sim**. Há muitas delas.

Independência Estatística

- O critério
confiança($X \rightarrow Y$) = suporte(Y)

é equivalente a:

- $P(Y|X) = P(Y)$
- $P(X,Y) = P(X) \times P(Y)$

Se $P(X,Y) > P(X) \times P(Y)$: X & Y são positivamente correlacionados

Se $P(X,Y) < P(X) \times P(Y)$: X & Y são negativamente correlacionados

Medidas que levam em conta a dependência estatística

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

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

Lift é usado para regras,
enquanto o interesse é
usado para conjuntos de
itens

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

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

Exemplo: Lift/Interesse

	Café	<u>Café</u>	
Chá	15	5	20
<u>Chá</u>	75	5	80
	90	10	100

Regra de Associação: Chá → Café

Confiança = $P(\text{Café}|\text{Chá}) = 0.75$

mas $P(\text{Café}) = 0.9$

⇒ Lift = $0.75/0.9 = 0.8333$ (< 1 , portanto, negativamente associado)

Assim, é bastante usar a confiança/lift para a eliminação?

Muitas medidas foram propostas, qual é a melhor?

#	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)	$\frac{\sum_j \max_k P(A_j, B_k) + \sum_k \max_j P(A_j, B_k) - \max_j P(A_j) - \max_k P(B_k)}{2 - \max_j P(A_j) - \max_k P(B_k)}$
3	Odds ratio (α)	$\frac{P(A,B)P(\bar{A},\bar{B})}{P(A,\bar{B})P(\bar{A},B)}$
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)	$\max \left(P(A, B) \log \left(\frac{P(B A)}{P(B)} \right) + P(\bar{A}\bar{B}) \log \left(\frac{P(\bar{B} \bar{A})}{P(\bar{B})} \right), \right. \\ \left. P(A, B) \log \left(\frac{P(A B)}{P(A)} \right) + P(\bar{A}\bar{B}) \log \left(\frac{P(\bar{A} \bar{B})}{P(\bar{A})} \right) \right)$
9	Gini index (G)	$\max \left(P(A)[P(B A)^2 + P(\bar{B} A)^2] + P(\bar{A})[P(B \bar{A})^2 + P(\bar{B} \bar{A})^2] \right. \\ \left. - P(B)^2 - P(\bar{B})^2, \right. \\ \left. P(B)[P(A B)^2 + P(\bar{A} B)^2] + P(\bar{B})[P(A \bar{B})^2 + P(\bar{A} \bar{B})^2] \right. \\ \left. - P(A)^2 - P(\bar{A})^2 \right)$
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)	$\max \left(\frac{P(A)P(\bar{B})}{P(\bar{A}B)}, \frac{P(B)P(\bar{A})}{P(\bar{B}A)} \right)$
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)	$\frac{P(A,B) + P(\bar{A}\bar{B})}{P(A)P(B) + P(\bar{A})P(\bar{B})} \times \frac{1 - P(A)P(B) - P(\bar{A})P(\bar{B})}{1 - P(A,B) - P(\bar{A}\bar{B})}$
20	Jaccard (ζ)	$\frac{P(A,B)}{P(A) + P(B) - P(A,B)}$
21	Klogsen (K)	$\sqrt{P(\bar{A}, \bar{B})} \max(P(B A) - P(B), P(A B) - P(A))$