



REGRAS DE ASSOCIAÇÃO

Daricélio Moreira Soares

Regras de Associação (Transacionais)

Uma regra de associação representa um padrão de relacionamento entre itens de dados de um domínio de aplicação que ocorre com uma determinada frequência na base de dados.

$$X \Rightarrow Y$$

$\{\text{candidíase}\} \Rightarrow \{\text{pneumonia}\}$

$\{\text{café, leite}\} \Rightarrow \{\text{pão, manteiga, queijo}\}$

A primeira regra indica, com um determinado grau de certeza, que se o paciente contraiu candidíase, então também teve pneumonia.

Regras de Associação

Regras de associação possuem índices que indicam sua relevância e a validade.

O fator de suporte de uma regra $X \Rightarrow Y$ é definido pela porcentagem de transações que incluem todos os itens do conjunto $X \cup Y$.

→ Representa a fração das transações que satisfazem tanto o antecedente quanto o conseqüente da regra.
O suporte de uma regra tenta indicar sua relevância.

Regras de Associação

Seja R a regra $X \Rightarrow Y$.

Seja T o número de transações consideradas.

Seja $T_{X \cup Y}$ o número de transações que incluem os elementos de $X \cup Y$.

$$\text{Suporte}(R) = T_{X \cup Y} / T$$

<u>TID</u>	<u>Itens Comprados</u>	
101	leite, pão, suco	Suporte($\{\text{leite}\} \Rightarrow \{\text{suco}\}$) = $2 / 4 = 50\%$
792	leite, suco	Suporte($\{\text{suco}\} \Rightarrow \{\text{leite}\}$) = 50%
1130	leite, ovos	Suporte($\{\text{pão}\} \Rightarrow \{\text{suco}\}$) = ____
1735	pão, biscoito, café	Suporte($\{\text{pão}\} \Rightarrow \{\text{ovos}\}$) = ____
		Suporte($\{\text{pão, café}\} \Rightarrow \{\text{biscoito}\}$) = ____

Regras de Associação

O fator de confiança de uma regra $X \Rightarrow Y$ é definido pela porcentagem de transações que incluem os itens X e Y em relação a todas que incluem os itens de X.

→ Representa o grau de satisfatibilidade do conseqüente, em relação às transações que incluem o antecedente.
A confiança tenta indicar a validade da regra.

Regras de Associação

Seja R a regra $X \Rightarrow Y$.

Seja T_X o número de transações que incluem os elementos de X.

Seja $T_{X \cup Y}$ o número de transações que incluem os elementos de $X \cup Y$.

$$\text{Confiança}(R) = T_{X \cup Y} / T_X$$

<u>Id-T.</u>	<u>Itens Comprados</u>	
101	leite, pão, suco	$\text{Confiança}(\{\text{leite}\} \Rightarrow \{\text{suco}\}) = 2 / 3 = 67\%$
792	leite, suco	$\text{Confiança}(\{\text{suco}\} \Rightarrow \{\text{leite}\}) = 2 / 2 = 100\%$
1130	leite, ovos	$\text{Confiança}(\{\text{pão}\} \Rightarrow \{\text{suco}\}) = \underline{\hspace{2cm}}$
1735	pão, biscoito, café	$\text{Confiança}(\{\text{pão}\} \Rightarrow \{\text{ovos}\}) = \underline{\hspace{2cm}}$
		$\text{Confiança}(\{\text{pão, café}\} \Rightarrow \{\text{biscoito}\}) = \underline{\hspace{2cm}}$

Mineração de Regras de Associação

→ Entrada:

- Base de dados de transações;
- Suporte mínimo;
- Confiança mínima.

→ Saída:

- Todas as regras de associação que possuem suporte e confiança maiores ou iguais ao suporte e à confiança mínimos.

Exemplos de Regras Mineradas

■ Entrada:

- Base de Dados: POF-FGV (registro das compras mensais de 422 famílias cariocas em Junho de 1998)
- SupMin = 3%, ConfMin = 60%

■ Resultados:

- Foram mineradas **8.469** regras de associação envolvendo 2 itens!
- Exemplos:
 - {Strogonoff de Frango (caixa)} \Rightarrow {Lasanha (caixa)}
 - {Milho Verde em Conserva} \Rightarrow {Ervilhas em Conserva}
 - {Kiwi} \Rightarrow {Feijão Preto}
 - {Cenoura} \Rightarrow {Batata Inglesa}

Regras de Associação Quantitativas

Exemplo (base de dados sobre a AIDS):

$(\text{sexo}=\text{"M"}) \wedge (20 \leq \text{idade} \leq 30) \wedge (\text{opção-sex}=\text{"heterossex"}) \Rightarrow (\text{usuário-drogas}=\text{"S"})$

→ Esta regra indica, com confiança C,
que pacientes aidéticos heterossexuais, entre 20 e 30 anos,
do sexo masculino têm C% de chance de serem usuários de drogas.

Outra Notação (com Predicados):

$\text{sexo}(X, \text{"M"}) \wedge \text{idade}(X, [20, 30]) \wedge \text{opção-sex}(X, \text{"heterossex"}) \Rightarrow \text{usuário-drogas}(X, \text{"S"})$

Regras de Associação Quantitativas

- **Regras Multidimensionais**

- Base do IBGE

$(\text{Escolaridade} = \text{"Analfabeto"}) \wedge (\text{Sexo} = \text{"F"}) \wedge (\text{Idade} > 50) \Rightarrow$
 $(\text{Ocupação} = \text{"Doméstica"})$

- **Regras Híbridas**

- Base da FGV

$(\text{Cidade} = \text{"POA"}) \wedge (\text{Membros} = 1) \wedge (\text{Produto} = \text{"cerveja"}) \Rightarrow$
 $(\text{Produto} = \text{"salaminho"})$

Mineração de Regras Quantitativas

→ Entrada:

- Base de dados relacional;
- Suporte mínimo;
- Confiança mínima.

→ Saída:

- Todas as regras quantitativas que possuem suporte e confiança maiores ou iguais ao suporte e à confiança mínimos.