

Pravidlové systémy. Klasifikační pravidla. Asociační pravidla.

Petr Pošík

Katedra kybernetiky

ČVUT FEL

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

System AQ

Algoritmus učení AQ

Vlastnosti AQ

System CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Klasifikační pravidla

Agenda

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

- ✓ Jak vypadá pravidlo
- ✓ Množina pravidel vs seznam pravidel
- ✓ Systém AQ
- ✓ Systém CN2

Klasifikační strom vs. klasifikační pravidla

Klasifikační pravidla

Agenda

Klasifikační strom vs. klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

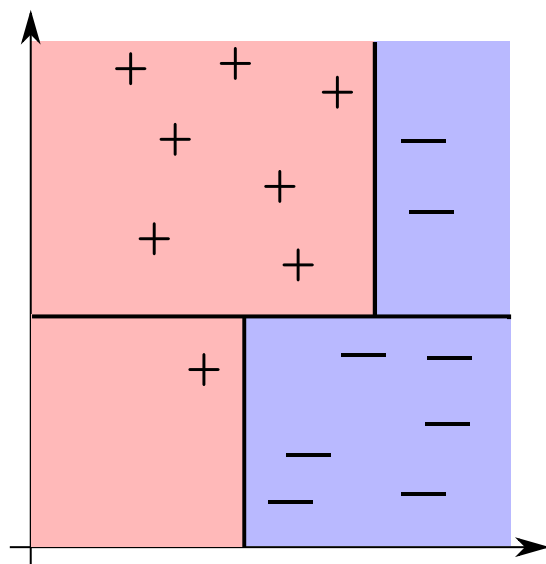
Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

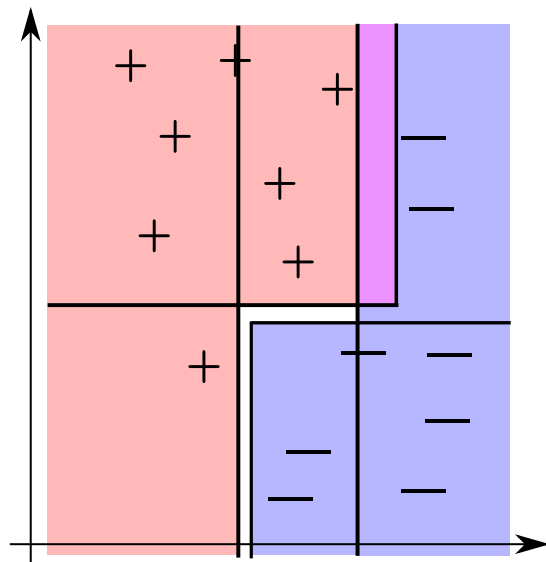
Klasifikační pravidla:
shrnutí

Asociační pravidla



Rozdělení prostoru stromem (splitting)

- ✓ ID3, C4.5, J48, See5
- ✓ Celý prostor je pokryt
- ✓ Rozhodnutí je všude jednoznačné



Pokrytí prostoru pravidly (covering)

- ✓ AQ, CN2
- ✓ V prostoru mohou existovat místa, kde neplatí žádné z pravidel: nutnost tzv. *default rule*
- ✓ Predikce jednotlivých pravidel mohou někde kolidovat: nutnost specifikovat, jak v takovém případě postupovat

Pravidlo č. i : $\text{if } condition_i \text{ then predict } class_i$

Množina pravidel (neuspořádaná): rule set, decision set

- ✓ nezáleží na pořadí vykonávání
- ✓ při ohodnocování se vyhodnotí podmínka všech; z vyhovujících se odvodí jedna finální predikce
 - ✗ např. tak, že se jako predikce použije nejčastější třída mezi všemi trénovacími příklady, které jsou pokryté alespoň jedním vyhovujícím pravidlem;
 - ✗ není-li příklad pokryt žádným pravidlem, použije se jako predikce nejčastější třída v trénovacích datech (default rule).
- ✓ Styl učení: vybereme třídu, pro níž chceme vytvořit pravidlo, a následně najdeme jeho podmínku

Seznam pravidel (uspořádaný): rule list, decision list

- ✓ záleží na pořadí vykonávání
- ✓ výstupem je predikce pravidla, které je splněno jako první; není-li splněno žádné, predikcí je nejčastější třída v trénovacích datech (default rule)
- ✓ horší interpretace, nutné brát v úvahu všechna dřívější pravidla
- ✓ Styl učení: najdeme nejlepší podmínku a zjistíme, kterou třídu popisuje

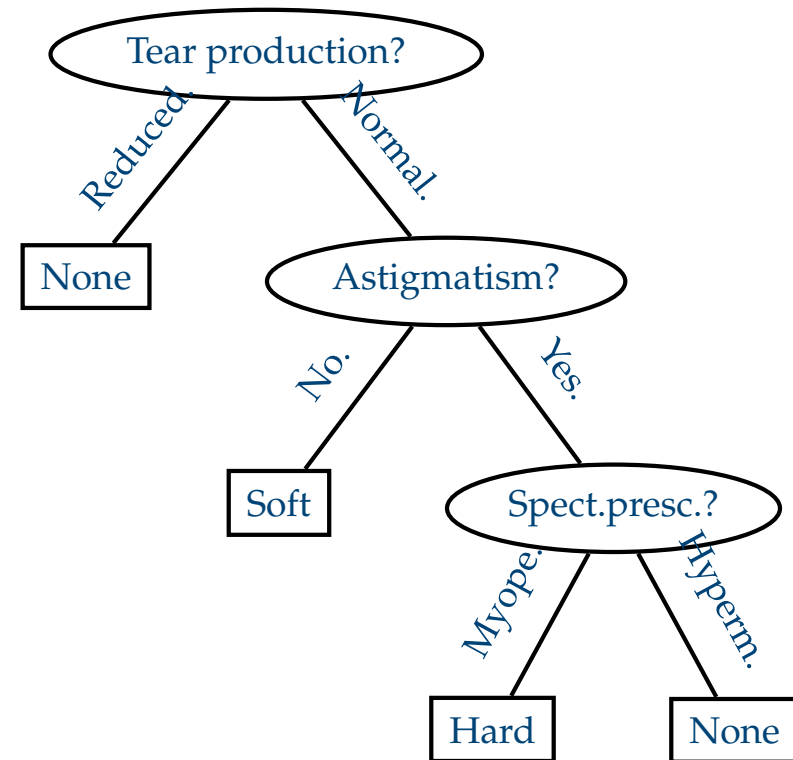
Příklad: kontaktní čočky

Všechny následující modely jsou ekvivalentní:

Seznam pravidel (záleží na pořadí vykonávání):

- ✓ if *Tear prod.* = *Reduced* then *class* = *None*.
- ✓ if *Astigmatism* = *No* then *class* = *Soft*.
- ✓ if *Spect.presc.* = *Myope.* then *class* = *Hard*.
- ✓ default: *class* = *None*.

Klasifikační strom:



Množina pravidel (nezáleží na pořadí vykonávání):

- ✓ if $Tear\ prod. = Reduced \vee (Tear\ prod. = Normal \wedge Astigmatism = Yes \wedge Spect.presc. = Hyperm.)$ then *class* = *None*.
- ✓ if $(Tear\ prod. = Normal \wedge Astigmatism = No)$ then *class* = *Soft*.
- ✓ if $(Tear\ prod. = Normal \wedge Astigmatism = Yes \wedge Spect.presc. = Myope.)$ then *class* = *Hard*.
- ✓ default: *class* = *None*.

AQ indukuje množinu pravidel: `if cover then predict class`

- ✓ pro každou třídu 1 pravidlo
- ✓ třída přiřazená ke coveru je nejčastější třída mezi trénovacími příklady pokrytými coverem

AQ indukuje množinu pravidel: *if cover then predict class*

- ✓ pro každou třídu 1 pravidlo
- ✓ třída přiřazená ke coveru je nejčastější třída mezi trénovacími příklady pokrytými coverem

Používaná terminologie:

- ✓ **Selector** je základní test na hodnotu atributu, příp. výrazu obsahujícího atributy, např.
 - ✗ [color = red \vee white \vee blue]
 - ✗ [temp \in 20..25 \vee 50..60]
 - ✗ [width & height = 5]
 - ✗ [length \times width & length \times height \in 36..40]

AQ indukuje množinu pravidel: *if cover then predict class*

- ✓ pro každou třídu 1 pravidlo
- ✓ třída přiřazená ke coveru je nejčastější třída mezi trénovacími příklady pokrytými coverem

Používaná terminologie:

- ✓ **Selector** je základní test na hodnotu atributu, příp. výrazu obsahujícího atributy, např.
 - ✗ [color = red \vee white \vee blue]
 - ✗ [temp \in 20..25 \vee 50..60]
 - ✗ [width & height = 5]
 - ✗ [length \times width & length \times height \in 36..40]
- ✓ **Complex** je konjunkce (AND) selektorů, např:
 - ✗ [color = red \vee white \vee blue][width & height = 5],a je splněný, jsou-li splněny všechny selektory.

AQ indukuje množinu pravidel: *if cover then predict class*

- ✓ pro každou třídu 1 pravidlo
- ✓ třída přiřazená ke coveru je nejčastější třída mezi trénovacími příklady pokrytými coverem

Používaná terminologie:

- ✓ **Selector** je základní test na hodnotu atributu, příp. výrazu obsahujícího atributy, např.
 - ✗ $[\text{color} = \text{red} \vee \text{white} \vee \text{blue}]$
 - ✗ $[\text{temp} \in 20..25 \vee 50..60]$
 - ✗ $[\text{width} \ \& \ \text{height} = 5]$
 - ✗ $[\text{length} \times \text{width} \ \& \ \text{length} \times \text{height} \in 36..40]$
- ✓ **Complex** je konjunkce (AND) selektorů, např:
 - ✗ $[\text{color} = \text{red} \vee \text{white} \vee \text{blue}][\text{width} \ \& \ \text{height} = 5]$,a je splněný, jsou-li splněny všechny selektory.
- ✓ **Cover** je disjunkce (OR) komplexů, např.:
 - ✗ $[\text{color} = \text{red} \vee \text{white}][\text{width} = 5] \vee [\text{temp} \in 20..25][\text{length} \times \text{width} \in 36..40]$,a je splněný, je-li splněn alespoň 1 z komplexů.

AQ indukuje množinu pravidel: $\text{if } cover \text{ then } predict \text{ class}$

- ✓ pro každou třídu 1 pravidlo
- ✓ třída přiřazená ke coveru je nejčastější třída mezi trénovacími příklady pokrytými coverem

Používaná terminologie:

- ✓ **Selector** je základní test na hodnotu atributu, příp. výrazu obsahujícího atributy, např.
 - ✗ $[color = red \vee white \vee blue]$
 - ✗ $[temp \in 20..25 \vee 50..60]$
 - ✗ $[width \ \& \ height = 5]$
 - ✗ $[length \times width \ \& \ length \times height \in 36..40]$
- ✓ **Complex** je konjunkce (AND) selektorů, např:
 - ✗ $[color = red \vee white \vee blue][width \ \& \ height = 5]$,a je splněný, jsou-li splněny všechny selektory.
- ✓ **Cover** je disjunkce (OR) komplexů, např.:
 - ✗ $[color = red \vee white][width = 5] \vee [temp \in 20..25][length \times width \in 36..40]$,a je splněný, je-li splněn alespoň 1 z komplexů.

Výraz (selektor, komplex, cover) **pokrývá** příklad, pokud je pro daný příklad splněn.

- ✓ Prázdný komplex (konjunkce 0 atributových testů) pokrývá všechny příklady.
- ✓ Prázdný cover (disjunkce 0 komplexů) nepokrývá žádný příklad.

Algoritmus učení AQ

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Pro každou třídu:

- ✓ vytvoří cover (disjunkci komplexů), který bude sloužit jako podmínka pravidla
- ✓ cover se tvoří iterativně: v každé iteraci
 - ✗ vytvoř 1 komplex a
 - ✗ jím pokryté příklady z trénovací množiny dále neber v úvahu.

Algoritmus 1: AQ: indukce pravidla pro 1 třídu

Vstup: Množina pozitivních (P) a negativních (N) příkladů

Výstup: Cover C pokrývající všechny příklady z P a žádný z N : $x \in P \Rightarrow C(x)$,
 $x \in N \Rightarrow \neg C(x)$

```
1 begin
2    $C \leftarrow \emptyset$ 
3   while  $\exists x \in P : \neg C(x)$  do
4     Zvol semínko  $s: s \in P \wedge \neg C(s)$ .
5     Vytvoř hvězdu  $H \leftarrow \text{GetStar}(s, N)$ , množinu komplexů pokrývajících  $s$  a
     nepokrývajících žádný příklad z  $N$ .
6     Zvol nejlepší komplex  $b \leftarrow \text{GetBest}(H)$  podle zvoleného kritéria.
7     Přidej nejlepší komplex  $b$  ke coveru  $C$ .
```

Heuristiky:

- ✓ “Semínko s vyber náhodně.”
- ✓ “Nejlepší je komplex pokrývající největší počet pozitivních případů.”

Algoritmus učení AQ (pokr.)

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Algoritmus 2: AQ: GetStar

Vstup: Semínko s ($s \in P$), negativní příklady (N), maximální velikost hvězdy $maxstar$.

Výstup: "Hvězda" H , množina komplexů pokrývajících s a nepokrývajících N .

```
1 begin
2    $H \leftarrow \{\emptyset\}$ 
3   while kterýkoli komplex z  $H$  pokrývá nějaký příklad z  $N$  do
4     Zvol negativní příklad  $n$  pokrytý některým komplexem z  $H$ .
5     /* Specializuj komplexy v  $H$  tak, aby nepokrývaly  $n$ : */
6     Vytvoř množinu  $E$  všech selektorů, které pokrývají  $s$ , ale ne  $n$ .
7      $H \leftarrow \{x \wedge y : x \in H, y \in E\}$ 
8     Odstraň z  $H$  všechny zbytečné komplexy.
9     while  $|H| > maxstar$  do
10    |   Odstraň z  $H$  nejhorší komplex.
```

Heuristiky:

- ✓ "Negativní příklady n vybírej podle vzrůstající vzdálenosti od s ."
- ✓ "Nejlepší je komplex s největším součtem pokrytých pozitivních a nepokrytých negativních příkladů."

Vlastnosti AQ

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

- ✓ K tvorbě hvězdy se používá tzv. *beam search* s velikostí paprsku *maxstar*.
 - ✗ Hladové prohledávání (*greedy search*) je *beam search* s velikostí paprsku 1.
- ✓ AQ končí v okamžiku, kdy jsou trénovací data ohodnocena zcela správně.
- ✓ Přeučení v případě šumu v datech!!!
- ✓ Novější varianty:
 - ✗ základní algoritmus je stejný
 - ✗ přeučení se zabraňuje předzpracováním dat nebo zjednodušením (prořezáním) výsledné sady pravidel

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

System AQ

Algoritmus učení AQ

Vlastnosti AQ

System CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

- ✓ používá seznam pravidel (decision list)
- ✓ také používá beam search ke hledání nejlepšího komplexu
- ✓ umožňuje přijmout i pravidla, které nejsou zcela konzistentní s trénovacími daty:
 - ✗ uvažuje *všechny* specializace komplexu
 - ✗ může mezi nimi vybírat na základě statistických měr
- ✓ odolnější vůči šumu a nekonzistencím v datech než AQ

Algoritmus učení CN2

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Postup učení:

- ✓ V každé iteraci najdi komplex pokrývající hodně příkladů jedné třídy C a pár příkladů z několika málo jiných tříd.
- ✓ Z trénovací sady odstraň příklady pokryté nalezeným komplexem.
- ✓ Na konec seznamu pravidel přidej `if $Complex(x)$ then predict C .`

Algoritmus 3: CN2: Indukce seznamu pravidel

Vstup: Trénovací množina T ohodnocených příkladů.

Výstup: Seznam pravidel L .

```
1 begin
2    $L \leftarrow \emptyset$ 
3   repeat
4     Najdi nejlepší komplex:  $b \leftarrow \text{GetBestComplex}(T)$ .
5     if  $b \neq \emptyset$  then
6       Vytvoř množinu příkladů pokrytých komplexem:  $T' \leftarrow \{x : x \in T \wedge b(x)\}$ .
7       Odstraň z trénovací sady pokryté příklady:  $T \leftarrow T - T'$ .
8       Označ nejčastější třídu v  $T'$  jako  $C$ .
9       Na konec seznamu  $L$  přidej pravidlo: if  $b(x)$  then predict  $C$ .
10  until nejlepší komplex  $b = \emptyset$  nebo  $T = \emptyset$ 
```

Algoritmus učení CN2 (pokr.)

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Algoritmus 4: AQ: FindBestComplex

Vstup: Trénovací množina T ohodnocených příkladů, množina S všech možných selektorů, maximální velikost hvězdy $maxstar$.

Výstup: Nejlepší komplex b .

```
1 begin
2   Inicializuj hvězdu a nejlepší komplex:  $H \leftarrow \{\emptyset\}, b \leftarrow \emptyset.$ 
3   while  $H \neq \emptyset$  do
4     Specializuj všechny komplexy v  $H$ :  $H' \leftarrow \{x \wedge y : x \in H, y \in S\}$ 
5     Z  $H'$  odstraň komplexy, které již byly v  $H$  (t.j. nejsou specializované).
6     Z  $H'$  odstraň komplexy nepokrývající žádný příklad (např. obsahují spor).
7     foreach  $c \in H'$  do
8       if  $c$  je lepší než  $b$  a  $c$  je statisticky významný then
9          $b \leftarrow c$ 
10    while  $|H'| > maxstar$  do
11      Odstraň z  $H'$  nejhorší komplex.
12     $H \leftarrow H'$ 
```

Heuristiky:

- ✓ “Nejlepší je komplex s *nejnižší entropií* rozdělení tříd mezi pokrytými příklady.”
- ✓ “Komplex je statisticky významný, je-li rozdělení tříd mezi pokrytými příklady významně jiné než v celé trénovací sadě.” (χ^2 test)

Klasifikační pravidla: shrnutí

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Modifikace CN2:

- ✓ použití Laplaceova odhadu přesnosti místo entropie: $\text{LaplaceAccuracy} = \frac{n_C + 1}{n_T + k}$
 - ✗ k je počet tříd v dané doméně, n_C je počet příkladů klasifikovaných do třídy C a n_T je celkový počet příkladů pokrytých pravidel.
- ✓ generování *neuspořádané* množiny pravidel

Klasifikační pravidla: shrnutí

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Modifikace CN2:

- ✓ použití Laplaceova odhadu přesnosti místo entropie: $\text{LaplaceAccuracy} = \frac{n_C + 1}{n_T + k}$
 - ✗ k je počet tříd v dané doméně, n_C je počet příkladů klasifikovaných do třídy C a n_T je celkový počet příkladů pokrytých pravidel.
- ✓ generování *neuspořádané* množiny pravidel

Učení pravidel:

1. **Konstrukce hypotézy:** najdi dobrou sadu n pravidel
 - ✓ obvykle zjednodušeno na *postupné hledání* n pravidel
2. **Konstrukce pravidla:** najdi pár (podmínka, třída)
 - ✓ zvol třídu a zkonstruuj pro ni vhodnou podmínku (AQ) nebo
 - ✓ zkonstruuj podmínku a přiřaď k ní vhodnou třídu (CN2).
3. **Konstrukce podmínky:** najdi sadu m atributových testů
 - ✓ obvykle se zjednodušuje *postupným přidáváním testů* do podmínky

Klasifikační pravidla: shrnutí

Klasifikační pravidla

Agenda

Klasifikační strom vs.
klasifikační pravidla

Pravidla

Příklad: kontaktní čočky

Systém AQ

Algoritmus učení AQ

Vlastnosti AQ

Systém CN2

Algoritmus učení CN2

Klasifikační pravidla:
shrnutí

Asociační pravidla

Modifikace CN2:

- ✓ použití Laplaceova odhadu přesnosti místo entropie: $\text{LaplaceAccuracy} = \frac{n_C + 1}{n_T + k}$
 - ✗ k je počet tříd v dané doméně, n_C je počet příkladů klasifikovaných do třídy C a n_T je celkový počet příkladů pokrytých pravidel.
- ✓ generování *neuspořádané* množiny pravidel

Učení pravidel:

1. **Konstrukce hypotézy:** najdi dobrou sadu n pravidel
 - ✓ obvykle zjednodušeno na *postupné hledání* n pravidel
2. **Konstrukce pravidla:** najdi pár (podmínka, třída)
 - ✓ zvol třídu a zkonstruuj pro ni vhodnou podmínku (AQ) nebo
 - ✓ zkonstruuj podmínku a přiřaď k ní vhodnou třídu (CN2).
3. **Konstrukce podmínky:** najdi sadu m atributových testů
 - ✓ obvykle se zjednodušuje *postupným přidáváním testů* do podmínky

Vlastnosti pravidlových systémů:

- ✓ Bývají srozumitelnější než klasifikační stromy.
- ✓ Postup konstrukce je obtížnější.

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravdel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Asociační pravidla

Transakce: Příklad

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Datová sada pro asociační pravidla

TID	Položky
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

Transakce: Příklad

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Datová sada pro asociační pravidla

TID	Položky
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

A: mléko

B: chléb

C: cereálie

D: cukr

E: vejce

Transakce: Příklad

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Datová sada pro asociační pravidla

TID	Položky
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

A: mléko

B: chléb

C: cereálie

D: cukr

E: vejce

TID	A	B	C	D	E
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	1	0	1
9	1	1	1	0	0

Instance = Transakce

TID: Transaction ID

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Item (položka): pár $atribut = hodnota$ nebo jen $hodnota$

- ✓ atributy se obvykle převádí na indikátory jednotlivých hodnot, např. místo $item = A$ používáme $A = true$

Položka a pravděpodobnost:

- ✓ Položka (item) je náhodný jev (buď se položka v transakci objeví nebo ne).

Itemset (množina položek): podmnožina všech možných položek

- ✓ Příklad: $X = \{A, B, E\}$ (na pořadí nezáleží)

Transakce: uspořádaná dvojice $(TID, itemset)$

Support (podpora)

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Podpora (support) množiny

položek X je podíl transakcí, které obsahují X .

✓ $\text{sup}(\{A\}) = \frac{6}{9}$

✓ $\text{sup}(\{B, C\}) = \frac{4}{9}$

✓ $\text{sup}(\{A, B, E\}) = \frac{2}{9}$

TID	A	B	C	D	E
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	1	0	1
9	1	1	1	0	0

Support (podpora)

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Podpora (support) množiny

položek X je podíl transakcí, které obsahují X .

- ✓ $\text{sup}(\{A\}) = \frac{6}{9}$
- ✓ $\text{sup}(\{B, C\}) = \frac{4}{9}$
- ✓ $\text{sup}(\{A, B, E\}) = \frac{2}{9}$

Podpora a pravděpodobnost:

- ✓ Podpora je odhad pravděpodobnosti výskytu náhodného jevu
- ✓ $\text{sup}(A) = \hat{p}(A)$
- ✓ $\text{sup}(\{A, B, C\}) = \hat{p}(A \wedge B \wedge C)$
- ✓ $\text{sup}(\{A\} \cup \{B\} \cup \{C\}) = \hat{p}(A \wedge B \wedge C)$

TID	A	B	C	D	E
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	1	0	1
9	1	1	1	0	0

Support (podpora)

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Podpora (support) množiny

položek X je podíl transakcí, které obsahují X .

- ✓ $\text{sup}(\{A\}) = \frac{6}{9}$
- ✓ $\text{sup}(\{B, C\}) = \frac{4}{9}$
- ✓ $\text{sup}(\{A, B, E\}) = \frac{2}{9}$

Podpora a pravděpodobnost:

- ✓ Podpora je odhad pravděpodobnosti výskytu náhodného jevu
- ✓ $\text{sup}(A) = \hat{p}(A)$
- ✓ $\text{sup}(\{A, B, C\}) = \hat{p}(A \wedge B \wedge C)$
- ✓ $\text{sup}(\{A\} \cup \{B\} \cup \{C\}) = \hat{p}(A \wedge B \wedge C)$

TID	A	B	C	D	E
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	1	0	1
9	1	1	1	0	0

Častá množina položek (frequent itemset) je taková množina X , jejíž podpora je vyšší než zvolený práh: $\text{sup}(X) \geq s_{\min}$.

Support (podpora)

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

"Subset property"

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Podpora (support) množiny

položek X je podíl transakcí, které obsahují X .

- ✓ $\text{sup}(\{A\}) = \frac{6}{9}$
- ✓ $\text{sup}(\{B, C\}) = \frac{4}{9}$
- ✓ $\text{sup}(\{A, B, E\}) = \frac{2}{9}$

Podpora a pravděpodobnost:

- ✓ Podpora je odhad pravděpodobnosti výskytu náhodného jevu
- ✓ $\text{sup}(A) = \hat{p}(A)$
- ✓ $\text{sup}(\{A, B, C\}) = \hat{p}(A \wedge B \wedge C)$
- ✓ $\text{sup}(\{A\} \cup \{B\} \cup \{C\}) = \hat{p}(A \wedge B \wedge C)$

TID	A	B	C	D	E
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	1	0	1
9	1	1	1	0	0

Častá množina položek (frequent itemset) je taková množina X , jejíž podpora je vyšší než zvolený práh: $\text{sup}(X) \geq s_{\min}$.

Jak najít všechny časté množiny?

- ✓ Naivní přístup: generovat všechny možné množiny X a kontrolovat, zda $\text{sup}(X) \geq s_{\min}$.
- ✓ Kolik je takových množin?

Support (podpora)

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Podpora (support) množiny

položek X je podíl transakcí, které obsahují X .

- ✓ $\text{sup}(\{A\}) = \frac{6}{9}$
- ✓ $\text{sup}(\{B, C\}) = \frac{4}{9}$
- ✓ $\text{sup}(\{A, B, E\}) = \frac{2}{9}$

Podpora a pravděpodobnost:

- ✓ Podpora je odhad pravděpodobnosti výskytu náhodného jevu
- ✓ $\text{sup}(A) = \hat{p}(A)$
- ✓ $\text{sup}(\{A, B, C\}) = \hat{p}(A \wedge B \wedge C)$
- ✓ $\text{sup}(\{A\} \cup \{B\} \cup \{C\}) = \hat{p}(A \wedge B \wedge C)$

TID	A	B	C	D	E
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	1	0	1
9	1	1	1	0	0

Častá množina položek (frequent itemset) je taková množina X , jejíž podpora je vyšší než zvolený práh: $\text{sup}(X) \geq s_{\min}$.

Jak najít všechny časté množiny?

- ✓ Naivní přístup: generovat všechny možné množiny X a kontrolovat, zda $\text{sup}(X) \geq s_{\min}$.
- ✓ Kolik je takových množin? 2^N , kde N je počet položek.
- ✓ Nešlo by prostor možných množin nějak prořezat?

“Subset property”

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Downward closure property, anti-monotonicity property:

- ✓ **Každá podmnožina časté množiny položek je častá!**
- ✓ Proč?

“Subset property”

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Downward closure property, anti-monotonicity property:

- ✓ **Každá podmnožina časté množiny položek je častá!**
- ✓ Proč?
- ✓ Předpokládejme, že $\{A, B\}$ je častá. Každý výskyt $\{A, B\}$ představuje také 1 výskyt $\{A\}$ a 1 výskyt $\{B\}$. $\{A\}$ i $\{B\}$ tedy také musejí být časté.
- ✓ Když pravidlo obrátíme: Množina s N prvky *může být* častá jen tehdy, když jsou časté všechny její podmnožiny o velikosti $N - 1, N - 2, \dots, 1$.
- ✓ Stačí, jsou-li časté všechny podmnožiny velikosti $N - 1$?

“Subset property”

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Downward closure property, anti-monotonicity property:

- ✓ **Každá podmnožina časté množiny položek je častá!**
- ✓ Proč?
- ✓ Předpokládejme, že $\{A, B\}$ je častá. Každý výskyt $\{A, B\}$ představuje také 1 výskyt $\{A\}$ a 1 výskyt $\{B\}$. $\{A\}$ i $\{B\}$ tedy také musejí být časté.
- ✓ Když pravidlo obrátíme: Množina s N prvky *může být* častá jen tehdy, když jsou časté všechny její podmnožiny o velikosti $N - 1, N - 2, \dots, 1$.
- ✓ Stačí, jsou-li časté všechny podmnožiny velikosti $N - 1$? **Ano.**

Téměř všechny algoritmy pro hledání asoc. pravidel využívají tuto vlastnost!

- ✓ Přesněji: využívají možnost prořezat prostor všech podmnožin množiny všech položek.
- ✓ “Žádná množina obsahující podmnožinu, která není častá, nemůže být častá.”

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Časté množiny velikosti větší než 1:

- ✓ Algoritmus Apriori
- ✓ Myšlenka: ze známých 1-prvkových množin nagenerej 2-prvkové, z 2-prvkových 3-prvkové, ...
- ✓ k -prvkovou množinu zkonstruuj sjednocením všech $(k - 1)$ -prvkových

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Časté množiny velikosti větší než 1:

- ✓ Algoritmus Apriori
- ✓ Myšlenka: ze známých 1-prvkových množin nagenerej 2-prvkové, z 2-prvkových 3-prvkové, ...
- ✓ k -prvkovou množinu zkonstruuj sjednocením všech $(k - 1)$ -prvkových

Příklad:

- ✓ Mějme pět častých 3-prvkových množin:
 $\{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}$
- ✓ Lexikografické uspořádání zlepšuje efektivitu
- ✓ Je množina $\{A, B, C, D\}$ kandidátem na častou 4-prvkovou množinu?

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Časté množiny velikosti větší než 1:

- ✓ Algoritmus Apriori
- ✓ Myšlenka: ze známých 1-prvkových množin nagenerej 2-prvkové, z 2-prvkových 3-prvkové, ...
- ✓ k -prvkovou množinu zkonstruuj sjednocením všech $(k - 1)$ -prvkových

Příklad:

- ✓ Mějme pět častých 3-prvkových množin:
 $\{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}$
- ✓ Lexikografické uspořádání zlepšuje efektivitu
- ✓ Je množina $\{A, B, C, D\}$ kandidátem na častou 4-prvkovou množinu?
Ano, protože všechny 3-prvkové podmnožiny jsou časté.

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Časté množiny velikosti větší než 1:

- ✓ Algoritmus Apriori
- ✓ Myšlenka: ze známých 1-prvkových množin nagenerej 2-prvkové, z 2-prvkových 3-prvkové, ...
- ✓ k -prvkovou množinu zkonstruuj sjednocením všech $(k - 1)$ -prvkových

Příklad:

- ✓ Mějme pět častých 3-prvkových množin:
 $\{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}$
- ✓ Lexikografické uspořádání zlepšuje efektivitu
- ✓ Je množina $\{A, B, C, D\}$ kandidátem na častou 4-prvkovou množinu?
Ano, protože všechny 3-prvkové podmnožiny jsou časté.
- ✓ Je množina $\{A, B, C, D\}$ častou 4-prvkovou množinou?

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Časté množiny velikosti větší než 1:

- ✓ Algoritmus Apriori
- ✓ Myšlenka: ze známých 1-prvkových množin nagenerej 2-prvkové, z 2-prvkových 3-prvkové, ...
- ✓ k -prvkovou množinu zkonstruuj sjednocením všech $(k - 1)$ -prvkových

Příklad:

- ✓ Mějme pět častých 3-prvkových množin:
 $\{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}$
- ✓ Lexikografické uspořádání zlepšuje efektivitu
- ✓ Je množina $\{A, B, C, D\}$ kandidátem na častou 4-prvkovou množinu?
Ano, protože všechny 3-prvkové podmnožiny jsou časté.
- ✓ Je množina $\{A, B, C, D\}$ častou 4-prvkovou množinou?
Nevíme, neznáme její podporu, nelze rozhodnout zda $\text{sup}(\{A, B, C, D\}) \geq s_{\min}$

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Časté množiny velikosti větší než 1:

- ✓ Algoritmus Apriori
- ✓ Myšlenka: ze známých 1-prvkových množin nagenerej 2-prvkové, z 2-prvkových 3-prvkové, ...
- ✓ k -prvkovou množinu zkonstruuj sjednocením všech $(k - 1)$ -prvkových

Příklad:

- ✓ Mějme pět častých 3-prvkových množin:
 $\{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}$
- ✓ Lexikografické uspořádání zlepšuje efektivitu
- ✓ Je množina $\{A, B, C, D\}$ kandidátem na častou 4-prvkovou množinu?
Ano, protože všechny 3-prvkové podmnožiny jsou časté.
- ✓ Je množina $\{A, B, C, D\}$ častou 4-prvkovou množinou?
Nevíme, neznáme její podporu, nelze rozhodnout zda $\text{sup}(\{A, B, C, D\}) \geq s_{\min}$
- ✓ Je množina $\{A, C, D, E\}$ kandidátem na častou 4-prvkovou množinu?

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Časté množiny velikosti 1:

- ✓ Najdi všechny 1-prvkové množiny položek s dostatečnou podporou

Časté množiny velikosti větší než 1:

- ✓ Algoritmus Apriori
- ✓ Myšlenka: ze známých 1-prvkových množin nagenerej 2-prvkové, z 2-prvkových 3-prvkové, ...
- ✓ k -prvkovou množinu zkonstruuj sjednocením všech $(k - 1)$ -prvkových

Příklad:

- ✓ Mějme pět častých 3-prvkových množin:
 $\{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}$
- ✓ Lexikografické uspořádání zlepšuje efektivitu
- ✓ Je množina $\{A, B, C, D\}$ kandidátem na častou 4-prvkovou množinu?
Ano, protože všechny 3-prvkové podmnožiny jsou časté.
- ✓ Je množina $\{A, B, C, D\}$ častou 4-prvkovou množinou?
Nevíme, neznáme její podporu, nelze rozhodnout zda $\text{sup}(\{A, B, C, D\}) \geq s_{\min}$
- ✓ Je množina $\{A, C, D, E\}$ kandidátem na častou 4-prvkovou množinu?
Ne, podmnožiny $\{A, D, E\}$ a $\{C, D, E\}$ nejsou časté.

Klasifikační vs. asociační pravidla

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Klasifikační pravidla

- ✓ Učení s učitelem:
- ✓ Jedna cílová proměnná
- ✓ Míra: **accuracy**

Asociační pravidla

- ✓ Učení bez učitele:
- ✓ Mnoho cílových proměnných
- ✓ Míra: **support, confidence, lift**

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání částých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Asociační pravidlo $R: X \Rightarrow Y$

- ✓ X a Y jsou disjunkttní množiny položek, Y neprázdná.
- ✓ “Obsahuje-li transakce množinu X , obsahuje také Y .”

Asociační pravidla

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Asociační pravidlo $R: X \Rightarrow Y$

- ✓ X a Y jsou disjunkttní množiny položek, Y neprázdná.
- ✓ “Obsahuje-li transakce množinu X , obsahuje také Y .”

Příklad: z časté množiny položek $\{A, B, C\}$ lze zkonstruovat následující pravidla:

- ✓ $A, B \Rightarrow C$
- ✓ $A, C \Rightarrow B$
- ✓ $B, C \Rightarrow A$
- ✓ $A \Rightarrow B, C$
- ✓ $B \Rightarrow A, C$
- ✓ $C \Rightarrow A, B$
- ✓ $\{\} \Rightarrow A, B, C$ nebo $true \Rightarrow A, B, C$

Míry asociačního pravidla

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Pravidlo $R : X \Rightarrow Y$

Podpora (support) pravidla R :

$$\text{sup}(R) = \text{sup}(X \cup Y) = \frac{a}{n}$$

	Y	$\neg Y$	Σ
X	a	b	r
$\neg X$	c	d	s
Σ	k	l	n

Míry asociačního pravidla

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Pravidlo $R : X \Rightarrow Y$

Podpora (support) pravidla R :

$$\text{sup}(R) = \text{sup}(X \cup Y) = \frac{a}{n}$$

Spolehlivost (confidence) pravidla R :

$$\text{conf}(R) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X)} = \frac{a}{r}$$

✓ Spolehlivost je odhad podmíněné pravděpodobnosti:

$$\text{conf}(R) = \hat{p}(Y|X) = \frac{\hat{p}(X \wedge Y)}{\hat{p}(X)}$$

	Y	$\neg Y$	Σ
X	a	b	r
$\neg X$	c	d	s
Σ	k	l	n

Míry asociačního pravidla

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Pravidlo $R : X \Rightarrow Y$

Podpora (support) pravidla R :

$$\text{sup}(R) = \text{sup}(X \cup Y) = \frac{a}{n}$$

Spolehlivost (confidence) pravidla R :

$$\text{conf}(R) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X)} = \frac{a}{r}$$

✓ Spolehlivost je odhad podmíněné pravděpodobnosti:

$$\text{conf}(R) = \hat{p}(Y|X) = \frac{\hat{p}(X \wedge Y)}{\hat{p}(X)}$$

Zdvih (lift) pravidla R :

$$\text{lift}(R) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X) \times \text{sup}(Y)} = \frac{\text{conf}(R)}{\text{sup}(Y)} = \frac{\hat{p}(Y|X)}{\hat{p}(Y)} = \frac{\frac{a}{r}}{\frac{k}{n}}$$

- ✓ Zdvih je poměr pozorované podpory pravidla vůči hodnotě, kterou bychom pozorovali, kdyby X a Y byly nezávislé.
- ✓ Je-li $\text{lift}(R) > 1$, X a Y jsou pozitivně korelované.
- ✓ Je-li $\text{lift}(R) < 1$, X a Y jsou negativně korelované.
- ✓ Je-li $\text{lift}(R) = 1$, X a Y jsou nezávislé.

	Y	$\neg Y$	Σ
X	a	b	r
$\neg X$	c	d	s
Σ	k	l	n

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Silné asociační pravidlo:

- ✓ Pravidlo R , pro něž $\text{sup}(R) \geq s_{\min}$ a $\text{conf}(R) \geq c_{\min}$

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Silné asociační pravidlo:

- ✓ Pravidlo R , pro něž $\text{sup}(R) \geq s_{\min}$ a $\text{conf}(R) \geq c_{\min}$

Hledání silných pravidel:

- ✓ Hlavní myšlenka: “subset property” (opět)
- ✓ Najdi množiny položek X , kde $\text{sup}(X) \geq s_{\min}$.
- ✓ Pro každou častou množinu položek X :
 - ✗ Pro každou její podmnožinu Y :
 - ✓ Vytvoř všechna pravidla $R: X - Y \Rightarrow Y$, otestuj zda $\text{conf}(R) \geq c_{\min}$

Hledání silných asoc. pravidel

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Silné asociační pravidlo:

- ✓ Pravidlo R , pro něž $\text{sup}(R) \geq s_{\min}$ a $\text{conf}(R) \geq c_{\min}$

Hledání silných pravidel:

- ✓ Hlavní myšlenka: “subset property” (opět)
- ✓ Najdi množiny položek X , kde $\text{sup}(X) \geq s_{\min}$.
- ✓ Pro každou častou množinu položek X :
 - ✗ Pro každou její podmnožinu Y :
 - ✓ Vytvoř všechna pravidla $R: X - Y \Rightarrow Y$, otestuj zda $\text{conf}(R) \geq c_{\min}$

Vysoká podpora a spolehlivost pravidla často nestačí!

- ✓ Předp., že X a Y jsou nezávislé a že $\text{sup}(X) \approx p(X) = 0.9$ a $\text{sup}(Y) \approx p(Y) = 0.8$.
- ✓ Potom:
 - ✗ $\text{sup}(X \cup Y) \approx p(X \wedge Y) = 0.72$,
 - ✗ $\text{conf}(X \Rightarrow Y) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X)} = 0.8$ a $\text{conf}(Y \Rightarrow X) = \frac{\text{sup}(X \cup Y)}{\text{sup}(Y)} = 0.9$
 - ✗ Ačkoli obě pravidla mohou být silná, neříkají nám prakticky nic. Viz lift:
 - ✗ $\text{lift}(X \Rightarrow Y) = \frac{\text{conf}(X \Rightarrow Y)}{\text{sup}(Y)} = 1$ a $\text{lift}(Y \Rightarrow X) = \frac{\text{conf}(Y \Rightarrow X)}{\text{sup}(X)} = 1$
- ✓ K vyfiltrování **skutečně užitečných** pravidel musíme použít jiné míry, než podporu a spolehlivost: *lift, leverage, conviction, ...*

Možná rozšíření asociačních pravidel

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání částých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Hledání zajímavých *rozdílů* v pravidlech:

- ✓ Nalezeno pravidlo mléko \Rightarrow chléb, ale nikoli pravidlo sojové mléko \Rightarrow chléb.
Co to znamená?

Využití hierarchické struktury produktů:

- ✓ nápoj \rightarrow mléko \rightarrow nízkotučné mléko
- ✓ Hledání asociací na všech úrovních

Sekvence položek v čase:

- ✓ Když nejdřív X , tak později Y .

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociační pravidla

Asociační pravidla

Míry asociačního pravidla

Hledání silných asoc. pravidel

Možná rozšíření asociačních pravidel

Aplikace

Asociační pravidla: souhrn

Analýza nákupního koše:

- ✓ Umístění zboží v prodejnách, cílené nabídkové akce
- ✓ Elektronické obchody: nabídka podobných titulů (viz Amazon)
- ✓ Legendární případ “pivo a pleny”

Analýza propojení:

- ✓ Odhalování struktury v různých “sociálních” sítích na základě četnosti kontaktů

Součást systémů pro podporu rozhodování:

- ✓ např. $\{car = porsche, gender = male, age < 20\} \Rightarrow \{risk = high, insurance = high\}$

Hledání neobvyklých událostí:

- ✓ WSARE: What is strange about recent events

Asociační pravidla: souhrn

Klasifikační pravidla

Asociační pravidla

Transakce: Příklad

Definice

Support (podpora)

“Subset property”

Hledání častých množin položek

Klasifikační vs. asociční pravidla

Asociační pravidla

Míry asocičního pravidla

Hledání silných asocičních pravidel

Možná rozšíření asocičních pravidel

Aplikace

Asociační pravidla: souhrn

- ✓ Časté množiny položek (podpora)
- ✓ Subset property
- ✓ Algoritmus Apriori
- ✓ Asociační pravidla (spolehlivost)
- ✓ Filtrování silných pravidel
- ✓ Aplikační oblasti