

REGRAS DE ASSOCIAÇÃO

Cristiane Neri Nobre

Mineração de Regras de Associação

- Regras de associação são parecidas com regras de classificação (ou árvores de decisão).
 - Elas relacionam atributos que “co-ocorrem” em conjuntos de itens (*itemsets*)
- Podem ser encontradas da mesma forma que regras de classificação.
- Entretanto,
 - Qualquer atributo pode ocorrer no lado direito de uma regra
 - As regras de associação têm a forma de regras <se *antecedente* então *consequente*>, em que antecedente e consequente são *itemsets*.
 - Por exemplo, se o cliente compra **pão** então também compra **manteiga**.
 - Estas regras são calculadas a partir dos dados e são de natureza probabilística.

Mineração de Regras de Associação

- O número de regras de associação é muitíssimo maior que o de regras de classificação!
- O grau de incerteza de uma regra é dado pela **confiança** da regra
- Algumas das conclusões que se podem tirar são do tipo:
 - *20% das pessoas que compram café também compram bolachas.*

Esta informação, quando conhecida, poderá resultar num conjunto de ações que vão desde a promoção conjunta de artigos até alteração da sua localização no supermercado.

Mineração de Regras de Associação

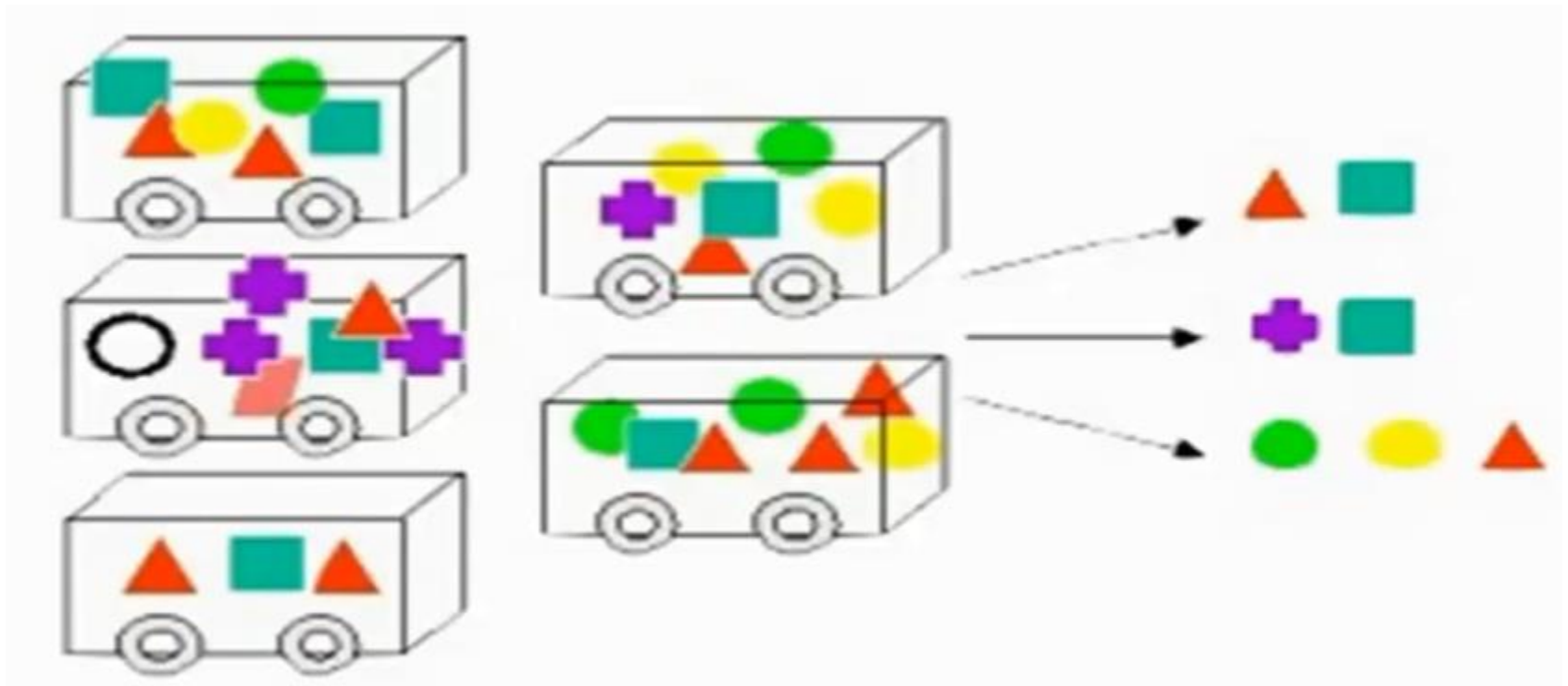
Exemplo 1: Prateleiras de mercado

- Descoberta de padrões entre fralda e cerveja nos finais de semana

parte significativa das compras de homens, às sextas-feiras à noite, que inclui fraldas, inclui também cerveja.
$$\{\text{fralda}\} \Rightarrow \{\text{cerveja}\}$$
- o cliente que compra pão e manteiga, 80% das vezes compra leite.
$$\{\text{pão, manteiga}\} \Rightarrow \{\text{leite}\}$$

Mineração de Regras de Associação

Exemplo 2:



Conceitos relacionados à criação de regras de associação

- **Item**

- É um par [atributo-valor].

- **ItemSets**

- São combinações de pares [atributo-valor] com um valor mínimo de **suporte**.

Mas o que significa o conceito de suporte?

- Na verdade para geração e avaliação da qualidade de regras de associação, temos muitas métricas de avaliação:
- **Alguns exemplos:**
 - Suporte
 - Confiança
 - Lift

Mas o que significa o conceito de suporte?

- **Cobertura (ou suporte) de uma regra**

Número de transações que contem todos os itens da transação dividido pelo número de transações

- **Exemplo de suporte**

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

1 item	2 itens	3 itens
{a}:7	{a,c}:4	{a,c,d}:3
{b}:3	{a,d}:5	{a,c,e}:3
{c}:7	{a,e}:6	{a,d,e}:4
{d}:6	{b,c}:3	
{e}:7	{c,d}:4	
	{c,e}:4	
	{d,e}:4	

O significa o conceito de Confiança?

- **Acurácia (confiança) de uma regra ($A \rightarrow B$)**

Indica a proporção de vezes que, em uma transação contendo o elemento A, também tem B.

$$\text{Confiança } (A \rightarrow B) = \frac{P(A \cup B)}{P(A)} = \frac{\text{suporte}(A \cup B)}{\text{suporte}(A)}$$

O significa o conceito de Confiança?

Qual a confiança da regra:

$$\text{Confiança } (A \rightarrow B) = \frac{P(A \cup B)}{P(A)} = \frac{\text{suporte}(A \cup B)}{\text{suporte}(A)}$$

Quem levou o produto A também levou o produto D ($A \rightarrow D$) ?

$$\text{Confiança da Regra} = \frac{5}{7} = 0.71$$

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

1 item	2 itens	3 itens
{a}:7	{a,c}:4	{a,c,d}:3
{b}:3	{a,d}:5	{a,c,e}:3
{c}:7	{a,e}:6	{a,d,e}:4
{d}:6	{b,c}:3	
{e}:7	{c,d}:4	
	{c,e}:4	
	{d,e}:4	

O significa o conceito de Lift?

- **Lift (coeficiente de interesse) de uma regra ($A \rightarrow B$)**

O quanto mais frequente torna-se B quando A ocorre

$$Lift = \frac{confian\c{a} (A \rightarrow B)}{suporte(B)}$$

O significa o conceito de Lift?

Qual a lift da regra:

$$Lift = \frac{\text{confiança}(A \rightarrow B)}{\text{suporte}(B)}$$

Quem levou o produto A também levou o produto D ($A \rightarrow D$) ?

$$\text{Confiança da Regra } (A \rightarrow D) = \frac{5}{7} \quad \text{Lift da Regra } (A \rightarrow D) = \frac{5/7}{6}$$

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

1 item	2 itens	3 itens
{a}:7	{a,c}:4	{a,c,d}:3
{b}:3	{a,d}:5	{a,c,e}:3
{c}:7	{a,e}:6	{a,d,e}:4
{d}:6	{b,c}:3	
{e}:7	{c,d}:4	
	{c,e}:4	
	{d,e}:4	

Investiguem outras métricas de avaliação

Veja este artigo que compara 21 métricas!

<http://www.cse.msu.edu/~ptan/papers/IS.pdf>

Veja este também (discute sobre medidas objetivas e subjetivas)

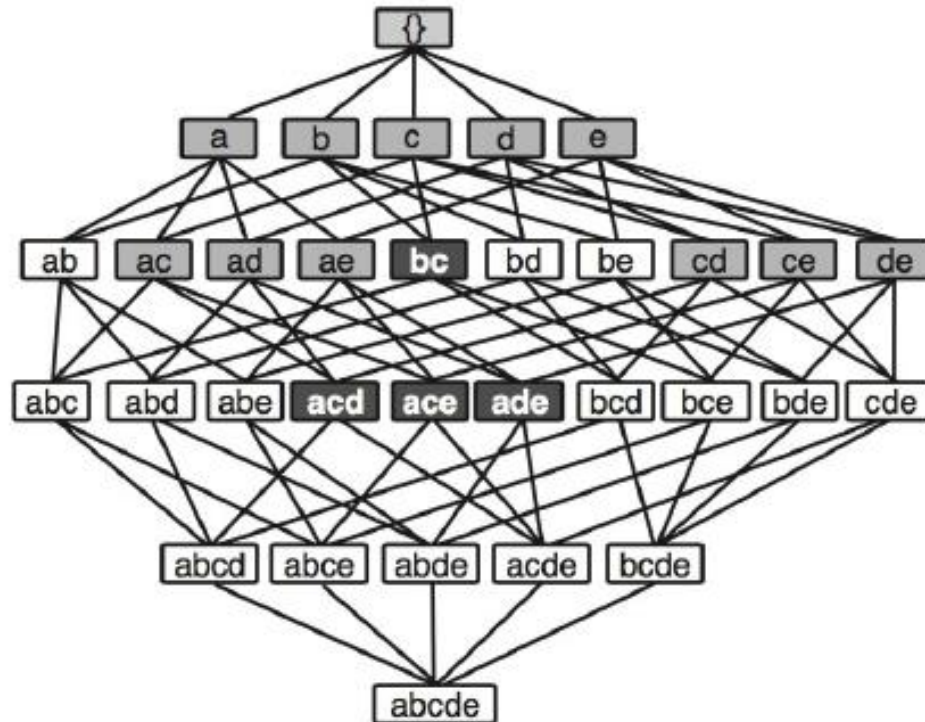
www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/download/79/64/0

Espaço de busca para problemas de associação

O espaço de busca de todos os possíveis conjuntos de itens para um conjunto de itens A contém exatamente $2^{|A|}$ *ItemSets* diferentes.

Ou $2^{|A|}-1$ itens sem incluir o conjunto vazio.

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



Espaço de busca para problemas de associação

A principal propriedade explorada pela maioria dos algoritmos de mineração de conjuntos de itens frequentes é que o suporte é monotonicamente decrescente com relação ao número de itens de um *itemset*.

Ou seja, **o conjunto de suporte de um conjunto de itens diminui sempre que se acrescenta um novo item.**

Formalmente, considere X e Y dois conjuntos de itens em um banco de dados de transações T sobre I , assim $X, Y \subseteq I$.

É fácil mostrar que $X \subseteq Y \Rightarrow \text{suporte}(Y) \leq \text{suporte}(X)$.

Espaço de busca para problemas de associação

Essa regra é uma consequência imediata de que o conjunto de suporte de X está incluído no conjunto de suporte de Y .

Então se um *itemset* é pouco frequente, todos os seus superconjuntos devem ser pouco frequentes

- Adicionalmente, **a propriedade todos os subconjuntos de um conjunto de itens frequentes são frequentes** é válida
- Essa é a propriedade da monotonicidade do suporte
- Uma função diz-se monotônica em x , se $x_1 < x_2$ implica que $f(x_1) < f(x_2)$

Neste vídeo, aprendemos:

1. O que são regras de associação
2. Quais as métricas utilizadas para avaliação das regras

No próximo vídeo, veremos o funcionamento de um algoritmo de associação chamado **APriori**