Granular Computing 9999
pages 1–5

METODY SZTUCZNEJ INTELIGENCJI - PROJEKTY

PB

1 Projekt z wyznaczania reduktów zbioru

Liczba osób realizujących projekt: 1-2 osoby

1. Wczytanie danych w formatach arf, tab
2. Wybór atrybutów, które mają zostać uwzględnione podczas wyszukania reduktów
3. Wykonanie poszczególnych kroków algorytmu z wypisaniem wyników pośrednich

1.1 Nazewnictwo

$(x_1, x_2, ...)$ - zbiór obiektów, reprezentujących dane
$x_i = \{x_i^1, x_i^2, .., x_i^p\}$, gdzie x_i^j oznacza atrybut o indeksie j obiektu x_i.
U przestrzeń wszystkich obiektów
X - podzbiór zbioru wszystkich obiektów U
x_i - obiekt należący do podzbioru wszystkich obiektów U
A - zbiór wszystkich atrybutów, cech, właściwości
a_i - atrybut należący do zbioru atrybutów A
V_{a_i} - zbiór wszystkich wartości atrybutu a_i (nazywany dziedziną a_i)
$V(a_i)$ - zbiór wszystkich wartości atrybutu a_i (nazywany dziedziną a_i)
B - niepusty podzbiór A ($B \subseteq A$)
$LOW(X_B)$ - dolna aproksymacja X względem B
$\underline{X}_B$ - dolna aproksymacja X względem B
$UPP(X_B)$ - górna aproksymacja X względem B
$\overline{X}_B$ - górna aproksymacja X względem B
AS_B - standardowa przestrzeń aproksymacyjna
$AS_{\#,\$}$ - sparametryzowana przestrzeń aproksymacyjna
$R_{a_i}(X)$ - przybliżoność ze wzdgledu na $\{a_i\}$
$Rough_{a_j}(a_i)$ - średnia przybliżoność atrybutu a_i względem atrybutu $\{a_j\}$
$MR(a_i)$ - minimalna przybliżoność atrybutu a_i
MMR - minimalna wartość MR wszystkich atrybutów
$IND(B)$ - relacja nierozróżnialności
$[x_i]_{IND(B)}$ - klasa równoważności obiektu x_i w relacji $IND(B)$, nazywana także zbiorem elementarnym w B
$(C_1, C_2, \ldots, C_K)$ - klasy, skupienia w danym pogrupowaniu danych
$Card(X)$ - liczebność zbioru X
$|X|$ - liczebność zbioru X
$P(U)$ - zbiór potęgowy zbioru U

2 Algorytm - Expansion Algorithm

Przykład, dla podanego systemu informacyjnego:

	a	b	c	d	E

x_1	1	0	2	1	1
x_2	1	0	2	0	1
x_3	1	2	0	0	2
x_4	1	2	2	1	0
x_5	2	1	0	0	2
x_6	2	1	1	0	2
x_7	2	1	2	1	1

Tablica 2: System informacyjny

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1							
x_2	d						
x_3	b,c,d	b,c					
x_4	b	b,d	c,d				
x_5	a,b,c,d	a,b,c	a,b,d	a,b,c,d			
x_6	a,b,c,d	a,b,c,d	a,b,c	a,b,c,d	c		
x_7	a,b	a,b,d	a,b,c,d	a,b	c,d	c,d	

Tablica 3: System informacyjny - macierz rozróżnialności

W tabeli decyzyjnej, dwa atrybuty x oraz y nazywane są silnie równoważnymi jeżeli występują zawsze razem w maceirzy rozróżnialnosci. Każdy element można rozpatrywać w postaci alternatywy, to znaczy, jeżeli element macierzy rozróżnialności ma postać a, b, c wtedy można go zapisać jako $a \vee b \vee c$. Funkcja rozróżnialności ma postać koniunkcji wyrażeń w postaci alternatywy macierzy rozróżnialności.

EXPANSION LAW

1. odszukać atrybut X występujący najczęściej (przynajmniej raz)
2. wykonać operacje AND z X i z wszystkimi pozostałymi OR z macerzy nierozróżnialnosci, które nie zawierają atrybutu X
3. zastosować koniunkcję AND z elementami OR wszystkich elementów, w których jeżeli element zawiera X, wyeliminować X.
4. połączyć elementy otrzymane z punktu (2) i (3)

Przykład: niech będzie zadana macierz rozróżnialnosci: $\{\{a, b, e\}, \{a, b\}, \{a, c\}, \{d\}\}$. Relacja rozróżnialności zadana jest w postaci: $\{a \vee b \vee e\} \wedge \{a \vee b\} \wedge \{a \vee c\} \wedge d$

W podanej relacji, element a występuje często. Stosujemy operację AND z a i d. Otrzymujemy $\{a\} \vee \{d\} = \{a, d\}$ - nazwijmy jako komponent I. Po zastosowaniu operacji AND dla $b \vee c$, b i c otrzymujemy: $(b \vee e) \wedge (b) \wedge (c) = \{b, c\} \wedge \{b, c, e\}$ - jako komponent II. W postaci połączonej komponnetu I oraz II mamy: $\{a, d\}, \{b, c\}, \{b, c, e\}$

Algorithm 1: Algorytm Ekspansji

Data: System Informacyjny
Result: Redukt systemu informacyjnego
Zdefiniować funkcję rozróżnialności $f = f_1 \wedge f_2 \wedge ... \wedge f_k$

Step 1 Zastosować prawo absorbcji do wyeliemonowania wszystkich wyrażeń w postaci alternatywy, które stanowią nadzbiór pozostałych wyrażeń w postaci alternatywy.

Step 2: Zastapić zbiór silnie równoważnych atrybutów zmienną zastępczą.

Step 3: Wybrać atrybut, występujący najczęściej w zbiorach w postaci koniunkcji (przynajmniej dwa razy), i zastosować prawo ekspansji.

Step 4: Powórzyć kroki od 1 do 3, do chwili gdy nie można zastsować prawa ekspansji dla każdego komponentu.

Step 5: Zastapić wszystkie silnie równoważne klasy ich odpowiadającymi atrybutami.

Step 6: Wyznaczyć redukt dla każdego komponentu.

Step 7: Wypisać zintegrowany redukt.

Algorithm 2: Algorytm Ekspansji - przykład

Data: System Informacyjny
Result: Redukt systemu informacyjnego - przykład

Relacja rozróżnialności:
$F = \{a \vee b \vee c \vee f\} \wedge \{b \vee d\} \wedge \{a \vee d \vee e \vee f\} \wedge \{a \vee b \vee c \vee d\} \wedge \{b \vee d \vee e \vee f\} \wedge \{c \vee d\}$

Zastosowanie prawa absorbcji:
$\{b \vee d\} \subseteq \{a \vee b \vee c \vee d\} \Rightarrow \{b \vee d\} \wedge \{a \vee b \vee c \vee d\} = \{b \vee d\}$

Zastosowanie prawa absorbcji:
$\{b \vee d\} \subseteq \{b \vee d \vee e \vee f\} \Rightarrow \{b \vee d\} \wedge \{b \vee d \vee e \vee f\} = \{b \vee d\}$

Po powyższych przekształceniach, relacja rozróżnialności przyjmuje postać:
$F = \{a \vee b \vee c \vee f\} \wedge \{b \vee d\} \wedge \{a \vee d \vee e \vee f\} \wedge \{c \vee d\}$

Relacja silnej równoważności:
Atrybuty $\{a, f\}$ silnie równoważne. Oznaczanie $M = \{a \vee f\}$
relacja rozróżnialności przyjmuje postać:
$F = \{M \vee b \vee c\} \wedge \{b \vee d\} \wedge \{M \vee d \vee e\} \wedge \{c \vee d\}$

Atrybut d występuje często. Zastosowanie prawa ekspansji:
$F = [\{d\} \wedge \{M \vee b \vee c\}] \wedge [\{M \vee b \vee c\} \wedge \{b\} \wedge \{M \vee e\} \wedge \{c\}]$

Zastosowanie prawa absorbcji dla drugiego komponentu:
$F = [d \wedge \{M \vee b \vee c\}] \wedge [\{b\} \wedge \{M \vee e\} \wedge \{c\}]$
W ten sposób wszystkie komponenty znajdują się w prostej postaci.

Zastąpienie M przez $a \vee f$ daje w wyniku:
$F = [\{d\} \wedge \{M \vee b \vee c\}] \wedge [\{b\} \wedge \{M \vee e\} \wedge \{c\}]$

Redukt pierwszego komponentu: $R_1 = \{a, d\}, \{d, f\}, \{b, d\}, \{b, c\}$
Redukt drugiego komponentu: $R_2 = \{a, b, c\}, \{b, c, f\}, \{b, c, e\}$

Redukt wynikowy przyjmuje postać:
$R = \{a, d\}, \{d, f\}, \{b, d\}, \{b, c\}, \{a, b, c\}, \{b, c, f\}, \{b, c, e\}$