
Mineração de padrões frequentes por meio de 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**.

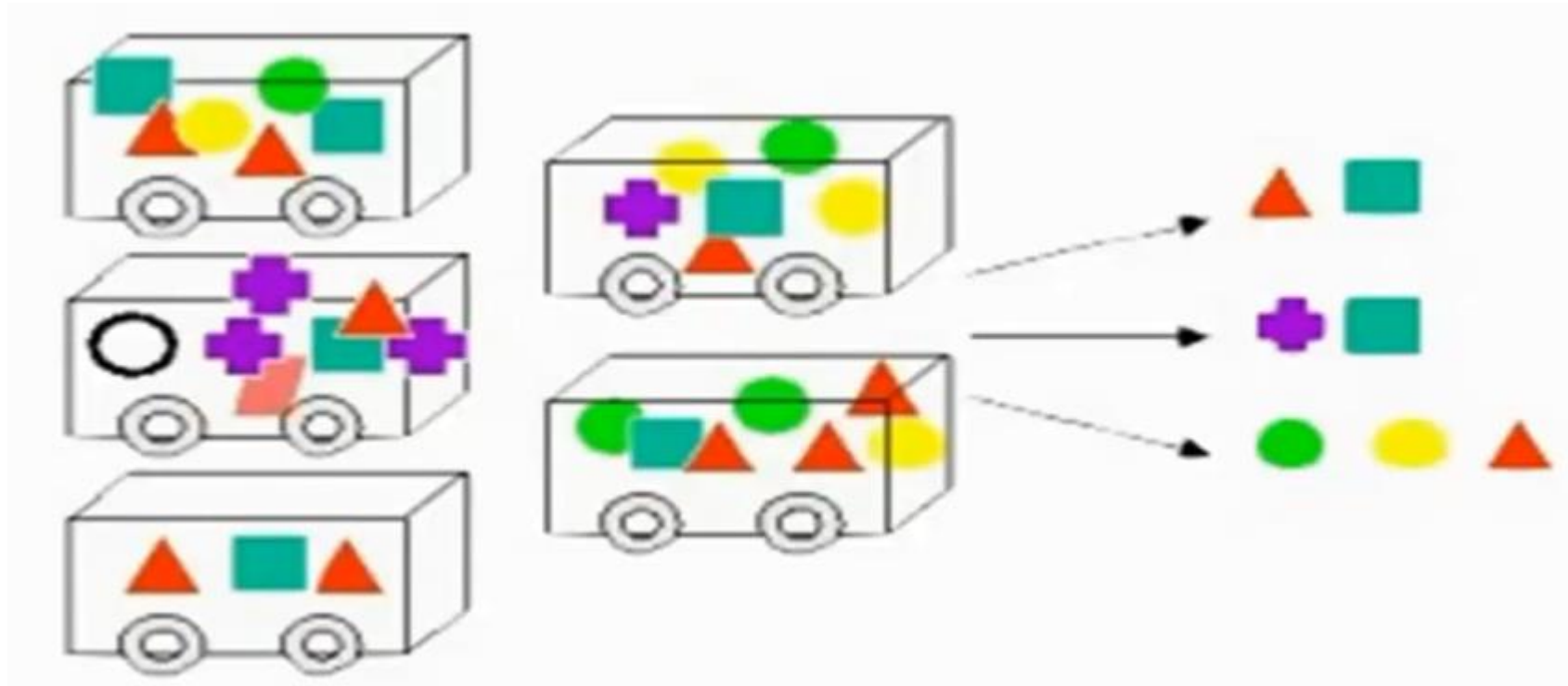
$\{\text{fralda}\} \Rightarrow \{\text{cerveja}\}$

- 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:



Avaliação da qualidade do modelo

➤ Cobertura (*suporte*) de uma regra

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

➤ Acurácia (*confiança*) de uma regra

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

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

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

O quanto mais frequente torna-se B quando A ocorre

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

Avaliação da qualidade do modelo

➤ **Item**

- É um par [atributo-valor].

➤ **ItemSets**

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

Mineração de Regras de Associação

- Seja o exemplo:

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}

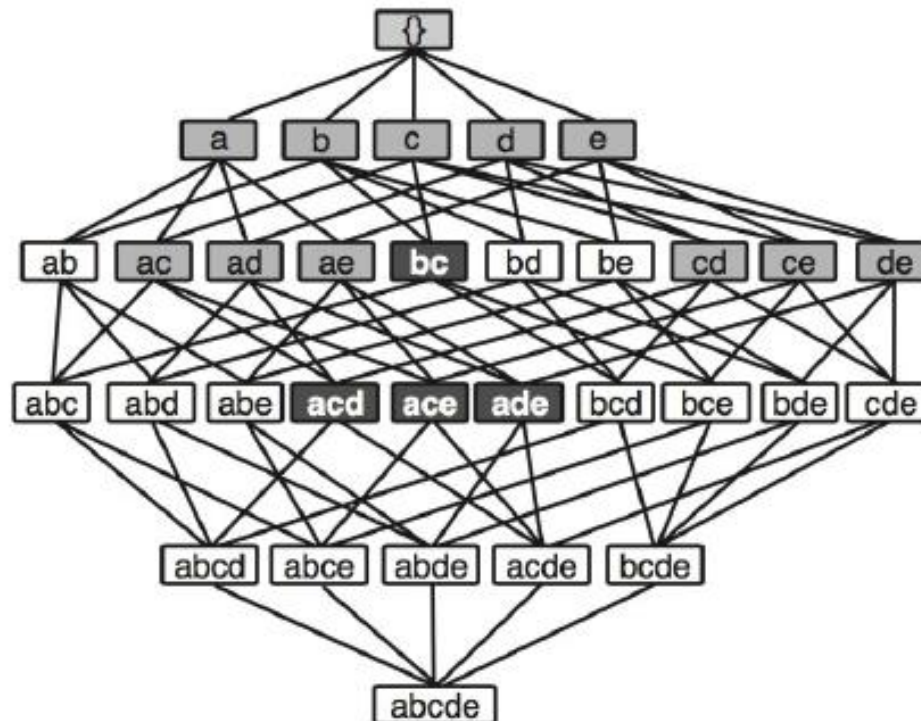
0 Items	1 item	2 itens	3 itens
Φ : 10	{a}:7	{a,c}:4	{a,c,d}:3
	{b}:3	{a,d}:5	{a,c,e}:3
	{c}:7	{a,e}:6	{a,d,e}:4
	{d}:6	{b,c}:3	
	{e}:7	{c,d}:4	
		{c,e}:4	
		{d,e}:4	

Mineração de Regras de Associação

O espaço de busca

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}



Mineração de Regras de Associação

Como funciona o algoritmo de extração de regras de associação:

Algoritmo *Apriori* (Agrawal *et al* 1994)?

Mineração de Regras de Associação

➤ Duas fases:

1. Gerar os *item sets* com o **suporte** mínimo especificado;
2. Para cada *itemset*, gerar as regras com a **confiança** mínima especificada.

Mineração de Regras de Associação

Exemplo 1: Prateleiras de mercado

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Mineração de Regras de Associação

Vamos calcular o **suporte** dos itemSets?

Lembrando que **suporte** de uma regra é o número de transações que contem todos os itens da transação dividido pelo número total de transações

Mineração de Regras de Associação

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Produto	Suporte
Leite	$2/10 = 0.2$
Café	$3/10 = 0.3$
Cerveja	$2/10 = 0.2$
Pão	$5/10 = 0.5$
Manteiga	$5/10 = 0.5$
Arroz	$2/10 = 0.2$
Feijão	$2/10 = 0.2$

Mineração de Regras de Associação

Considerando o suporte ≥ 3

Como serão os valores para os ItemSets?

Mineração de Regras de Associação

ItemSet = 1?

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Produto	Suporte
Leite	$2/10 = 0.2$
Café	$3/10 = 0.3$
Cerveja	$2/10 = 0.2$
Pão	$5/10 = 0.5$
Manteiga	$5/10 = 0.5$
Arroz	$2/10 = 0.2$
Feijão	$2/10 = 0.2$

Mineração de Regras de Associação

ItemSet = 2?

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Produto	Suporte
Café e pão	$3/10 = 0.3$
Café e manteiga	$3/10 = 0.3$
Pão e Manteiga	$4/10 = 0.4$

Mineração de Regras de Associação

ItemSet =3?

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Produto	Suporte
Café, pão e manteiga	$3/10 = 0.3$

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Segunda fase

- Para cada *itemSet* faça:
 - ✓ Gere todas as regras possíveis
 - ✓ Mantenha somente as regras com a **confiança** mínima exigida

Mineração de Regras de Associação

Como calcular a confiança?

➤ **Confiança** de uma regra

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

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

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 2**

Produto	Suporte
Café e pão	$3/10 = 0.3$
Café e manteiga	$3/10 = 0.3$
Pão e Manteiga	$4/10 = 0.4$

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

{Café, pão}

Se Café → Pão

Se Pão → Café

{Café, manteiga}

Se Café → manteiga

Se Manteiga → Café

{Pão, manteiga}

Se Pão → manteiga

Se Manteiga → Pão

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 2**

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Qual a confiança das regras?

{Café, pão}

Se Café → Pão

confiança = $3/3 = 1$

Se Pão → Café

confiança = $3/5 = 0.6$

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 2**

Qual a confiança das regras?

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

{Café, manteiga}

Se Café → manteiga

confiança = $3/3 = 1$

Se Manteiga → Café

confiança = $3/5 = 0.6$

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 2**

Qual a confiança das regras?

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

{Pão, manteiga}

Se Pão → manteiga

confiança = $4/5 = 0.8$

Se Manteiga → Pão

confiança = $4/5 = 0.8$

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 2**

Se assumirmos **confiança** ≥ 0.8 ,
teremos as seguintes regras:

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Se Café \rightarrow Pão

confiança = $3/3 = 1$

Se Café \rightarrow manteiga

confiança = $3/3 = 1$

Se Pão \rightarrow manteiga

confiança = $4/5 = 0.8$

Se Manteiga \rightarrow Pão

confiança = $4/5 = 0.8$

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 3**

Produto	Suporte
Café, pão e manteiga	$3/10 = 0.3$

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

{Café, pão, manteiga}

Se Café e pão → manteiga

Se Café e manteiga → pão

Se pão e manteiga → café

Se café → pão e manteiga

Se pão → café e manteiga

Se manteiga → café e pão

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 3**

Produto	Suporte
Café, pão e manteiga	$3/10 = 0.3$

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

{Café, pão, manteiga}

Se Café e pão → manteiga

confiança = $3/3 = 1$

Se Café e manteiga → pão

confiança = $3/3 = 1$

Se pão e manteiga → café

confiança = $3/4 = 0.75$

Mineração de Regras de Associação

Como gerar as regras a partir dos itemSets?

Com o **ItemSets = 3**

Qual a confiança das regras?

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

{Café, pão, manteiga}

Se café → pão e manteiga

$$\text{confiança} = 3/3 = 1$$

Se pão → café e manteiga

$$\text{confiança} = 3/5 = 0.6$$

Se manteiga → café e pão

$$\text{confiança} = 3/5 = 0.6$$

Mineração de Regras de Associação

Quais as regras iremos selecionar, considerando a **confiança** ≥ 0.8

Num	Regras	Confiança
1	Se Café \rightarrow Pão	$3/3 = 1$
2	Se Café \rightarrow manteiga	$3/3 = 1$
3	Pão \rightarrow manteiga	$4/5 = 0.8$
4	Manteiga \rightarrow Pão	$4/5 = 0.8$
5	Se Café e pão \rightarrow manteiga	$3/3 = 1$
6	Se Café e manteiga \rightarrow pão	$3/3 = 1$
7	Se café \rightarrow pão e manteiga	$3/3 = 1$

Mineração de Regras de Associação

Como avaliar as melhores regras a partir destas 7 regras?

Iremos calcular o lift da regra:

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

Mineração de Regras de Associação

Iremos calcular o lift da regra:

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

Regras	Confiança	Lift
Se Café → Pão	3/3 = 1	1/0.5=2
Se Café → manteiga	3/3 = 1	1/0.5 = 2
Pão → manteiga	4/5 = 0.8	0.8/0.5 = 1.6
Manteiga → Pão	4/5 = 0.8	0.8/0.5=1.6
Se Café e pão → manteiga	3/3 = 1	1/0.5= 2
Se Café e manteiga → pão	3/3 = 1	1/0.5=2
Se café → pão e manteiga	3/3 = 1	1/0.4=2.5

Produto	Suporte
Leite	2/10 = 0.2
Café	3/10 = 0.3
Cerveja	2/10 = 0.2
Pão	5/10 = 0.5
Manteiga	5/10 = 0.5
Arroz	2/10 = 0.2
Feijão	2/10 = 0.2

Produto	Suporte
Café e pão	3/10 = 0.3
Café e manteiga	3/10 = 0.3
Pão e Manteiga	4/10 = 0.4

Mineração de Regras de Associação

Qual o significado do lift?

Regras	Confiança	Lift	Significado
Se Café → Pão	$3/3 = 1$	$1/0.5=2$	Clientes que compram café tem duas vezes mais chances de comprarem pão

Valor de **lift** igual a 1 para a regra ($A \rightarrow B$), indica que A e B são independentes

Um valor de **lift superior a 1** indicam que A e B aparecem mais frequentemente juntos do que esperado. Isso significa que a ocorrência de A tem um efeito positivo sobre a ocorrência de B.

Um valor de **lift menor do que 1** indica que A e B aparecem com menos frequência do que o esperado em conjunto, indicando que a ocorrência de A tem um efeito negativo sobre a ocorrência de B

Mineração de Regras de Associação

Depois é só ordenar as regras pelo lift! De forma decrescente!

Investiguem outras métricas!

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

Investiguem outros algoritmos de regras de associação:

eclat, fp-grow!

Mineração de Regras 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 $X \subseteq Y \Rightarrow \text{suporte}(Y) \leq \text{suporte}(X)$.

Mineração de Regras 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 $f(x_1) < f(x_2)$

Outro exemplo

➤ Veja um exemplo com uma base do WEKA

Abra o arquivo **Supermarket**, uma base de dados real da Nova Zelândia

Esta base de dados relaciona produtos na **semana**, ou seja, 7 dias.

Faça a análise das regras...

Se você desejar encontrar padrões que se repetem 10 vezes por dia, qual seria o suporte sugerido?

Para esta situação, faça:

$$10(10 \text{ vezes}) * 7(7 \text{ dias}) = 70 / \text{número de instâncias} = 70 / 4627 = 0.015$$

Este é o suporte sugerido...

Mineração de Regras de Associação

- O problema principal dos algoritmos de regras de associação é que o especialista precisa dar valor às regras geradas
- Para isso, pode ser necessário jogar atributos fora, aqueles que aparecem com muitas frequência, por exemplo.

Referências

- Capítulo 10 do livro
- Katti Faceli. Inteligência Artificial, Uma abordagem de Aprendizado de Máquina, LTC, 2015.