Bogusław Guzik
Milda Burzała

PAKIET PROGRAMOWY OBLICZANIA MIERNIKÓW SYNTETYCZNYCH I DENDRYTÓW PRZY DWUSZCZEBLOWYM PODZIALE ZBIORU CECH

Wstęp

Niech P oznacza zbiór m obiektów, a C niech oznacza zbiór cech opisujących te obiekty. Przyjmujemy, że dokonano dwuszczeblowego podziału zbioru cech. Najpierw w zbiorze C wyodrębniono niepuste, rozłączne i wyczerpujące zbiór C podzbiory cech U_u (u=1,2,...,n) zwane dziedzinami. Z kolei w każdej dziedzinie wyodrębniono niepuste, rozłączne i wyczerpujące zbiór U_u podzbiory cech S_s^u (s=1,2,...,s_u) zwane sferami.
W prezentowanym programie obliczeniowym G1DØ zakłada się, że dziedzinę traktuje się jako zbiór cech (elementem składowym dziedziny jest cecha). Możliwe jest też traktowanie dziedziny jako zbioru sfer (elementem składowym dziedziny jest wówczas sfera). Rozróżnienie między dziedziną jako zbiorem cech a dziedziną jako zbiorem sfer jest ważne z punktu widzenia sposobu obliczania miernika syntetycznego i miernika zróżnicowania obiektów. W pierwszym przypadku są one funkcją określoną na wartościach cech tworzących dziedzinę, w drugim - funkcją określoną na wartościach odpowiednio: mierników syntetycznych, mierników zróżnicowania dla sfer tworzących dziedzinę.

1. Możliwości programu G1DØ

Program realizuje m.in. następujące funkcje:
1) Przeprowadza normalizację cech wg zadeklarowanych przez użytkownika sposobów normalizacji.
2) Ustala wartości wzorców poszczególnych cech wg reguł wybranych przez użytkownika.

3) Oblicza wartości miernika syntetycznego i jego charakterystyki statystyczne. Typy mierników syntetycznych deklaruje użytkownik programu.
4) Porządkuje obiekty na podstawie wartości miernika syntetycznego.
5) W przypadku danych w postaci szeregów czasowych (obiektami są wówczas lata), oblicza się macierz różnic między wartościami miernika syntetycznego dla poszczególnych obiektów-lat oraz charakterystyki statystyczne różnic dla danego roku względem pozostałych lat.
6) Oblicza macierz zróżnicowania (np. odległości) między obiektami oraz charakterystyki statystyczne zróżnicowania. Typ mierników zróżnicowania deklaruje użytkownik programu.
7) Wyznacza dendryty wrocławskie i ich charakterystyki statystyczne przy danym typie miernika zróżnicowania.
8) Oblicza tzw, „przewagi" obiektu nad pozostałymi obiektami oraz charakterystyki statystyczne przewag wg typu przewag deklarowanych przez użytkownika.
9) Wyznacza dendryty wrocławskie dla danej macierzy „przewag" oraz ich charakterystyki stytystyczne.
10) Ponadto program tworzy trzy wyjściowe taśmy magnetyczne (lub zbiory wynikowe typu MT) zawierające m.in.:
- wartości mierników syntetycznych,
- dendryty,
- macierze oryginalnych wartości cech.

Daje to możliwości dalszego przetwarzania tych informacji bez konieczności przeprowadzania obliczeń od początku.
Obliczenia wymienione w pkt. 3-9 realizowane są z osobna dla poszczególnych sfer i dziedzin.
Mierniki zróżnicowania oraz „przewagi" zdefiniowane będą w dalszej części pracy.

2. Przygotowanie danych wejściowych

Program G1DØ wykorzystuje dwa zbiory danych wejściowych. Pierwszy zbiór (typu ✱TR - czytnik tasiemki papierowej) zawiera macierz wyników obserwacji cech tworzących zbiór C. O cechach tych zakłada się, że mają one postać stymulant. Drugi zbiór (typu ✱CR - czytnik kart perforowanych 80-kolumnowych) zawiera dane sterujące obliczeniami programu G1DO.

2.1. Macierz wyników obserwacji cech - stymulant

Przejście z cech pierwotnych na cechy-stymulanty używane w wielowymiarowej analizie porównawczej wymaga zwykle przekształceń zmiennych pierwotnych. Jeden rodzaj przekształceń to przekształcenia doprowadzające cechy-destymulanty lub cechy-nominanty do postaci cech-stymulant. Inny rodzaj przekształceń to przekształcenia doprowadzające cechy pierwotne do postaci np. wskaźników natężenia, indeksów, cech agregatowych itp. Ponieważ

takich możliwych przekształceń jest wiele, przyjęto, że macierz wyników obserwacji cech-stymulant tworzona jest w osobnym przebiegu przez pakiet XDS3 („Analiza statystyczna") znajdującym się na fabrycznym wyposażeniu EMC Odra 1305. Pakiet ten umożliwia w dogodny sposób wykonanie praktycznie wszystkich spotykanych w WAP przekształceń cech.
Program G1DØ zakłada, że macierz wyników obserwacji cech-stymulant drukowana jest w przebiegu pakietu XDS3 instrukcją

POBS,66,SSSS

gdzie SSSS - przyjęta w obliczeniach pakietu XDS3 nazwa macierzy oryginalnych cech stymulant. Zakłada się, że zbiór wynikowy pakietu XDS3 jest typu *LP. Brak miejsca nie pozwala na szczegółowe omówienie przygotowania danych do obliczeń pakietu XDS3. Informacje na ten temat znaleźć można np. w [1], [2]. Ograniczenia dla zbioru wyjściowego zawierającego macierz cech-stymulant: liczba cech $\leqslant 500$, liczba obiektów $\leqslant 50$.

2.2. Dane sterujące obliczeniami programu G1DØ

Dane te zapisywane są jako rekordy kart perforowanych 80-kolumnowych (jako rekordy zbioru typu *CR) i tworzą pewne bloki.

- Blok problemu

Blok ten ma postać jednego rekordu:
- kolumny 1-4 : słowo PROB
- kolumny 8-15 : ustalona przez użytkownika nazwa problemu TPROB
- kolumny 16-23 : ustalona przez użytkownika wewnętrzna nazwy wyjściowej taśmy magnetycznej TMUKL (lub zbioru typu *MT), na którą zapisywane będą parametry zadania oraz macierz wyników obserwacji
- kolumny 24-31: wewnętrzna nazwa wyjściowej taśmy magnetycznej TMMS (lub zbioru typu *MT), na którą zapisywane będą informacje o miernikach syntetycznych
- kolumny 32-39: wewnętrzna nazwa wyjściowej taśmy magnetycznej TMDE (lub zbioru typu *MT), na którą zapisywane będą informacje o dendrytach
- kolumna 43: liczba typu INTEGER sygnalizująca poziom druku wyników NDR
- kolumny 44-47: identyfikator szeregów czasowych - słowo TIME lub pole puste
- kolumna 51: identyfikator ID obliczeń macierzy przewag - znak różny od spacji lub spacja.

Nazwy TPROB, TMUKL, TMMS, TMDE, nie mogą być ciągiem spacji. Rodzaj informacji drukowanych w zależności od wartości NDR wyjaśniony będzie w paragrafie „Tabulogram wynikowy".
Jeśli w polu 44-47 występuje słowo TIME, program traktuje szeregi wartości zmiennych jako szeregi czasowe (a obiekty jako momenty czasu). Jeśli w kolumnie 51 występuje znak różny od spacji, oznacza to, że obok obliczeń dendrytów na podstawie macierzy zróżnicowania, deklaruje się obliczenia dendrytów na podstawie tzw. macierzy przewag. Sprawy te wyjaśnimy szczegółoweij w dalszej części pracy.

Przykład

PROB---ANALIZA$_{\nabla}$PARAM$_{\nabla\nabla\nabla}$MSYNTE$_{\nabla\nabla}$DENDRYT$_{\nabla}$--- 3TIME---1

Nazwa problemu: ANALIZA$_{\nabla}$, wewnętrzna nazwa zbioru wyjściowego z parametrami zadania PARAM$_{\nabla\nabla\nabla}$, wewnętrzna nazwa zbioru wyjściowego zawierającego mierniki syntetyczne MSYNTE$_{\nabla\nabla}$, wewnętrzna nazwa zbioru wyjściowego zawierającego dendryty DENDRYT$_{\nabla}$, poziom druku wyników NDR=3, dane należy traktować jako szeregi czasowe (słowo TIME). Oprócz dendrytów wyznaczonych na podstawie macierzy odległości wyznaczone będą dendryty na podstawie macierzy przewag, gdyż identyfikator ID=1 (nie jest spacją).

- Blok nazw cech - stymulant

Blok ten ma następującą postać:

* 1-y rekord: kolumny 1-4 słowo TCOL
* następne rekordy:
  - kolumny 1-4: słowo $_{\nabla}$COL
  - kolejne pola 4-znakowe od pola 8-11 aż do pola 64-67 (tzn. pola 8-11, 12-15, 16-19, itd...) zawierają co najwyżej czteroznakowe nazwy kolejnych zmiennych w wejściowej macierzy cech - stymulant SSSS.

UWAGA (!)

W danym rekordzie można umieścić od Ø do 15 nazw. Nazwy nie mogą się powtarzać. Pierwsze pole puste (4 spacje) spośród pól przewidzianych na nazwy traktowane jest jako koniec listy nazw w danym rekordzie.

Liczba cech wymienionych w bloku TCOL musi być równa liczbie cech w macierzy wydrukowanej przez pakiet XDS3 instrukcją POBS,66,SSSS. Pierwsza nazwa w bloku TCOL traktowana jest jako nazwa pierwszej cechy w macierzy SSSS, druga - jako nazwa drugiej cechy w macierzy SSSS itd. Program G1DO używa nazw cech deklarowanych w bloku TCOL, stąd nazwy cech w bloku TCOL nie muszą być identyczne z nazwami używanymi przy tworzeniu macierzy cech-stymulant za pomocą pakietu XDS3.

Ograniczenia dla liczby nazw LZ w bloku TCOL: $1 \leqslant LZ \leqslant 500$.

Przykład

TCOL

$_{\nabla}$COL---ZM1$_{\nabla}$ZM2$_{\nabla}$XYF$_{\nabla}$47A$_{\nabla}$Z1O$_{\nabla}$

$_{\nabla}$COL---X1$_{\nabla\nabla}$X2$_{\nabla\nabla}$X47$_{\nabla\nabla}$XXY

Rozpatruje się 9 cech stymulant nazwanych ZM1$_{\nabla}$, ZM2$_{\nabla}$, XYF$_{\nabla}$, 47A$_{\nabla}$, Z1Ø$_{\nabla}$, X1$_{\nabla\nabla}$, X2$_{\nabla\nabla}$, X47$_{\nabla}$, $_{\nabla}$XXY. Wartości pierwszej cechy w macierzy wyników obserwacji cech-stymulant przyporządkowane będą cesze ZM1, wartości drugiej - cesze ZM2, wartości czwartej - cesze 47A, wartości dziewiątej - - cesze $_{\nabla}$XXY.

- Blok nazw obiektów

* 1-y rekord: kolumny 1-4 słowo TROW

* następne rekordy:
  - kolumny 1-4: słowo $_{\vee}$ROW
  - kolejne pola czteroznakowe od pola 8-11 do pola 64-67 zawierają co najwyżej 4-znakowe nazwy kolejnych obiektów.

Wymagania co do umieszczenia nazw opisano w Uwadze (1).
Liczba obiektów wymienionych w bloku TROW musi być równa liczbie obiektów w macierzy tworzonej przez pakiet XDS3 instrukcją POBS, 66,SSSS. Pierwsza nazwa w bloku TROW traktowana jest jako nazwa pierwszego obiektu (nazwa pierwszego wiersza macierzy cech-stymulant), druga - jako nazwa drugiego obiektu (nazwa drugiego wiersza macierzy cech-stymulant) itd. Nazwy używane w bloku TROW mogą odbiegać od nazw obiektów używanych przy tworzeniu macierzy cech-stymulant za pomocą pakietu XDS3.

Ograniczenia dla liczby nazw LTR w bloku TROW: $2 \leqslant LTR \leqslant 5Ø$.

Przykład
TROW
$_{\vee}$ROW---B1$_{\vee\vee}$B2$_{\vee\vee}$B3$_{\vee\vee}$
$_{\vee}$ROW---B4$_{\vee\vee}$ B5$_{\vee\vee}$B6$_{\vee\vee}$B7AA1B1Ø
W zadaniu występuje 8 obiektów B1$_{\vee\vee}$, B2$_{\vee\vee}$, B3$_{\vee\vee}$, B4$_{\vee\vee}$, B5$_{\vee\vee}$, B6$_{\vee\vee}$, B7AA, 1B10. Pierwszy wiersz macierzy wyników obserwacji cech-stymulant zawiera wartości stymulant w obiekcie B1, drugi wiersz - wartości stymulant w obiekcie B2, ósmy wiersz - wartości stymulant w obiekcie 1B1Ø.

● Bloki dziedzin i sfer

W programie przyjmuje się dwuszczeblowy podział zbioru cech na jedną bądź więcej tzw. dziedzin, a w ramach danej dziedziny - na jedną lub więcej tzw. sfer. Sfera składa się z jednej lub kilku cech.
Struktura bloku dla danej dziedziny jest następująca:
* 1-y rekord dla dziedziny:
  - kolumny 1-4: słowo UKLA
  - kolumny 7-11: co najwyżej czteroznakowa nazwa dziedziny.

Następnie występują podbloki danych określających sfery danej dziedziny:
** 1-y rekord podbloku dla danej sfery
  - kolumny 1-4: słowo SFER
  - kolumny 7-11: co najwyżej czteroznakowa nazwa sfery

** następne rekordy podbloku dla danej sfery

Następne rekordy zawierają nazwy cech tworzących daną sferę. Rekordy te opracowywane są tak jak rekordy określające nazwy cech w bloku TCOL. Pojedynczy (pierwszy lub kontynuacyjny) rekord ma następującą strukturę:
  - kolumny 1-4: słowo $_{\vee}$COL
  - kolejne pola 4-znakowe: począwszy od pola 8-11 aż do pola 64-67 zawierają nazwy cech tworzących sferę.

Wymagania co do sposobu umieszczania nazw określono w Uwadze (1). Zbiór cech wymienionych jako cechy tworzące wszystkie dziedziny i sfery nie mu-

si wyczerpywać całej listy cech podanej w bloku TCOL. Wśród nazw cech tworzących sfery i dziedziny nie mogą wystąpić nazwy nie wymienione w bloku TCOL. Nazwa cechy podanej dla danej sfery nie może być powtórzona wśród cech danej sfery lub innych sfer. Nazwy dziedzin oraz sfer nie mogą się powtarzać.

Ograniczenia:

- liczba dziedzin LUK, $1 \leqslant LUK \leqslant 10$
- liczba wszystkich sfer LSF, $1 \leqslant LSF \leqslant 50$
- liczba sfer w dziedzinie LSFU, $1 \leqslant LSFU \leqslant 50$
- liczba sfer we wszystkich sferach LZU, $1 \leqslant LZU \leqslant 400$
- liczba cech w danej sferze LZS, $1 \leqslant LZS \leqslant 400$

Uwaga: Jeśli obliczenia mają być wykonane dla jednej macierzy cech-stymulant (np. macierzy „ogółem"), to należy zadeklarować jedną dziedzinę i jedną sferę.

Przykład

UKLA---UKL1
SFER---$SF1_{\nabla}$
$_{\nabla}$COL---$ZM1_{\nabla}$ XYF_{∇} $_{\nabla}XXY$
SFER---$SF2_{\nabla}$
$_{\nabla}$COL---$ZM2_{\nabla}$
$_{\nabla}$COL---$47A_{\nabla}$
UKLA---$UK2_{\nabla}$
SFER---SF1Ø
$_{\nabla}$COL---$Z1Ø_{\nabla}$ $X47_{\nabla}$

Oznacza to, że zadeklarowano dwie dziedziny UKL1, $UK2_{\nabla}$. Dziedzina UKL1 dzieli się na dwie sfery $SF1_{\nabla}$ (zawierającą cechy $ZM1_{\nabla}$, XYF_{∇}, $_{\nabla}XXY$) oraz SF2 (zawierającą cechy $ZM2_{\nabla}$, $47A_{\nabla}$)..Dziedzina $UK2_{\nabla}$ składa się z jednej sfery SF1Ø zwierającej cechy $Z1Ø_{\nabla}$, $X47_{\nabla}$. Spośród cech wymienionych w bloku TCOL (zob. odpowiedni przykład) cech X1, X2 nie przypisano do żadnej sfery.

● Blok typu normalizacji cech

Program umożliwia obliczenia dla jednej bądź więcej reguł normalizacji cech, wybranych przez użytkownika z listy dwudziestu reguł normalizacji.

* 1-y rekord:
  - kolumny 1-4: słowo NORM
* następne rekordy
  - kolumny 1-4: słowo $_{\nabla}$TNO
  - kolejne pola czteroznakowe począwszy od pola 8-11 aż do pola 64-67 zawierają justowane do lewej kody typów normalizacji.

Wymagania co do sposobu umieszczania nazw określa Uwaga (!). Niech y_{ij} oznacza oryginalną wartość j-ej cechy-stymulanty zaobserwowaną w obiekcie i-tym, a x_{ij} niech oznacza znormalizowaną wartość tej cechy w obiekcie i-tym. Niech dalej M_j jest wartością maksymalną, m_j jest wartością minimalną, R_j jest rozstępem, $\bar{y}_j$ jest średnią arytmetyczną, s_j jest odchyleniem standardowym j-ej oryginalnej cechy-stymulanty.

W obliczeniach za pomocą programu G1DO można wykorzystać jedną bądź więcej normalizacji o ogólnej postaci

$$x_{ij} = \frac{y_{ij} - a_j}{b_j}$$

gdzie a_j=Ø, m_j, $\bar{y}_j$, M_j natomiast b_j=M_j,m_j,R_j,$\bar{y}_j$,s_j.
Kody typów normalizacji w zależności od a_j, b_j są następujące

b_j \ a_j	Ø	m_j	$\bar{y}_j$	M_j
M_j	$1MA_\nabla$	$2MA_\nabla$	$3MA_\nabla$	$4MA_\nabla$
m_j	$1MI_\nabla$	$2MI_\nabla$	$3MI_\nabla$	$4MI_\nabla$
R_j	$1RO_\nabla$	$2RO_\nabla$	$3RO_\nabla$	$4RO_\nabla$
$\bar{y}_j$	$1SR_\nabla$	$2SR_\nabla$	$3SR_\nabla$	$4SR_\nabla$
s_j	$1ST_\nabla$	$2ST_\nabla$	$3ST_\nabla$	$4ST_\nabla$

Np. kod $1MA_\nabla$ odpowiada normalizacji $x_{ij} = y_{ij}/M_j$, a kod 3ST odpowiada normalizacji $x_{ij} = \frac{y_{ij} - \bar{y}_j}{s_j}$.

Przykład

NORM
$_\nabla$TNO---$1MA_\nabla 3MI_\nabla$
$_\nabla$TNO---$3ST_\nabla$
Rozpatruje się trzy rodzaje normalizacji cech-stymulant:
$1MA_\nabla$, $3MI_\nabla$, $3ST_\nabla$.

- Blok wzorców cech-stymulant

Program umożliwia obliczenie mierników syntetycznych bezwzorcowych oraz wzorcowych.

Niech y_{oj} oznacza wzorcową wartość oryginalnej cechy-stymulanty. Możliwe są następujące typy wzorców:

a) wzorzec kodowany symbolem MAX_∇

$$y_{oj} = \max_{1 \leqslant i \leqslant m} y_{ij} \qquad (i=1,\ldots,m \text{ - numery obiektów})$$

b) wzorzec kodowany symbolem MIN_∇

$$y_{oj} = \min_{1 \leqslant i \leqslant m} y_{ij}$$

c) wzorcem jest obiekt (i_*) wybrany przez użytkownika

$$y_{oj} = y_{i_* j}$$

Ten typ wzorca koduje się nazwą obiektu (np. B7AA, jeśli wzorcem jest obiekt nazwany B7AA)

d) wzorzec o wartościach ustalonych przez użytkownika. Wzorzec ten kodowany jest czteroznakową nazwą, z której dwa pierwsze znaki tworzą ciąg US.

Wzorcowe wartości y_{oj} typu a,b,c program ustala automatycznie. Rekordy bloku typu wzorca opracowywane są w następującej formie:

* 1-y rekord:
  - kolumny 1-4: słowo WZOR
* następne rekordy (rekordy nazw wzorców)
  - kolumna 1-4: słowo $_{\nabla}$TWZ
  - kolejne pola czteroznakowe począwszy od pola 8-11 zawierają justowane od lewej czteroznakowe kody wzorców.

Wymagania co do sposobu umieszczania nazw podano w Uwadze (!). W danym przebiegu można zadeklarować obliczenia bez wzorca lub z wzorcem (jednym bądź więcej). Jeśli obliczenia będą dokonywane bez wzorców, to w rekordach nazw wzorców $_{\nabla}$TWZ nie może wystąpić żaden różny od spacji znak w kolumnach od 8 do 72.

* rekordy wartości wzorców ustalonych typu US

Jeśli w rekordach nazw wzorców wymieniono kody wzorców ustalonych, to bezpośrednio po kompletnej sekwencji rekordów TWZ muszą wystąpić rekordy zawierające wartości wzorców ustalonych.

Dla danego wzorca ustalonego rekordy danych są następujące:

- 1-y rekord:
  - kolumny 1-4: słowo WWZU
  - kolumny 8-11: justowana do lewej nazwa wzorca ustalonego
- następne rekordy

Następne rekordy zawierają wzorcowe wartości kolejnych oryginalnych cech-stymulant przy rozpatrywanym wzorcu ustalonym. Wartości te należy podawać jako liczby stałoprzecinkowe. Umieszcza się je co najwyżej do kolumny 80. Zakłada się, że kolejność liczb odpowiada kolejności pojawiania się cech w bloku dziedzin i sfer.

Ograniczenia:
Liczba wzorców nie może przekraczać 9. Liczba wzorców ustalonych nie może przekraczać 5.

Przykład

```
WZOR
 ▽TWZ---MAX▽ MIN ▽ B7AAUS2▽
 ▽TWZ---US1▽
 WWZU---US1▽
     7.Ø 3.2 1.3 4.9
     2.1 1.Ø Ø
 WWZU---US2▽
     3.2 4.8 2.1 3.6 2.8 5.2 9.1
```

Deklarowano 5 wzorców: MAX_{∇}, MIN_{∇}, B7AA, $US2_{\nabla}$, $US1_{\nabla}$. Wzorzec B7AA oznacza, że za wzorcowe wartości cech oryginalnych wzięte będą wartości występujące dla obiektu B7AA. Deklarowano dwa wzorce ustalone $US1_{\nabla}$, $US2_{\nabla}$. Wzorcowe wartości kolejnych cech oryginalnych wg wzorca $US1_{\nabla}$ są następujące: 7.Ø, 3.2, 1.3, 4.9, 2.1, 1.Ø, Ø. Wobec kolejności cech deklarowanej w bloku dziedzin i sfer (zob. odpowiedni przykład) wartości te związane są z cechami odpowiednio: $ZM1_{\nabla}$, XYF_{∇} , $_{\nabla}XXY$, $ZM2_{\nabla}$, $47A_{\nabla}$, $Z1Ø_{\nabla}$, $X47_{\nabla}$.

• Blok typów miernika syntetycznego

* 1-y rekord:

- kolumny 1-4: słowo MSYN

* następne rekordy

W następnych rekordach podaje się nazwy mierników syntetycznych. Pojedynczy rekord ma postać:

- kolumny 1-4: słowo $_{\nabla}TMS$
- kolejne pola czteroznakowe począwszy od pola 8-11 zawierają nazwy mierników syntetycznych.

Wymagania co, do trybu umieszczania nazw podaje Uwaga(!).

W bloku typów miernika syntetycznego można zadeklarować do 9 typów miernika syntetycznego. Jeśli nie żąda się obliczeń mierników syntetycznych, to w rekordach $_{\nabla}TMS$ w kolumnach 8-72 muszą wystąpić tylko spacje.

Kody mierników syntetycznych

Niech N oznacza liczbę cech tworzących dany układ cech (dziedzinę bądź sferę). Niech dalej x_{oj} oznacza znormalizowaną wartość wzorcową cechy j-ej. Wartości x_{oj} wyznaczane są przez program na podstawie wartości oryginalnych y_{oj} wg reguły normalizacji deklarowanej dla cech. Oznaczmy jeszcze

$$h_{ij} = \begin{cases} -1 & \text{jeśli } x_{ij} < x_{oj} \\ 1 & \text{jeśli } x_{ij} \geqslant x_{oj} \end{cases}$$

W programie można zadeklarować obliczanie jednego bądź więcej mierników syntetycznych. Kody możliwych mierników i ich definicje są następujące:

1) miernik bezwzorcowy SAB_{∇} (średnia arytmetyczna bezwzorcowa)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} x_{ij}$$

2) miernik wzorcowy LOB_{∇} (liniowe odchylenie bezwzględne)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} h_{ij} \, | x_{ij} - x_{oj} |$$

3) miernik wzorcowy KOB_v (kwadratowe odchylenie bezwzględne)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} h_{ij} (x_{ij} - x_{oj})^2$$

4) miernik wzorcowy POB (pierwiastkowe odchylenie bezwzględne)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} h_{ij} \sqrt{|x_{ij} - x_{oj}|}$$

5) miernik wzorcowy WOB_v (wykładnicze odchylenie bezwzględne)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} h_{ij} (\exp |x_{ij} - x_{oj}| - 1)$$

6) miernik wzorcowy LOW_v (liniowe odchylenie względne)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} h_{ij} z_{ij}$$

gdzie

$$z_{ij} = \frac{|x_{ij} - x_{oj}|}{|x_{oj}|}$$

7) miernik wzorcowy KOW (kwadratowe odchylenie względne)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} h_{ij} (z_{ij})^2$$

8) miernik wzorcowy POW (pierwiastkowe odchylenie względne)

$$M_i = \frac{1}{N} \sum_{j=1}^{N} h_{ij} \sqrt{z_{ij}}$$

9) miernik wzorcowy WOW_v (wykładnicze odchylenie względne)

$$M_j = \frac{1}{N} \sum_{j=1}^{N} h_{ij} (\exp z_{ij} - 1)$$

Mierniki LOW_v, KOW_v, POW_v, WOW_v obliczane są tylko wtedy, gdy wszystkie wartości wzorcowe x_{oj} różnią się od zera i są tego samego znaku, a ponadto, gdy wszystkie wartości x_{ij} (i=1,...,m; j=1,...,N) są tego samego znaku.

Przedstawione mierniki wzorcowe osiągają wartość 0, gdy średnio rzecz biorąc obiekt i-ty jest tak samo rozwinięty jak pseudo-obiekt wzorcowy (w szczególności $M_i = 0$, jeśli dla j=1,...,N jest $x_{ij} = x_{oj}$). Jeśli $M_i > 0$, to obiekt i-ty jest średnio rzecz biorąc lepiej rozwinięty od obiektu-wzorca, a jeśli $M_i < 0$, to obiekt i-ty jest średnio rzecz biorąc gorzej rozwinięty od obiektu-wzorca.

Przykład

MSYN
$_{\nabla}$TMS---SAB$_{\nabla}$KOB$_{\nabla}$
$_{\nabla}$TMS---LOW$_{\nabla}$
Zadeklarowano obliczanie miernika bezwzorcowego SAB$_{\nabla}$ oraz mierników wzorcowych KOB$_{\nabla}$, LOW$_{\nabla}$.

● Blok typów zróżnicowania obiektów

Program umożliwia obliczenie dendrytów przy różnych miernikach zróżnicowania obiektów (miernikami tymi najczęściej są odległości). Struktura rekordów deklarujących mierniki zróżnicowania jest następująca:

* 1-y rekord:
  - kolumny 1-4: słowo DEND
* następne rekordy:

W następnych rekordach podaje się kody mierników zróżnicowania, przy których należy wyznaczać dendryty wrocławskie.
Pojedynczy rekord ma postać:
- kolumny 1-4: słowo $_{\nabla}$TOD
- kolejne pola 4-znakowe począwszy od pola 8-11 zawierają justowane do lewej kody mierników zróżnicowania.

Sposób umieszczania nazw określa Uwaga (!).

W programie można skorzystać z jednego lub więcej mierników zróżnicowania. Jeśli nie zakłada się obliczenia dendrytów, to w rekordach TOD na pozycjach od 7 do 68 muszą wystąpić same spacje. Dopuszczalne mierniki zróżnicowania i ich kody są następujące:

1) zróżnicowanie wg odległości bezwzględnej (miejskiej) BEZ$_{\nabla}$

$$d_{ik} = \frac{1}{N} \sum_{j=1}^{N} |x_{ij} - x_{kj}|$$

2) zróżnicowanie wg odległości euklidesowej EUK

$$d_{ik} = \frac{1}{N} \sqrt{\sum_{j=1}^{N} (x_{ij} - x_{kj})^2}$$

3) zróżnicowanie pierwiastkowe SQR_v

$$d_{ik} = \frac{1}{N}\left(\sum_{j=1}^{N} \sqrt{|x_{ij} - x_{kj}|}\right)^2$$

4) zróżnicowanie wykładnicze EXP_v

$$d_{ik} = \ln\left(\frac{1}{N}\sum_{j=1}^{N} \exp|x_{ij} - x_{kj}|\right)$$

5) zróżnicowanie wg różnicy kwadratów RKW_v

$$d_{ik} = \frac{1}{N}\sqrt{\sum_{j=1}^{N} |x_{ij}^2 - x_{kj}^2|}$$

6) zróżnicowanie wg różnic pierwiastków RSQ_v

$$d_{ik} = \frac{1}{N}\sum_{j=1}^{N}\left(\sqrt{|x_{ij}|} - \sqrt{|x_{kj}|}\right)^2$$

(jest to zmodyfikowana odległość Jeffreysa-Matusita).

7) zróżnicowanie wg różnic wykładniczych REX_v

$$d_{ik} = \ln\left(1 + \frac{1}{N}\sum_{j=1}^{N} |\exp|x_{ij}| - \exp|x_{kj}||\right)$$

8) zróżnicowanie wg wskaźnika Clarka CLA_v

$$d_{ik} = \sqrt{\sum_{j=1}^{N}\left(\frac{x_{ij} - x_{kj}}{x_{ij} + x_{kj}}\right)^2}$$

9) zróżnicowanie wg wskaźnika Canberra CAN_v

$$d_{ik} = \sum_{j=1}^{N}\left|\frac{x_{ij} - x_{kj}}{x_{ij} + x_{kj}}\right|$$

Mierniki RKW_v , RSQ_v , REX_v , CLA_v , CAN_v obliczane są tylko wtedy, gdy wszystkie x_{ij} $(i=1,\ldots,m;\ j=1,\ldots,N)$ są bądź nieujemne, bądź niedodatnie. Ponadto wskaźniki CLA_v , CAN_v obliczane są tylko na podstawie takich j, że $x_{ij} + x_{kj} \neq 0$.

Przykład

DEND

$_{\nabla}$TOD---BEZ$_{\nabla}$ EUK$_{\nabla}$

$_{\nabla}$TOD---SQR$_{\nabla}$ CAN$_{\nabla}$

Oznacza to, że deklarowano obliczanie dendrytów przy czterech definicjach zróżnicowania obiektów: BEZ$_{\nabla}$, EUK$_{\nabla}$, SQR$_{\nabla}$, CAN$_{\nabla}$.

„Przewagi" obiektów

W programie uwzględniono możliwość wyznaczania dendrytów na podstawie tzw. macierzy „przewag" obiektów. Dany typ przewagi związany jest z danym typem zróżnicowania. Po wyznaczeniu dendrytu dla danego typu zróżnicowania przechodzi się do obliczeń odpowiedniej macierzy przewag obiektów i konstruowania na tej podstawie dendrytów, jeśli parametr ID w bloku problemu PROB jest różny od spacji.

Element p_{ik} macierzy przewag określa „przewagę" obiektu i nad obiektem k. Jeśli $p_{ik} > 0$, to średnio rzecz biorąc obiekt i jest lepiej rozwinięty od obiektu k. Jeśli $p_{ik} < 0$, to średnio rzecz biorąc obiekt i jest gorzej rozwinięty od obiektu k. Przypadek p_{ik} oznacza, iż oba obiekty są podobnie rozwinięte.

Niech

$$h_{ik}^{j} = \begin{cases} 1 & \text{jeśli} \quad x_{ij} \geqslant x_{kj} \\ -1 & \text{jeśli} \quad x_{ij} < x_{kj} \end{cases}$$

Niech dalej W_{ik} oznacza pewną wartość, natomiast

$$z_{ik} = \begin{cases} 1 & \text{jeśli} \quad W_{ik} \geqslant 0 \\ -1 & \text{jeśli} \quad W_{ik} < 0 \end{cases}$$

Wartości p_{ik} zdefiniowano w następujący sposób:

y zróżnicowaniu BEZ

$$p_{ik} = \frac{1}{N} \sum_{j=1}^{N} h_{ik}^{j} \left| x_{ij} - x_{kj} \right|$$

2) przy zróżnicowaniu EUK

$$P_{ik} = \frac{1}{N} \quad z_{ik} \quad \sqrt{|W_{ik}|}, \quad \text{gdzie} \quad W_{ik} = \sum_{j=1}^{N} h_{ik}^{j} (x_{ij} - x_{kj})^2$$

3) przy zróżnicowaniu SQR

$$p_{ik} = \frac{1}{N} \quad z_{ik} \quad W_{ik}^2 , \qquad \text{gdzie} \quad W_{ik} = \sum_{j=1}^{N} h_{ik}^{j} \sqrt{|x_{ij} - x_{kj}|}$$

4) przy zróżnicowaniu EXP

$$p_{ik} = z_{ik} \ln |W_{ik}| \quad , \quad \text{gdzie} \quad W_{ik} = \frac{1}{N} \sum_{j=1}^{N} h_{ik}^{j} \exp |x_{ij} - x_{kj}|$$

5) przy zróżnicowaniu RKW

$$p_{ik} = \frac{1}{N} z_{ik} \sqrt{|W_{ik}|} \quad , \quad \text{gdzie} \quad W_{ik} = \sum_{j=1}^{N} h_{ik}^{j} |x_{ij}^{2} - x_{kj}^{2}|$$

6) przy zróżnicowaniu RSQ

$$p_{ik} = \frac{1}{N} \sum_{j=1}^{N} h_{ik}^{j} \left(|x_{ij}| - |x_{kj}| \right)^{2}$$

7) przy zróżnicowaniu REX

$$p_{ik} = z_{ik} \ln |W_{ik}| \quad , \quad \text{gdzie} \quad W_{ik} = 1 + \frac{1}{N} \sum_{j=1}^{N} h_{ik}^{j} (\exp|x_{ij}| - \exp|x_{kj}|)$$

8) przy zróżnicowaniu CLA

$$p_{ik} = z_{ik} \sqrt{|W_{ik}|} \quad , \text{gdzie} \quad W_{ik} = \sum_{j=1}^{N} h_{ik}^{j} \left(\frac{x_{ij} - x_{kj}}{x_{ij} + x_{kj}} \right)^{2}$$

9) przy zróżnicowaniu CAN

$$p_{ik} = \sum_{j=1}^{N} h_{ik}^{j} \left| \frac{x_{ij} - x_{kj}}{x_{ij} + x_{kj}} \right|$$

Macierz przewag $[p_{ik}]$ i,k=1,...,m zawiera elementy dodatnie i ujemne, przy czym jest to macierz symetryczna względem modułów, bowiem $p_{ik} = -p_{ki}$. Dotyczące macierzy przewag $[p_{ik}]$ dendryty wyznaczone są na podstawie macierzy modułów przewag.

Blok końca danych

- kolumny 1-3: słowo GET.

2.3. Struktura obliczeń generowanych przez dane sterujące

Obliczenia wykonywane przez program G1DO mają strukturę „pętlową". Najbardziej zewnętrzną pętlą jest pętla dotycząca typu normalizacji. Wewnątrz pętli dla danego typu normalizacji znajduje się pętla dla typu wzorca , a wewnątrz pętli typu wzorca znajduje się pętla typu miernika

syntetycznego. Po wykonaniu obliczeń dla wszystkich mierników syntetycznych i wszystkich wzorców przy danej regule normalizacji, program przechodzi do wykonania pętli dotyczących obliczeń dendrytów przy deklarowanych typach zróżnicowania i przewag.

3. Instrukcja operatorska dla pracy pod kontrolą George-3

```
LO ▽PROGRAM▽ G1DØ
AS ▽#CR,DANE
AS ▽#TR,MATR
AS ▽#LP,WYNI
LF ▽ WYNI #LP,PA
AS ▽#MT10,PARA(WRITE)
AS ▽#MT11,MSYNT(WRITE)
AS ▽#MT12,DENDR(WRITE)
AS ▽#MT7,ROB7(WRITE)
AS ▽#MT8,ROB8(WRITE)
TI ▽nnnMINS
EN ▽Ø
```

gdzie: DANE - nazwa zbioru zawierającego dane sterujące (zob. paragraf 2.2)

MATR - nazwa zbioru zawierającego macierz oryginalnych wyników obserwacji cech-stymulant (jest to nazwa zbioru wynikowego z obliczeń pakietem XDS3 - zob. paragraf 2.1)

WYNI - nazwa zbioru wynikowego

PARA - zewnętrzna nazwa zbioru wyjściowego, dla którego zapisywane będą parametry zadania oraz macierz oryginalnych wartości cech tworzących poszczególne dziedziny

MSYNT - zewnętrzna nazwa zbioru wyjściowego, do którego zapisywane będą wartości miernika, syntetycznego

DENDR - zewnętrzna nazwa zbioru wyjściowego, do którego zapisywane będą dendryty

ROB7, ROB8 - nazwy zbiorów roboczych

nnn - przewidywany czas zajęcia procesora w minutach (np. nnn=1Ø).

Nazwy zbiorów są alfanumeryczne co najwyżej 12-znakowe. Nazwy PARA, MSYNT, DENDR niekoniecznie muszą być identyczne z odpowiednimi nazwami wewnętrznymi TMUKL (parametry zadania), TMMS (mierniki syntetyczne), TMDE (dendryty) deklarowanymi w bloku problemu. Identyczność odpowiednich nazw (np. PARA=TMUKL) jest jednak wygodna dla rozpoznania zbiorów.

4. Tabulogram wynikowy

Rodzaj informacji wyprowadzanych na tabulogram (do zbioru typu #LP) zależy od wartości parametru druku NDR. Zakłada się, że w bloku PROB wartość NDR=Ø,1,2,3. Jeśli deklarowana wartość jest spoza tego zakresu, to program ustala NDR=3.

Program drukuje m.in.

a) Macierz obserwacji pierwotnych cech stymulant oraz oryginalne wartości wzorów (jeśli NDR $\geqslant$).

b) Macierz znormalizowanych wartości cech i znormalizowanych wartości wzorców (NDR $\geqslant$ 2).

c) Informacje dotyczące mierników syntetycznych przy ustalonej sferze (dziedzinie), ustalonej normalizacji, ustalonym wzorcu:
   1. Tablica wartości mierników systetycznych oraz miejsce obiektu wg danego miernika syntetycznego (NDR $\geqslant$ Ø).
   2. Macierz przewag wg miernika syntetycznego. Elementem p_{ij} tej macierzy jest różnica $M_j - M_i$ (NDR $\geqslant$ 2).
   3. Różnice wartości mierników syntetycznych dla obiektów sąsiednich. Wydruk jest realizowany, gdy w bloku problemu w kolumnach 44-47 podano słowo TIME (NDR $\geqslant$ 1).

d) Informacje dotyczące zróżnicowań, przewag i dendrytów.
   1. Macierz zróżnicowań obiektów (macierz przewag obiektów) oraz rangi danego obiektu względem innych. Obiekt najmniej różniący się od danego ma rangę 1 (NDR $\geqslant$ 2).
   2. Zróżnicowanie (przewagi) dla obiektów sąsiednich. Wydruk ma miejsce, gdy w bloku problemu w kolumnach 44-47 występuje słowo TIME (NDR $\geqslant$ 1).
   3. Macierz modułów przewag (NDR=3).
   4. Dendryty wrocławskie konstruowane na macierzy zróżnicowania lub macierzy modułów przewag. W ostatnim przypadku dla krawędzi $\langle i,k \rangle$ podaje się m.in. identyfikator IDE(POCZ) obiektu początkowego krawędzi i. Jeśli IDE=1, to przewaga $p_{ik} \geqslant 0$, a jeśli IDE = -1, to $p_{ik} < 0$. (NDR $\geqslant$ 0),

Program wyprowadza jeszcze wiele innych informacji.

Uwaga: Na tabulogramie słowo UKLAD jest synonimem DZIEDZINY.
Dane dotyczące sfery sygnalizowane są przez SFERA =␣, UKLAD =␣, a dane dotyczące dziedziny sygnalizowane są przez SFERA = OGOL, UKLAD = ␣, gdzie ␣ - odpowiednia nazwa.

5. Informacja o programie

- nazwa programu: PROGRAM ▽G1DØ
- język: FORTRAN 1900
- program skompilowano za pomocą kompilatora XFIV w trybie adresacji 22 AM DRM przy poziomie śledzenia TRACE Ø
- pamięć 24K
- długość: ok. 19000 (około 2800 instrukcji)
- program G1DØ jest programem nakładanym.