

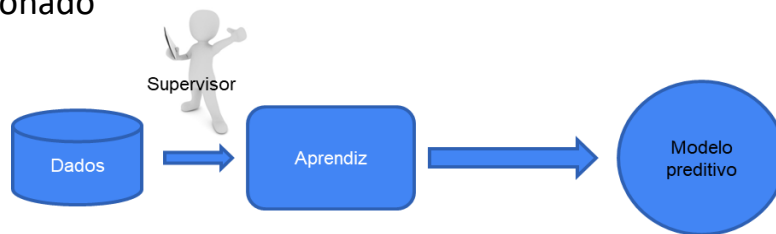
Extração de Regras de Associação

- Inteligência Artificial
- Heloisa de Arruda Camargo

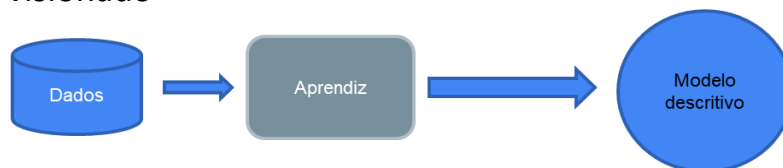
1

AM não supervisionado - contexto

- AM supervisionado



- AM não supervisionado



IA-2022

2

AM não supervisionado - contexto

• AM supervisionado

Exemplo	Tempo	Temperatura	Umidade	Vento	Classe
1	sol	2	72	forte	sim
2	sol	28	91	forte	não
3	sol	22	70	fraco	sim
4	sol	23	95	fraco	não
5	sol	30	85	fraco	não
6	nublado	23	90	forte	sim
7	nublado	29	78	fraco	sim
8	nublado	19	65	forte	não
9	nublado	26	75	fraco	sim
10	nublado	20	87	forte	sim
11	chuva	22	95	fraco	sim
12	chuva	19	70	forte	não
13	chuva	23	80	forte	não
14	chuva	25	81	fraco	sim
15	chuva	21	80	fraco	sim

IA-2022

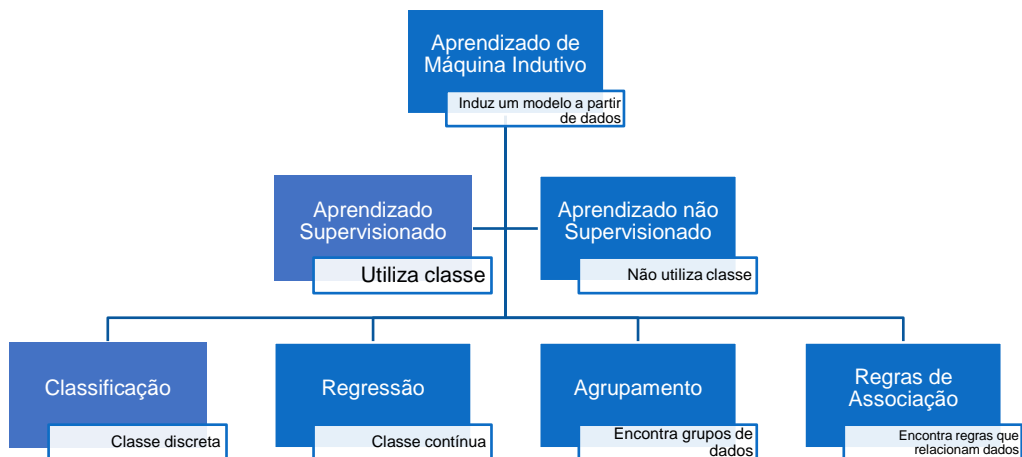
• AM não supervisionado

Exemplo	Tempo	Temperatura	Umidade	Vento	Classe
1	sol	2	72	forte	sim
2	sol	28	91	forte	não
3	sol	22	70	fraco	sim
4	sol	23	95	fraco	não
5	sol	30	85	fraco	não
6	nublado	23	90	forte	sim
7	nublado	29	78	fraco	sim
8	nublado	19	65	forte	não
9	nublado	26	75	fraco	sim
10	nublado	20	87	forte	sim
11	chuva	22	95	fraco	sim
12	chuva	19	70	forte	não
13	chuva	23	80	forte	não
14	chuva	25	81	fraco	sim
15	chuva	21	80	fraco	sim

3

3

Hierarquia do AM indutivo



Indução - forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto particular de exemplos.

IA-2022

4

AM não supervisionado

Resumo – principais características

Os exemplos são definidos por um **conjunto de atributos** SEM atributo-meta

Os exemplos são **não-rotulados**

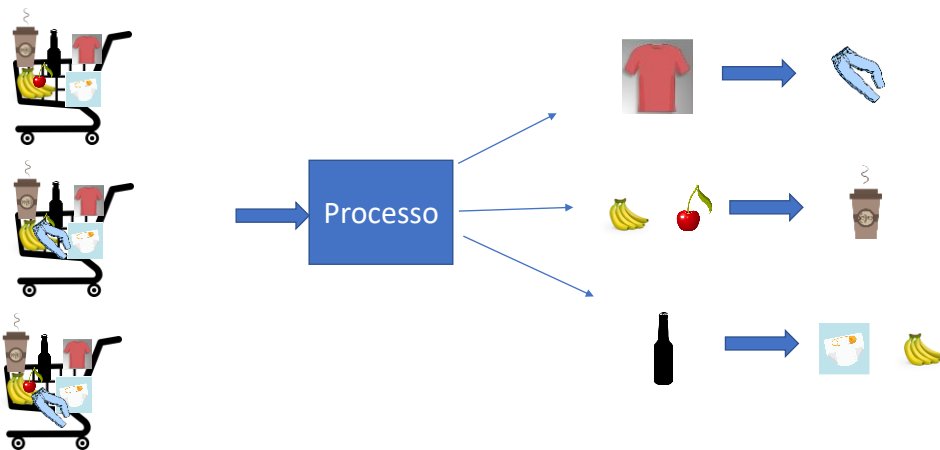
Descreve, de forma concisa, **estruturas** ou **relacionamentos** encontrados nos dados

Consiste na identificação de **comportamentos intrínsecos** do conjunto de dados

5

O que é geração de regras de associação?

- Processo que identifica **associações entre atributos** dos dados, com base na frequência com que ocorrem juntos



IA-2022

6

Regras de associação



- **Objetivo**
 - Encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados
- **Exemplo:**
 - Uma Base de Dados de um supermercado teria como regra o fato de que 80% dos clientes que compram um produto Q, também adquirem, na mesma ocasião o produto W. Em que 80% é o fator de confiabilidade da regra
- **Problema:**
 - Analisar um grande volume de conhecimento extraído no formato de regra

IA-2022

7

Regras de associação - Conceitos importantes

- **Itemset:** Conjunto de atributos ou itens ordenados lexicograficamente
 - Exemplos:
 - `{a, b, c}`
 - `{1, 2, 3}`
 - `{Antonio, Felipe, Marcos}`
 - `{produto1, produto2, produto3}`
- **k-itemset:** Itemset com k elementos
 - Exemplos:
 - 2-itemset: `{a, b}`
 - 3-itemset: `{1, 2, 3}`
- **Transações:** Conjunto de itens (produtos) que representam compras de clientes
 - Exemplos:
 - T1: `{produto1, produto2}`
 - T2: `{produto1, produto2, produto3}`
 - T3: `{produto3}`

IA-2022

8

Regras de associação - Conceitos importantes

- **Itemset frequente:** Ocorre nas transações com uma frequência determinada
- **Suporte do itemset:** Porcentagem de transações onde esse itemset aparece

Itemset:
{açúcar, leite, manteiga, pão}

Quais itemsets têm frequência pelo menos de 50%?

Transações:

T1	leite, pão
T2	açúcar, leite, manteiga
T3	leite, manteiga, pão
T4	leite, manteiga

{leite}	100%
{manteiga}	75%
{pão}	50%
{leite, manteiga}	75%
{leite, pão}	50%

9

Regra de Associação - Definição

- Seja D uma Base de Dados composta por um conjunto de itens

$$A = \{a_1, a_2, \dots, a_m\}$$

ordenados lexicograficamente

- e por um conjunto de transações

$$T = \{t_1, t_2, \dots, t_n\}$$

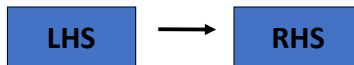
- em que cada transação t_i é composta por um conjunto de itens tal que t_i está contido em A

IA-2022

10

Regra de Associação - Definição

- Uma **regra de associação** é uma implicação na forma



- em que LHS (Left Hand Side) e RHS (Right Hand Side) são conjuntos de itens que estão contidos em A e a **intersecção de LHS e RHS é vazia**

Regras de associação
{leite, pão} → {manteiga}
{pão} → {manteiga, leite}
{leite, pão} → {açúcar, manteiga}

IA-2022

11

Suporte e Confiança da regra

- Suporte de uma regra:
- A regra $LHS \rightarrow RHS$ tem **suporte s** se em s% das transações em D ocorre $LHS \cup RHS$.

Regra: {leite, pão} → {manteiga}

$\text{sup}(\{leite, pão\} \rightarrow \{manteiga\}) = \text{sup}(\{leite, pão, manteiga\}) = 25\%$

Regra: {leite} → {manteiga}

$\text{sup}(\{leite\} \rightarrow \{manteiga\}) = \text{sup}(\{leite, manteiga\}) = 75\%$

IA-2022

12

T1	leite, pão
T2	açúcar, leite, manteiga
T3	leite, manteiga, pão
T4	leite, manteiga

{leite}	100%
{leite, pão}	50%
{leite, manteiga}	75%
{leite, manteiga, pão}	25%
{manteiga, pão}	25%

Suporte e Confiança da regra

- Confiança de uma regra:

A regra $LHS \rightarrow RHS$ ocorre no conjunto de transações T com **confiança c** se c% das transações em T em que ocorre LHS também ocorre RHS.

$$\text{confiança} = \frac{\text{suporte}(\{LHS \cup RHS\})}{\text{suporte}(\{LHS\})}$$

Regra: {leite, pão} \rightarrow {manteiga}

$\text{conf}(\{leite, pão\} \rightarrow \{manteiga\}) = \text{conf}(\{leite, pão, manteiga\})$

$$\text{conf}(\{leite, pão, manteiga\}) = \frac{\text{suporte}(\{leite, pão, manteiga\})}{\text{suporte}(\{leite, pão\})} = \frac{25}{50} = 0,5$$

IA-2022

13

T1	leite, pão
T2	açúcar, leite, manteiga
T3	leite, manteiga, pão
T4	leite, manteiga

{leite}	100%
{leite, pão}	50%
{leite, manteiga}	75%
{leite, manteiga, pão}	25%
{manteiga, pão}	25%

Extraindo regras de associação

$A = \{\text{bermuda, calça, camiseta, sandália, tênis}\}$

$T = \{1, 2, 3, 4\}$

Encontrar regras com:

Suporte Mínimo = 50% (2 transações).

Confiança Mínima = 50%.

(suporte e confiança definidos pelo usuários)

Transações	Itens Comprados
1	calça, camiseta, tênis
2	camiseta, tênis
3	bermuda, tênis
4	calça, sandália

Itemsets Frequentes	Suporte
{tênis}	75%
{calça}	50%
{camiseta}	50%
{camiseta, tênis}	50%

IA-2022

14

Extraíndo regras de associação

- Suporte Mínimo = 50% (2 transações)
- Confiança Mínima = 50%

tênis → camiseta

Itemsets	Suporte
Frequentes	
{tênis}	75%
{calça}	50%
{camiseta}	50%
{camiseta, tênis}	50%

suporte = suporte({tênis, camiseta}) = 50%

$$\text{confiança} = \frac{\text{suporte}(\{\text{tênis, camiseta}\})}{\text{suporte}(\{\text{tênis}\})} = \frac{50}{75} = 66,6\%$$

IA-2022

15

Geração de Regras de Associação

Esquema básico dos algoritmos:

Dada uma Base de Dados D composta por um **conjunto de itens** $A = \{a_1, a_2, \dots, a_m\}$ ordenados lexicograficamente e por um **conjunto de transações** $T = \{t_1, t_2, \dots, t_n\}$, em que, cada transação t_i é composta por um conjunto de itens tal que t_i está contido em A

- Encontrar todos os itemsets que possuem suporte maior que um **suporte mínimo** especificado pelo usuário (itemsets frequentes).
- Utilizar todos os itemsets frequentes para gerar todas as regras de associação que possuem confiança maior do que a **confiança mínima** especificada pelo usuário.

IA-2022

16

Algoritmo Apriori

IA-2022

17

Algoritmo Apriori

Encontra todos os k -itemsets frequentes contidos em uma base de dados

- Possui três fases:
 - geração de candidatos,
 - poda de candidatos e
 - validação dos candidatos (cálculo do suporte)
- Executado de forma iterativa:
 - Os itemsets de tamanho k são calculados a partir dos itemsets frequentes de tamanho $k-1$ que já foram calculados no passo anterior (a partir dos itemsets frequentes de tamanho $k-2$, etc).
- As três fases são repetidas em cada iteração

IA-2022

18

Algoritmo Apriori

- Fase da geração dos candidatos de tamanho k
 - Já temos o conjunto L_{k-1} de itemsets frequentes de tamanho $k-1$
 - Gerar os itemsets candidatos (não necessariamente frequentes) de tamanho k a partir do conjunto L_{k-1} .

Para que um itemset de tamanho k tenha chance de ser frequente, todos os itemsets de tamanho $k-1$ contidos no itemset de tamanho k devem também ser frequentes, portanto, **deverão pertencer ao conjunto L_{k-1}**



L_{k-1} = itemsets frequentes de tamanho $k-1$

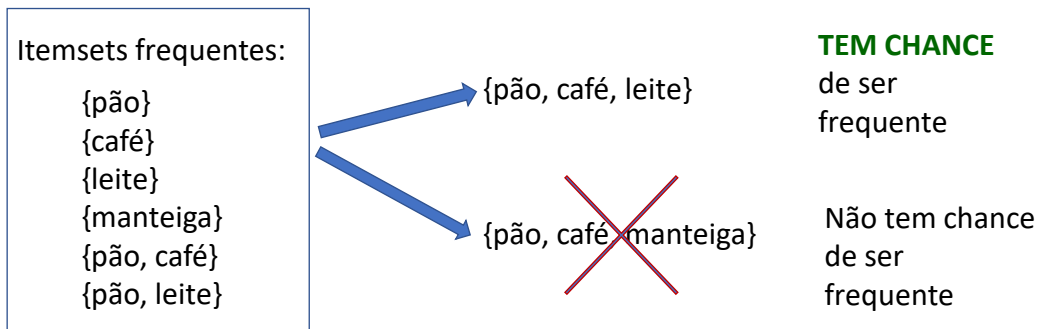
C'_k = itemsets candidatos de tamanho k

IA-2022

19

Algoritmo Apriori

- Fase da geração dos candidatos de tamanho k



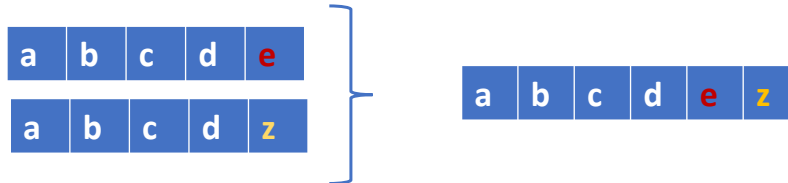
IA-2022

20

Algoritmo Apriori

• Fase da poda dos candidatos

Sabemos que se um itemset de possuir um subconjunto de itens (um subitemset) de tamanho $k-1$ não frequente ele poderá ser descartado, pois não terá a menor chance de ser frequente.



L_{k-1} = itemsets frequentes de tamanho $k-1$

C'_k = itemsets candidatos de tamanho k

Supondo que o itemset $\{b, c, d, e, z\}$ não esteja em L_{k-1} , o itemset $\{a, b, c, d, e, z\}$ pode ser podado

IA-2022

21

Algoritmo Apriori

• Fase do cálculo do suporte

- O banco de dados com as transações é varrido
- Para cada transação t de D verifica-se quais são os candidatos suportados por t e para estes candidatos incrementa-se de uma unidade o contador do suporte.
- Os itemsets com suporte maior ou igual que o suporte mínimo ficam no conjunto L_k

IA-2022

22

Geração de k-itemsets frequentes (entrada: suporte mínimo)

1. $L_1 := \{1\text{-itemsets frequentes}\};$
2. **para** ($k := 2; L_{k-1} \neq \emptyset; k++$) **faça**
3. $C_k := \text{apriori-gen}(L_{k-1});$ **//Gera novos conjuntos candidatos**
4. **para todo** (transações $t \in T$) **faça**
5. $C_t := \text{subset}(C_k, t);$ **//Conjuntos candidatos contidos em t**
6. **para todo** candidato $c \in C_t$ **faça**
7. $c.\text{count}++;$
8. **fim-para**
9. **fim-para**
10. $L_k := \{c \in C_k \mid c.\text{count} \geq \text{sup-min}\};$
11. **fim-para**
12. Resposta $:= \cup_k L_k$

IA-2022

23

Exemplo

$A = \{\text{bermuda, calça, camiseta, sandália, tênis}\}$

Transações	Itens comprados
1	calça, camiseta, tênis
2	camiseta, tênis, calça, sandália
3	bermuda, tênis
4	calça, sandália
5	camiseta, calça, tênis, bermuda

Suporte mínimo: 60%

IA-2022

24

Exemplo

$A = \{\text{bermuda, calça, camiseta, sandália, tênis}\}$

Transações	Itens comprados
1	calça, camiseta, tênis
2	camiseta, tênis, calça, sandália
3	bermuda, tênis
4	calça, sandália
5	camiseta, calça, tênis, bermuda

1-itemset	Suporte
bermuda	40%
calça	80%
camiseta	60%
sandália	40%
tênis	80%

Suporte mínimo: 60%

1-itemsets: $\{\text{bermuda}\}, \{\text{calça}\}, \{\text{camiseta}\}, \{\text{sandália}\}, \{\text{tênis}\}$

$L_1 = \{\{\text{tênis}\}, \{\text{calça}\}, \{\text{camiseta}\}\}$
 $(\{\text{bermuda}\} \text{ e } \{\text{sandália}\})$ não são frequentes

IA-2022

25

Exemplo - Geração de candidatos (passo 3 do algoritmo):

- A função apriori-gen (L_{k-1}) do algoritmo faz a geração e poda dos itemsets candidatos
- Itemsets de tamanho k são gerados a partir do conjunto L_{k-1}

Tran.	Itens comprados
1	calça, camiseta, tênis
2	camiseta, tênis, calça, sandália
3	bermuda, tênis
4	calça, sandália
5	camiseta, calça, tênis, bermuda

- **Geração de Candidatos a itemset com 2 itens:**

- $L_1 = \{\{\text{calça}\}, \{\text{camiseta}\}, \{\text{tênis}\}\}$
- Candidatos: $C_2 = \{\text{calça, camiseta}\}, \{\text{calça, tênis}\}, \{\text{camiseta, tênis}\}$

IA-2022

26

Exemplo - Poda de candidatos (passo 3 do algoritmo):

- **Poda** de candidatos a itemset com 2 itens:
- Retirar do conjunto de candidatos aqueles itemsets que contiverem itemsets de tamanho $k-1$ que não são frequentes

- Para $k=2$ não ocorre a poda

Tran.	Itens comprados
1	calça, camiseta, tênis
2	camiseta, tênis, calça, sandália
3	bermuda, tênis
4	calça, sandália
5	camiseta, calça, tênis, bermuda

- Os itemsets candidatos encontrados ficam no conjunto C_2

$$C_2 = \{\{calça, camiseta\}, \{calça, tênis\}, \{camiseta, tênis\}\}$$

IA-2022

27

Exemplo - Validação de candidatos (passos 4-9 do algoritmo):

Tran.	Itens comprados
1	calça, camiseta, tênis
2	camiseta, tênis, calça, sandália
3	bermuda, tênis
4	calça, sandália
5	camiseta, calça, tênis, bermuda

2-itemsets	Suporte
{calça, camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%

- **Validação** de Candidatos a itemset com 2 itens (passos de 4 a 9 do algoritmo):
- Percorrer as transações e contar o número de ocorrências de cada um.
- Criar o conjunto L_2 com aqueles que tem suporte mínimo
 - $L_2: \{\{calça, camiseta\}, \{calça, tênis\}, \{camiseta, tênis\}\}$

IA-2022

28

Exemplo - Geração de candidatos (passo 3 do algoritmo):

Tran.	Itens comprados
1	calça, camiseta, tênis
2	camiseta, tênis, calça, sandália
3	bermuda, tênis
4	calça, sandália
5	camiseta, calça, tênis, bermuda

- **Geração de Candidatos a itemset com 3 itens:**
 - $L2 = \{\{calça, camiseta\}, \{calça, tênis\}, \{camiseta, tênis\}\}$
 - Candidato: $\{\{calça, camiseta, tênis\}\}$
- **Poda:** não ocorre
- **Validação:** suporte do itemset $\{calça, camiseta, tênis\} = 60\%$

IA-2022

29

Exemplo - final:

- Não é possível formar itemsets de tamanho 4, já que temos somente um itemset frequente de tamanho 3
- Neste ponto o algoritmo termina, com os seguintes itemsets frequentes:

Itemsets frequentes	Suporte
$\{calça\}$	80%
$\{camiseta\}$	60%
$\{tênis\}$	80%
$\{calça, camiseta\}$	60%
$\{calça, tênis\}$	60%
$\{camiseta, tênis\}$	60%
$\{camiseta, calça, tênis\}$	60%

IA-2022

30

Geração de regras de associação

- As regras de associação são geradas a partir de um conjunto contendo todos k-itemsets frequentes, com $k \geq 2$

No exemplo:

Itemsets frequentes	Suporte
{calça}	80%
{camiseta}	60%
{tênis}	80%
{calça,camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{camiseta, calça, tênis}	60%

Itemsets que podem gerar regras:

Itemsets frequentes	Suporte
{calça,camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{camiseta, calça, tênis}	60%

IA-2022

31

Geração de regras de associação – procedimento básico

- A partir de um itemset I de tamanho k , gera todos os subconjuntos a de I e gera regras da forma:

$$a \rightarrow (I-a)$$

$L = \{\text{calça, camiseta, tênis}\}$

Subconjunto do itemset	Gera a regra
{calça}	{calça} \rightarrow {camiseta, tênis}
{camiseta}	{camiseta} \rightarrow {calça, tênis}
{tênis}	{tênis} \rightarrow {calça, camiseta}
{calça, camiseta}	{calça, camiseta} \rightarrow {tênis}
{calça, tênis}	{calça, tênis} \rightarrow {camiseta}
{camiseta, tênis}	{camiseta, tênis} \rightarrow {calça}

IA-2022

32

Algoritmo para gerar regras de associação

- As regras são geradas recursivamente:
- **L = {calça, camiseta, tênis}**
- Considera primeiro {calça, camiseta} gera calça, camiseta → tênis
 - Considera {calça} gera calça → camiseta, tênis
 - Considera {camiseta} gera camiseta → calça, tênis

IA-2022

33

Algoritmo para gerar regras de associação

- para** todo (k-itemset frequente I_k , $k \geq 2$) **faça**
2. Call genrules (I_k, I_k);
 3. **fim-para**
 - // O procedimento genrules gera todas as regras válidas sobre I_k**
 4. **procedure genrules** (I_k : k-itemset frequente, a_m : m-itemset frequente)
 5. gerar subconjuntos não vazios de um itemset frequente com m-1 itens
 6. **para todo** itemset gerado construir regras
 7. **se** confiança da regra gerada \geq confiança mínima
 8. imprimir regra
 9. **se** (m-1>1) **Call** genrules (I_k, a_{m-1});
 - //Gera regras com subconjuntos de a_{m-1} como antecedente**
 10. **fim-se**
 11. **fim-para**

IA-2022

34

Algoritmo para gerar regras de associação

- Serão geradas todas as regras possíveis a partir dos itemsets frequentes com $k \geq 2$

Itemsets frequentes	Suporte
{calça}	80%
{camiseta}	60%
{tênis}	80%
{calça,camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{calça, camiseta, tênis}	60%

Tran.	Itens comprados
1	calça, camiseta, tênis
2	camiseta, tênis, calça, sandália
3	bermuda, tênis
4	calça, sandália
5	camiseta, calça, tênis, bermuda

IA-2022

35

Algoritmo para gerar regras de associação

Itemsets frequentes com $k \geq 2$:

{calça,camiseta}
 {calça, tênis}
 {camiseta, tênis}
 {calça, camiseta, tênis}

O algoritmo **gen-rules** será aplicado a cada um deles

IA-2022

36

Algoritmo para gerar regras de associação

- Para $I_k = \{\text{calça, camiseta}\}$ é chamado:
 - $\text{genrules}(\{\text{calça, camiseta}\}, \{\text{calça, camiseta}\})$
 - gera 1-itemsets não vazios que são subconjuntos desse 2-itemset:
 - $A = \{\{\text{calça}\}, \{\text{camiseta}\}\}$
 - constrói regras com esses itemsets no lado esquerdo e calcula a confiança:
 - Regra 1: **calça -> camiseta**
 - Regra 2: **camiseta -> calça**

IA-2022

37

Algoritmo para gerar regras de associação

Regra 1: **calça -> camiseta**

Itemsets frequentes	Suport
{calça}	80%
{camiseta}	60%
{tênis}	80%
{calça, camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{calça, camiseta, tênis}	60%

$\text{suporte}(R1) = \text{suporte}(\{\text{calça, camiseta}\}) = 60\%$

$\text{confiança}(R1) = \text{suporte}(\{\text{calça, camiseta}\}) / \text{suporte}(\{\text{calça}\})$

$= 60/80 = 75\%$

IA-2022

38

Algoritmo para gerar regras de associação

Regra 2:

camiseta -> calça

Itemsets frequentes	Suporte
{calça}	80%
{camiseta}	60%
{tênis}	80%
{calça,camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{calça, camiseta, tênis}	60%

$\text{suporte}(R2) = \text{suporte}(\{\text{calça, camiseta}\}) = 60\%$

$\text{confiança}(R2) = \text{suporte}(\{\text{calça, camiseta}\}) / \text{suporte}(\{\text{camiseta}\})$

$= 60/60 = 100\%$

IA-2022

39

Algoritmo para gerar regras de associação

• Mesmo processo se repete para os itemsets:

- {calça, tênis}
 - Regra 3: calça → tênis
 - Regra 4: tênis → calça
- {camiseta, tênis}
 - Regra 5: camiseta → tênis
 - Regra 6: tênis → camiseta

IA-2022

40

Algoritmo para gerar regras de associação

- Para $I_k = \{\text{calça, camiseta, tênis}\}$ é chamado:

□ $\text{genrules}(\{\text{calça, camiseta, tênis}\}, \{\text{calça, camiseta, tênis}\})$

- gera 2-itemsets que são subconjuntos desse 3-itemset:
 - $A = \{\{\text{calça, camiseta}\}, \{\text{calça, tênis}\}, \{\text{camiseta, tênis}\}\}$
 - (A contém itemsets de tamanho 2)
- constrói regras com cada um desses itemsets no lado esquerdo e calcula a confiança

IA-2022

41

Algoritmo para gerar regras de associação

Para $a_2 = \{\text{calça, camiseta}\}$

Regra 7: **calça, camiseta -> tênis**

Itemsets frequentes	Suporte
{calça}	80%
{camiseta}	60%
{tênis}	80%
{calça, camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{calça, camiseta, tênis}	60%

$\text{suporte}(R7) = \text{suporte}(\{\text{calça, camiseta, tênis}\}) = 60\%$

$\text{confiança}(R7) = \text{suporte}(\{\text{calça, camiseta, tênis}\}) / \text{suporte}(\{\text{calça, camiseta}\})$

$= 60/60 = 100\%$

IA-2022

42

Algoritmo para gerar regras de associação

- Para o mesmo itemset considerado ($\{\text{calça}, \text{camiseta}\}$) gera 1-itemsets que são subconjuntos dele:

- $A = \{\{\text{calça}\}, \{\text{camiseta}\}\}$
- (A contém itemsets de tamanho 1)

- constrói regras com cada um desses itemsets no lado esquerdo e calcula a confiança:

- Para $a_1 = \{\text{calça}\}$:

Regra 8: **calça** -> **camiseta**, **tênis**

- Para $a_1 = \{\text{camiseta}\}$:

Regra 9: **camiseta** -> **calça**, **tênis**

IA-2022

43

Resumindo...

- Para $I_k = \{\text{calça}, \text{camiseta}, \text{tenis}\}$
- É gerado $A = \{\{\text{calça}, \text{camiseta}\}, \{\text{calça}, \text{tênis}\}, \{\text{camiseta}, \text{tênis}\}\}$
- $\{\text{calça}, \text{camiseta}\}$ dá origem a
 - Regra 7: calça, camiseta -> tênis
 - Regra 8: calça -> camiseta, tênis
 - Regra 9: camiseta -> calça, tênis
- $\{\text{calça}, \text{tênis}\}$ dá origem a
 - Regra 10: calça, tênis -> camiseta
 - Regra 11: calça -> tênis, camiseta (REPETIDA!)
 - Regra 12: tênis -> calça, camiseta
- $\{\text{camiseta}, \text{tênis}\}$ dá origem a
 - Regra 13: camiseta, tênis -> calça
 - Regra 14: camiseta -> tênis, calça (REPETIDA!)
 - Regra 15: tênis -> calça, camiseta (REPETIDA!)

IA-2022

44

Heurísticas para seleção de regras de associação

- **Coeficiente de interesse** ou **Lift**
- Reflete a noção estatística de independência entre duas variáveis aleatórias

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

- Lift igual a 1 A e B são independentes
- Lift inferior a 1 A e B são negativamente correlacionados
- Lift superior a 1 A e B são correlacionados

IA-2022

45

Heurísticas para seleção de regras de associação

Lift

Regra 1: **calça → camiseta**

suporte(R1) = suporte({calça, camiseta}) = 0,6

confiança(R1) = suporte({calça, camiseta}) / suporte({calça}) = 0,6/0,8 = 0,75

Lift(R1) = suporte({calça, camiseta}) / suporte({calça}) * suporte({camiseta})
= 0,6 / (0,8 * 0,6) = 0,6 / 0,48 = 1,25

Itemsets frequentes	Suporte
{calça}	80%
{camiseta}	60%
{tênis}	80%
{calça, camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{calça, camiseta, tênis}	60%

Lift maior que 1: os itens estão relacionados

IA-2022

46

Heurísticas para seleção de regras de associação

• Convicção

- Mede o quão convincente é a regra
- Pode ser interpretada como a frequência de erro da regra
- Frequência esperada de A ocorrer sem B

$$Conv(A \rightarrow B) = \frac{1 - suporte(B)}{1 - confiança(A \rightarrow B)}$$

- Convicção igual a 1 A e B são independentes
- Convicção inferior a 1 a ligação entre A e B é aleatória
- Convicção superior a 1 A e B são correlacionados

IA-2022

47

Heurísticas para seleção de regras de associação

Convicção

Regra 1: **calça -> camiseta**

$$suporte(R1) = suporte(\{calça, camiseta\}) = 0,6$$

$$confiança(R1) = suporte(\{calça, camiseta\}) / suporte(\{calça\}) = 0,6/0,8 = 0,75$$

$$Conv(R1) = 1 - suporte(\{camiseta\}) / 1 - confiança(calça \rightarrow camiseta) \\ = (1 - 0,6) / 0,75 = 0,4 / 0,75 = 0,53$$

Itemsets frequentes	Suporte
{calça}	80%
{camiseta}	60%
{tênis}	80%
{calça, camiseta}	60%
{calça, tênis}	60%
{camiseta, tênis}	60%
{calça, camiseta, tênis}	60%

Convicção menor que 1: os itens não estão relacionados

IA-2022

48



Bons estudos!