

Universidade Federal do Acre  
Programa de Pós-Graduação em Ciência da Computação - PPgCC

## Mineração de Dados



# Regras de Associação

---

Prof. Dr. Daricélio Soares

# Onde estamos?

- Tarefas Descritivas

- Padrões e regras descrevem características importantes dos dados com os quais se está trabalhando.

- Mineração de Dados Indireta

- Através de uma técnica de mineração, extraem-se padrões significativos que serão posteriormente avaliados.
- O resultado da mineração complementa o conhecimento do especialista e deverá ser examinado e avaliado por este.

# 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.
- Seja  $I = \{i_1, i_2, \dots, i_n\}$  o conjunto de itens do domínio da aplicação.
- Uma regra de associação  $R$  definida sobre  $I$  é uma implicação da forma

$$X \rightarrow Y$$

- onde  $X \subset I$ ,  $Y \subset I$ ,  $X \neq \emptyset$ ,  $Y \neq \emptyset$  e  $X \cap Y = \emptyset$ .  $X$  é o antecedente da regra e  $Y$  é o conseqüente.

# Regras de Associação (Transacionais)

$$\mathbf{X} \rightarrow \mathbf{Y}$$

$$\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \dots \wedge \mathbf{X}_n \Rightarrow \mathbf{Y}_1 \wedge \mathbf{Y}_2 \wedge \dots \wedge \mathbf{Y}_m$$

$$\{\text{candidíase}\} \rightarrow \{\text{pneumonia}\}$$

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

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

# Medidas de Interesse

- Seja  $R$  a regra  $X \rightarrow Y$ .
- Seja  $T$  o número de transações consideradas.
- Seja  $T(XUY)$  o número de transações que incluem os elementos de  $X \cup Y$ .

$$\text{Suporte}(R) = T(XUY) / T$$

# Medidas de Interesse

- 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  $X \cup Y$ .

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

# Mineração de Regras de Associação

## Input

Base de dados de transações (?);  
Suporte mínimo;  
Confiança mínima.

## Output

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

# Regras de Associação Multidimensionais

- Regras de associação quantitativas são utilizadas quando se deseja minerar padrões em bases de dados relacionais (formadas por atributos quantitativos e atributos categóricos).

Atributos Categóricos		Atributos Quantitativos			
Id	Sexo	Profissão	Salário	Idade	...



# Regras de Associação Multidimensionais

- 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.

# Meta-Regras

## Mineração Baseada em Restrições

- Permitem a especificação do tipo de regras que se deseja minerar.
- Podem funcionar como restrições definidas pelo usuário.
- Podem representar hipóteses a serem confirmadas

### Meta-Regra

$$\text{idade}(X, [30, 39]) \wedge \text{renda}(X, [4K, 6K]) \rightarrow \text{compras}(Y, \text{"educational software"})$$

# Questões?

