

Resumo IA - AB2

Regras de Associação:

- Transacionais
- Quantitativas (Relacionais ou Multidimensionais)
- Meta-Regras

Regras de Associação Transacionais:

- Representa um padrão de relacionamento entre itens de dados de um domínio de aplicação que ocorre com uma determinada frequência na base de dados.

Exemplo:

```
{candidíase} ⇒ {pneumonia}
```

```
{café, leite} ⇒ {pão, manteiga, queijo}
```

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

```
X ⇒ Y
```

X é o antecedente da regra e Y é o consequente

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 quando o consequente da regra.
- O suporte de uma regra tenta indicar sua **relevância**.

Exemplos:

Suporte:

Seja R a regra $X \Rightarrow Y$. Seja R a regra $X \Rightarrow Y$. Seja T o número de transações consideradas. Seja T X U Y o número de transações que incluem os elementos de $X \cup Y$.

$$\text{Suporte}(R) = T(x \cup y) / T$$

TID	Itens Comprados
101	leite, pão, suco
792	leite, suco
1130	leite, ovos
1735	pão, biscoito, café

```

Suporte({leite} ⇒ {suco}) = 2 / 4 = 50%
Suporte({suco} ⇒ {leite}) = 2/4 => 50%
Suporte({pão} ⇒ {suco}) = 1/4 => 25%
Suporte({pão} ⇒ {ovos}) = 0/4 => 0%
Suporte({pão, café} ⇒ {biscoito}) = 1/4 => 25%

```

Seja X um conjunto de itens.

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

TID	Itens Comprados
101	leite, pão, suco
792	leite, suco
1130	leite, ovos
1735	pão, biscoito, café

$$\text{Suporte}(X) = T_x / T$$

```

Suporte({leite}) = 3/4 = 75%
Suporte({leite, suco}) = 2 / 4 = 50%
Suporte({pão, suco}) = 1/4 => 25%
Suporte({pão, ovos}) = 0/4 => 0%
Suporte({pão, café, biscoito}) = 0/4 => 25%

```

Confiança:

- O **fator de confiança** de uma regra $X \Rightarrow Y$ é definido pela porcentagem de transações que incluem os itens de Y dentre aquelas que incluem os itens de X;
 - Representa o grau de satisfatibilidade do consequente, em relação às transações que incluem o antecedente.
- A confiança tenta indicar a **validade da regra**.

Exemplo:

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 $X \cup Y$.

$$\text{Confiança}(R) = T(x \cup y) / T_x$$

TID	Itens Comprados
101	leite, pão, suco
792	leite, suco
1130	leite, ovos
1735	pão, biscoito, café

```

Confiança({leite} ⇒ {suco}) = 2 / 3 = 67%
Confiança({suco} ⇒ {leite}) = 2 / 2 = 100%
Confiança({pão} ⇒ {suco}) = 1/2 => 50%
Confiança({pão} ⇒ {ovos}) = 0/2 => 0%
Confiança({pão, café} ⇒ {biscoito}) = 1/1 => 100%

```

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.

Exemplo:

- 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)} ⇒ {Lasanha (caixa)}
 - {Milho Verde em Conserva} ⇒ {Ervilhas em Conserva}
 - {Kiwi} ⇒ {Feijão Preto}
 - {Cenoura} ⇒ {Batata Inglesa}
- O suporte indica o grau de relevância que quer se dar para as associações na base.

- Um suporte mínimo baixo pode ser usado para focar na especificidade.
- Um suporte mínimo alto pode ser usado para focar na generalidade.
- A confiança indica grau de validade de aquelas associações:
 - Trabalhar com uma confiança baixa pode mostrar regras novas que podem ou não fazer sentido. Trabalhar com uma confiança alta vai fazer com que a maioria das regras extraídas façam sentido e tenham um grau de validade maior.

Regras de Associação Quantitativas:

Exemplo(base de dados sobre a AIDS)

$(\text{Transmissão-Sexual} = \text{"Não"}) \Rightarrow (\text{Drogas} = \text{"Sim"})$

- Esta regra indice, com certo grau de certeza, que se o contágio não foi de natureza sexual, o paciente utiliza drogas.
- Uma regra de associação quantitativa R definida sobre D é uma implicação da forma:

$X_1 \wedge X_2 \wedge \dots \wedge X_n \Rightarrow Y_1 \wedge Y_2 \wedge \dots \wedge Y_m$

onde $n \geq 1$, $m \geq 1$, X_i ($1 \leq i \leq n$) e Y_j ($1 \leq j \leq m$) são condições definidas sobre atributos distintos de D.

- Os conceitos de suporte e confiança se aplicam como nas regras convencionais.

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