Matematyczne aspekty modelowania pajęczynowego obiektów

A. AMELJAŃCZYK
e-mail:aameljanczyk@wat.edu.pl

Instytut Systemów Informatycznych
Wydział Cybernetyki WAT
ul. S. Kaliskiego 2, 00-908 Warszawa

W pracy omówiono możliwości wykorzystania tzw. metod modelowania pajęczynowego do modelowania i analizy jakościowej obiektów złożonych w procesach eksploracji danych. Zdefiniowano takie pojęcia jak: N-wymiarowa przestrzeń pajęczynowa, model pajęczynowy obiektu, adekwatność modelu oraz niektóre charakterystyki eksploracyjne modelu.

Słowa kluczowe: eksploracja danych, matematyczny model obiektu, modelowanie pajęczynowe, model pajęczynowy obiektu, model M-dokładny, dokładność modelu, adekwatność modelu.

1. Wprowadzenie

Matematyczne modele obiektów (systemów) służą najczęściej analizom jakościowym i porównawczym. Mogą być też podstawą budowy modeli optymalizacyjnych jak również podstawą definiowania funkcji rankingowych [1], [3]. Formalne modele opisowe znalazły też duże zastosowania w procesach i metodach eksploracji danych [5], [6], [7]. Obszarem szczególnie interesującym z tego punktu widzenia jest modelowanie i analiza danych medycznych, modelowanie stanu zdrowia pacjenta i tzw. jednostek chorobowych, wykorzystywane w komputerowych systemach wspomagania decyzji medycznych [2], [3], [9].

W procesie modelowania matematycznego obiektów bardzo ważną rolę odgrywa cel modelowania [4], [5]. Z celu modelowania wynikają przede wszystkim wymogi dotyczące konkretnego, formalnego języka modelowania oraz wymagana dokładność (adekwatność) modelu. Dysponując oryginałem obiektu o nazwie $x \in X$, w zależności od celu modelowania, możemy zbudować wiele różnych modeli $y = F(x)$, tego samego obiektu.

Zbiór nazw modelowanych obiektów, na ogół jest utożsamiane z pewnym podzbiorem X zbioru liczb naturalnych $\aleph$. Zbiór X jest wtedy zbiorem numerów obiektów. W szczególnych przypadkach zbiór X nie musi być skończonym podzbiorem liczb naturalnych. Takim przykładem może być zbiór możliwych „zestaw danych medycznych” [2], [3], [9].

Modelowany obiekt $x \in X$ charakteryzuje się na ogół wieloma różnymi cechami cząstkowymi, które decydują o jego globalnych (systemowych) własnościach. Szczegółowość (adekwatność) modelu – jakkolwiek definiowana jest oczywiście funkcją liczby cech (własności), które zostały uwzględnione w modelu. Im więcej cech uwzględnimy, tym na ogół model będzie bardziej dokładny (szczegółowy) i tym lepiej będzie „odzwierciedlał” oryginał. Taki model zapewne będzie jednak bardziej kosztowny, skomplikowany oraz bardziej niewygodny do prowadzenia badań i analizy. Problem ustalenia ile cech i które z nich należy wziąć pod uwagę jest oczywiście problemem wyboru odpowiedniego kompromisu między dokładnością modelu a jego złożonością i kosztem. Intuicyjnie można zgodzić się z opinią, że w każdym procesie modelowania istnieje pewna graniczna liczba cech do uwzględnienia, powyżej której, przy zadanym celu modelowania przyrost „jakości modelu” jest pomijalnie mały.

2. Uproszczony model opisowy obiektu $x \in X$

Oznaczmy symbolem x pewien obiekt ze zbioru X obiektów. Załóżmy dalej, że obiekt ten posiada maksymalnie N cech (własności), które mogą mieć znaczenie jeśli chodzi o adekwatność (jakość) modelu obiektu $x \in X$ z punktu widzenia przyjętego celu modelowania.

Załóżmy, że cechy świadczące o jakości obiektu są uporządkowane według ich ważności, zgodnie z porządkiem naturalnym.

Oczywiście wielkość liczby N zależy od „stopnia złożoności” obiektów ze zbioru X co zapiszemy $N = N(X)$. Jeśli przykładowo elementami zbioru X będą „proste obiekty" np. rodzaje bloczków budowlanych, gatunki serów, itp. to maksymalna liczba wyróżnionych cech może sięgać kilku.

Jeśli będą to natomiast złożone urządzenia elektroniczne czy też złożone mechanizmy, to liczba cech sięgać może przykładowo kilkunastu.

W przypadku modelowania bardzo złożonych obiektów jak np. stanu zdrowia pacjenta [3], [6] liczba ta może sięgać kilkuset i więcej.

Teoretycznie możemy rozpatrywać procesy modelowania gdzie liczba $N \to \infty$.
Symbolem $\aleph$ oznaczymy zbiór $\{1,\ldots,n,\ldots,N\}$ zaś symbolem $\mathcal{M}$ – zbiór $\{1,\ldots,m\ldots,M\} \subset \aleph$.

Określenie 2.1.

Modelem M-dokładnym $(M \le N)$ obiektu x nazwiemy model uwzględniający tylko M spośród N, najważniejszych cech obiektu $x \in X$.

Określenie 2.2.

Modelem dokładnym obiektu $x \in X$ nazwiemy model uwzględniający maksymalną liczbę N cech świadczących o jego jakości.

Określenie 2.3.

Funkcją modelowania opisowego nazywać będziemy funkcję $F: X \to \Re^N$, przyporządkowującą każdemu modelowanemu obiektowi $x \in X$ ciąg wartości poszczególnych jego cech. Tak więc modelem obiektu $x \in X$ będzie jego obraz

$$F(x) = (F_1(x),..,F_n(x),..,F_N(x)) \in \Re^N \quad (2.1)$$

dany funkcją modelowania F

gdzie $F_n(x)$ – wartość n-tej cechy modelowanego obiektu $x \in X$.

Załóżmy dalej, że dla każdego $n \in \aleph$

$$w_n \le F_n(x) \le w^n, \; x \in X \quad (2.2)$$

Fakt ten będziemy zapisywać:

$$y_n = F_n(x) \in \left[w_n, w^n\right] \subset \Re^1, n \in \aleph, x \in X \quad (2.3)$$

Ciągi liczbowe $y = (y_1,\ldots,y_n,\ldots,y_N) = F(x)$ nazywać też będziemy „danymi" o obiektach $x \in X$.

Niech przykładowo $N = 15$, $M = 9$. Modelem dokładnym obiektu $x \in X$ jest

$$F(x) = (1,2,1,1,3,4,8,7,2,5,4,3,1,0,0) \in \Re^{15}$$

zaś modelem M-dokładnym

$$F^M(x) = (1,2,1,1,3,4,8,7,2) \in \Re^9$$

Zadanie matematycznego modelowania opisowego obiektów x ze zbioru X możemy formalnie zapisać jako trójkę uporządkowaną:

$$Z_0 = (\Re^N, X, F) \quad (2.4)$$

Obraz zbioru Y zbioru X będzie zatem zbiorem modeli obiektów $x \in X$.

$$Y = F(X) = \left\{y = F(x) \in \Re^N \middle| x \in X\right\} \subset \Re^N \quad (2.5)$$

Klasą równoważności $X(y)$ obiektów ze względu na ustalone wartości cech $y \in Y$ nazywamy przeciwobraz zbioru jednoelementowego $\{y\} \subset Y$

$$X(y) = F^{-1}(\{y\}) = \{x \in X | F(x) = y\} \quad (2.6)$$

Jeśli w zbiorze klas równoważności $\{X(y) \subset X | y \in Y\}$, istnieje przynajmniej jedna klasa o liczności $|X(y)| > 1$, to zwiększenie ilości uwzględnianych cech może być przesłanką zwiększenia „dokładności modelowania". Pojawia się tutaj interesujący problem określenia takiej liczby N (o ile istnieje), że dla każdego $y \in Y$, $|X(y)| = 1$

3. N-wymiarowa przestrzeń pajęczynowa

Załóżmy, że dane jest zadanie modelowania opisowego: $Z_0 = (\Re^N, X, F)$, którego wynikiem jest zbiór $Y = F(X)$ modeli opisowych obiektów $x \in X$.

Możliwość interpretacji graficznej (tzw. „zobrazowania") takich modeli w celu analizy ich własności jest jednak bardzo ograniczona – maksymalnie do $N = 3$. O wiele większe możliwości graficznej interpretacji, a w konsekwencji analizy i „wydobywania" dodatkowej wiedzy z danych opisujących modelowane obiekty daje tzw. modelowanie pajęczynowe (web modelling).

Określenie 3.1.

N-wymiarową, ograniczoną przestrzenią pajęczynową nazywać będziemy parę uporządkowaną

$$P_r = (S_r, N)$$

gdzie S_r jest pewnym zbiorem wyznaczanym na płaszczyźnie, np. w układzie współrzędnych biegunowych w następujący sposób:

$$S_r = \{(d,\varphi) | \; 0 \le d \le r, 0 \le \varphi \le 2\Pi\} \quad (3.1)$$

Liczba $r > o$ zwana jest promieniem (zakresem) przestrzeni, zaś para (d,φ) to współrzędne punktu (elementu) zbioru S_r (d – odległość od tzw. bieguna, φ – kąt skierowany).

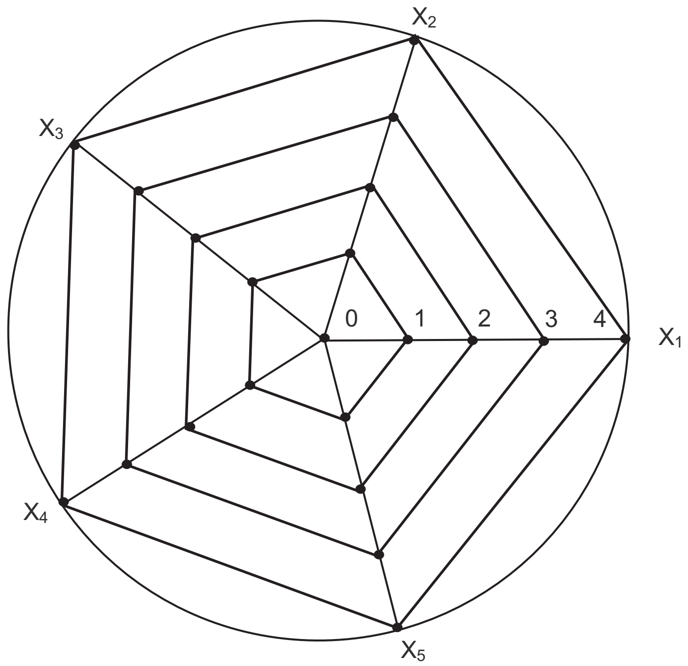
W przypadku gdy $r \to \infty$ otrzymamy N-wymiarową przestrzeń nieskończoną

$$P_\infty = (S_\infty, N).$$

Osie współrzędnych $0x_n, n = 1,\ldots, N$ przestrzeni pajęczynowej możemy wyznaczyć następująco:

$$0x_n = \left\{ (d, \beta_n) \middle| 0 \le d \le r, \beta_n = \frac{2\Pi(n-1)}{N} \right\},\ n \in \aleph \tag{3.2}$$

Na rys. 1. przedstawiona została pięciowymiarowa przestrzeń pajęczynowa z zaznaczoną „skalą odległości" (dla $r = 4$) oraz pięcioma „osiami współrzędnych" $0x_n$, $n \in \aleph = \{1, 2, 3, 4, 5\}$



Rys. 1. Pięciowymiarowa przestrzeń pajęczynowa o promieniu $r = 4$

Zbiór S_r bywa czasami definiowany wprost jako koło o promieniu r w układzie współrzędnych kartezjańskich:

$$S_r = \left\{ (s_1, s_2) \in \Re^2 \middle| s_1^2 + s_2^2 \le r^2 \right\} \tag{3.3}$$

Zamieniając współrzędne biegunowe na kartezjańskie otrzymamy:

$$S_r = \{ (s_1, s_2) \in \Re^2 | s_1 = d\cos\varphi, s_2 = d\sin\varphi,$$
$$0 \le d \le r, 0 \le \varphi \le 2\Pi \} \tag{3.4}$$

oraz osie współrzędnych:

$$0x_n = \{ (s_1 \cdot s_2) \in \Re^2 | s_1 = d\cos\varphi_n, s_2 = d\sin\varphi_n,$$
$$0 \le d \le r, \varphi_n = \frac{2\Pi(n-1)}{N} \},\ n \in \aleph \tag{3.5}$$

Określenie 3.2.

Funkcją modelowania pajęczynowego F^P nazywać będziemy odwzorowanie typu:

$$F^P : \mathrm{Y} \to 2^{Sr}$$

przyporządkowujące każdemu elementowi $y \in \mathrm{Y}$ zbiór $S_N(y)$ w postaci wieloboku o wierzchołkach:

$$A_n(y) = (y_n, \beta_n),\ n \in \aleph \tag{3.6}$$

Symbolicznie fakt ten zapiszemy:

$$S_N(y) = [A_1(y), \ldots A_N(y)] \subset S_r \tag{3.7}$$

Wierzchołki wieloboku $S_N(y)$ leżą na poszczególnych osiach współrzędnych przestrzeni pajęczynowej (patrz tzw. wykresy gwiazdowe [6], [7]).

Określenie 3.3.

Modelem pajęczynowym obiektu y (a w konsekwencji obiektu x, takiego, że $F(x) = y$) nazywać będziemy zbiór

$$S_N(y) = F^P(y) \subset S_r.$$

Określenie 3.4.

Zadaniem modelowania pajęczynowego Z_P nazywać będziemy trójkę uporządkowaną

$$Z_P = (P_r, \mathrm{Y}, F^P).$$

Wynikiem realizacji zadania modelowania pajęczynowego jest zbiór $F^P(\mathrm{Y}) = \mathrm{Y}_P$, którego elementami są modele pajęczynowe poszczególnych obiektów $y = F(x)$, $x \in X$

$$\mathrm{Y}_P = \left\{ S_N(y) \in 2^{Sr} \middle| y \in \mathrm{Y} \right\} \tag{3.8}$$

Przykład 1.

Niech $X = \{1, 2, 3\}$, $N = 5$.

Funkcja modelowania opisowego dana jest w tabeli 1. W tabeli tej zapisano również krańcowe wartości cech w_n, w^n, $n \in \aleph = \{1, 2, 3, 4, 5\}$

Wartość parametru r wyznaczamy następująco:

$$r = \max_{n \in \aleph} w^n = 6$$

Tabela 1.

$F(x)$ \ X	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$
1	2	3	$\frac{1}{2}$	1	4
2	1	1	2	4	1
3	3	4	6	4	5
w_n	1	1	$\frac{1}{2}$	1	1
w^n	3	4	6	4	5

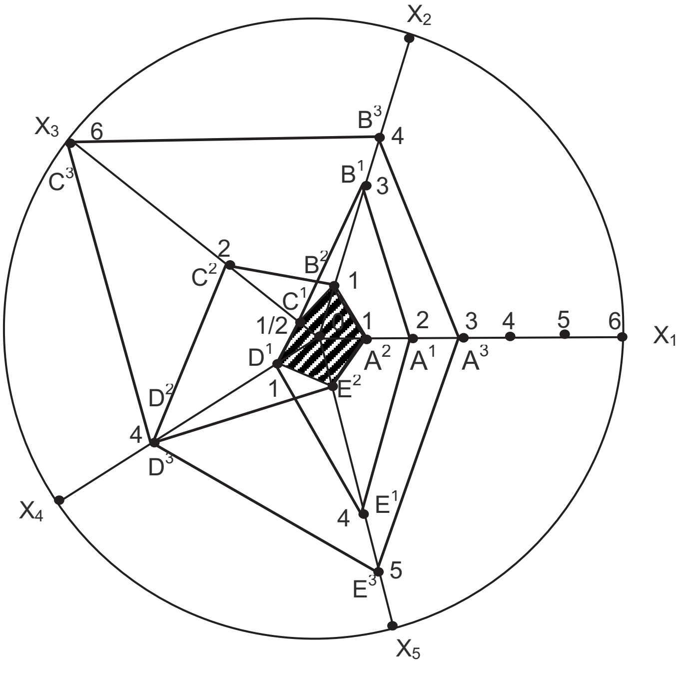
Pięciowymiarową przestrzeń z parametrem $r=6$ zdefiniujemy jako parę uporządkowaną

$$P_6=(S_6,5)\ ,$$

gdzie

$$S_6=\{(d,\varphi)\,|\,0\le d\le 6, 0\le\varphi\le 2\Pi\}$$

Na rys. 2. przedstawiono modele pajęczynowe obiektów $x\in\{1,2,3\}$



Rys. 2. Modele pajęczynowe obiektów $x\in X$

Na rysunku tym zaznaczono również modele „sztucznych” obiektów [1]:

$$w_X=(w_1,...,w_5)=(1,1,\frac{1}{2},1,1)$$

oraz $\quad w^X=(w^1,...,w^5)=(3,4,6,4,5)$,

Zauważmy, że $S_5(w_X)\subset S_5(x)$ dla każdego $x\in X$ oraz, że $S_5(w^X)=S_5(3)$ a ponadto zachodzi $S_5(x)\subset S_5(w^X)=S_5(3)$ dla każdego $x\in X$.

Modelami pajęczynowymi obiektów $x\in\{1,2,3\}$ są odpowiednie zbiory $S_5(1), S_5(2), S_5(3)$.

Są to następujące wieloboki zaznaczone na rys. 2.:

$$S_5(1)=\lfloor A^1,B^1,C^1,D^1,E^1\rfloor$$

$$S_5(2)=\lfloor A^2,B^2,C^2,D^2,E^2\rfloor$$

$$S_5(3)=\lfloor A^3,B^3,C^3,D^3,E^3\rfloor$$

Modele „obiektów sztucznych” [1], [2] w_X i w^X stanowią zbiory:

$S_5(w_X)=\lfloor A^2,B^2,C^1,D^1,E^2\rfloor$ – (zbiór ten został „zakreskowany”).

$$S_5(w^X)=\lfloor A^3,B^3,C^3,D^3,E^3\rfloor=S_5(3)$$

Zbiory te stanowią odpowiednio kres dolny i kres górny zbioru $Y_P=F^P(X)$ w przestrzeni pajęczynowej z relacją inkluzji [1].

Zauważmy, że modele $S_5(1), S_5(2), S_5(3)$ różnią się między sobą „kształtem”, polem powierzchni, „wzajemnym usytuowaniem”, długością obwodu, liczbą boków itp. w przestrzeni $P_6=(S_6,5)$.

Takich dodatkowych charakterystyk modeli $S_N(x)$ może być wiele. Przykładowymi charakterystykami opisującymi dodatkowe właściwości $S_N(x)$ są następujące:

$p(S_N(x))$ – „pole powierzchni” zbioru $S_N(x)$

$g(S_N(x))$ – „środek ciężkości” zbioru $S_N(x)$

$b(S_N(x))$ – „długość obwodu” $S_N(x)$

$\omega(S_N(x))$ – „liczba boków” (wieloboku) zbioru $S_N(x)$

$\delta_N(x)$ – „miara kąta wierzchołkowego” trójkątów tworzących zbiór $S_N(x)$ itp.

W przypadku ustalonych ciągów zbiorów $S_N(x), x\in X_k\subset X, k=1,2,...$ dodatkowymi charakterystykami mogą być „wzajemne usytuowania” elementów ciągów, wzajemne usytuowanie ich środków ciężkości, odległości środków ciężkości, części wspólne zbiorów $S_N(x), x\in X_k$, relacje inkluzji, różnica czy też suma zbiorów itp.

Wszystkie te charakterystyki przy odpowiedniej interpretacji mogą być bardzo cenne z punktu widzenia analizy jakościowej modelowanych obiektów jak też samego procesu modelowania.

Badanie tych charakterystyk pozwala bowiem „wydobyć” więcej informacji o obiektach $x\in X$ niż wynika to wprost z ich formalnych modeli opisowych $y=F(x), x\in X$ (danych w postaci ciągów liczb). Wprowadzając dodatkowo tzw. „gęstość cech” („ciężar właściwy" cech) w postaci $\eta_n(x)$ możemy dokonywać analizy jeszcze bardziej szczegółowej.

W dalszej części pracy, z racji ograniczonej jej objętości zajmiemy się głównie charakterystyką modelu $S_N(x)$ w postaci $p(S_N(x))$ oraz charakterystyką kątową $\delta_N(x)$.

Pole powierzchni zbioru $S_N(x)$ jest pewną funkcją „danych opisowych" obiektu $x \in X$ $(F(x) = y)$ oraz liczby N uwzględnionych cech.

Dla ustalonych N oraz $x \in X$ wartości charakterystyki $p(S_N(x))$ możemy określić następująco:

$$p(S_N(x)) = \alpha_N \left[\sum_{k=1}^{N-1} (y_k y_{k+1}) + y_N y_1 \right] \tag{3.9}$$

gdzie $$\alpha_N = \frac{1}{2} \sin \frac{2\Pi}{N} \tag{3.10}$$

zaś $y_k = F_k(x),\ k = 1,2,\dots,N$

Dodatkowe informacje dotyczące zbioru X modelowanych obiektów niosą też jak wcześniej wspomniano charakterystyki zbiorów:

$$S_N(w_X) \ \ i \ \ S_N(w^X)$$

Odpowiednio pola powierzchni tych zbiorów określamy następująco:

$$p(S_N(w_X)) = \alpha_N \left[\sum_{k=1}^{N-1} (w_k w_{k+1}) + w_N w_1 \right] = \gamma_X(N) \tag{3.11}$$

$$p(S_N(w^X)) = \alpha_N \left[\sum_{k=1}^{N-1} (w^k w^{k+1}) + w^N w^1 \right] = \gamma^X(N) \tag{3.12}$$

Modele $S_N(w_X)$ oraz $S_N(w^X)$ odgrywają ważną rolę w analizie jakościowej obiektów $x \in X$, gdyż zbiory te posiadają następujące ogólne własności:

$$S_N(w_X) \subset S_N(x) \text{ dla każdego } x \in X \tag{3.13.}$$

$$S_N(x) \subset S_N(w^X) \text{ dla każdego } x \in X \tag{3.14}$$

a ponadto

$$\gamma_X(N) \le p(S_N(x)) \le \gamma^X(N), \quad x \in X \tag{3.15}$$

Wartość $p(S_N(x))$ zależy oczywiście od kolejności uwzględnianych cech, stąd też w przypadku braku założenia ustalającego kolejność cech, stosowana jest charakterystyka $\overline{p}(S_N(x))$ będąca „uśrednieniem" pól powierzchni odpowiadających wszystkim możliwym kombinacjom uporządkowań cech (N!).

4. Znormalizowane przestrzenie pajęczynowe

Niech $F(x) = (F_1(x), \dots, F_n(x), \dots, F_N(x))$ – funkcja modelowania taka, że $F: X \to Y \subset \Re^N$.

O funkcji tej założymy, że $F_n(x) \ge 0$, $n \in \aleph$, $x \in X$. Jeśli by tak nie było to należałoby od wartości funkcji $F_n(x)$ odjąć odpowiednią liczbę c_n dla każdego $x \in X$ (przesunięcie skali) [1], [4] w następujący sposób :

$$\overline{F}_n(x) = F_n(x) - c_n, \ x \in X,$$

gdzie $c_n = \min\limits_{x \in X} \overline{F}_n(x), n \in \aleph$.

Dalej zakładać będziemy, że funkcja modelowania spełnia powyższy warunek.

Normalizacji funkcji modelowania można dokonać na wiele sposobów [1], [4].

Jednym z nich jest następujący sposób:

$$\overline{\overline{F}}_n(x) = \frac{\overline{F}_n(x)}{\overline{\overline{F}}_n} \tag{4.1}$$

gdzie

$$\overline{F}_n = \max_{x \in X} \overline{F}_n(x) \ne 0, \ n \in \aleph \tag{4.2}$$

W ten sposób znormalizowana funkcja modelowania ma taką własność, że

$$0 \le \overline{\overline{F}}_n(x) \le 1 \text{ dla każdego } n \in \aleph, x \in X \tag{4.3}$$

Zaś obiekty w_X i w^X mają postać:

$$w_X = (0, \dots, 0) \in \Re^N$$

$$w^X = (1, \dots, 1) \in \Re^N$$

Dalej będziemy zakładać, że funkcja F jest znormalizowana w sensie (4.3).

Określenie 4.1.

Znormalizowaną N-wymiarową przestrzenią pajęczynową nazywać będziemy parą uporządkowaną

$$P_r = (S_r, N)$$

gdzie $r = \max\limits_{n \in \aleph} w^n = 1$

Dla uproszczenia zapisu, w przypadku przestrzeni znormalizowanych będziemy pisać $P = (S, N)$.

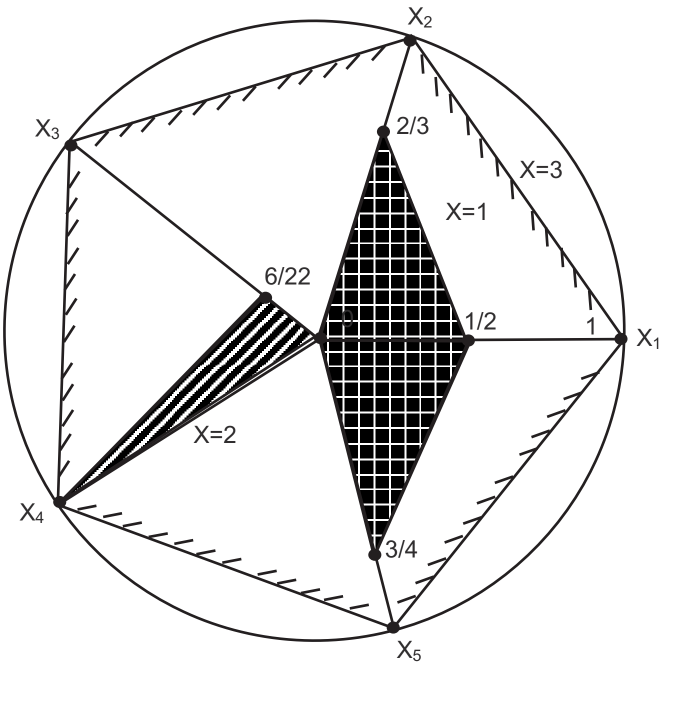
Przykład 2.

W tabeli 2. przedstawiono wartości znormalizowanej funkcji modelowania z przykładu 1. (przesunięcie skali c jest następujące $c = \left(1, 1, \frac{1}{2}, 1, 1\right)$.

Tabela 2. Wartość znormalizowanej funkcji modelowania

$F(x)$ \ X	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$
1	$\frac{1}{2}$	$\frac{2}{3}$	0	0	$\frac{3}{4}$
2	0	0	$\frac{6}{22}$	1	0
3	1	1	1	1	1
w_n	0	0	0	0	0
w^n	1	1	1	1	1

Na kolejnym rysunku przedstawione zostały znormalizowane modele pajęczynowe poszczególnych obiektów $x \in X$.



Rys. 3. Znormalizowane modele pajęczynowe obiektów $x \in X$

5. Problem szczegółowości i adekwatności modeli pajęczynowych

Wśród wielu dodatkowych charakterystyk modelu pajęczynowego $S_N(x)$ na szczególną uwagę zasługują charakterystyki, które mogą być wykorzystane do oceny jakościowej modelowanego obiektu lub też samego procesu modelowania.

Niniejsza praca jak już wcześniej wspomniano dotyczyć będzie charakterystyki $p(S_n(x))$, którą można głównie z racji jej konstrukcji (3.9) wykorzystać do badań jakościowych obiektów jak też wielokryterialnej analizy porównawczej oraz „charakterystyki kątowej" będącej wprost, funkcją maksymalnej liczby uwzględnionych w modelu cech.

Załóżmy dalej, że dysponujemy modelem *M*-dokładnym obiektu $x \in X$. Maksymalna możliwa liczba cech, które można teoretycznie uwzględnić w przypadku modelowania obiektów ze zbioru X to $N(X)$. Oczywiście $M \le N(X)$. Dla kompletności rozważań przyjmijmy, że $M = 0,1,\ldots,N(X)$. Jakość modelu $S_M(x)$ można określić definiując jego „rozbieżność" z oryginałem $x \in X$ lub zamiennie – „dokładność".

Określenie 5.1.

Rozbieżnością modelu M-dokładnego $S_M(x)$ z oryginałem $x \in X$ nazywać będziemy liczbę $\overline{D_M}(x) = 2\Pi - M\,\delta_N(x),\ M \le N(X)$

gdzie *M*- liczba cech aktualnie uwzględnionych w modelu ,zaś $\delta_N(x) = \dfrac{2\Pi}{N(X)}$ graniczna miara kątowa przyrostu „dokładności modelu" przy uwzględnieniu maksymalnej (dla elementów zbioru *X*) liczby cech.

Po przekształceniu mamy:

$$\overline{D}(x) = 2\Pi\left(1 - \frac{M}{N(X)}\right),\quad M \le N(X) \quad (5.1)$$

$$0 \le \overline{D_M}(x) \le 2\Pi,\quad M \le N(X) \quad (5.2)$$

Jeśli na $N(X)$ możliwych do uwzględnienia cech , uwzględnimy w modelu wszystkie, to rozbieżność takiego modelu z oryginałem będzie zerowa.

Jeśli uwzględnimy tylko jedną cechę to rozbieżność wyniesie

$$\overline{D_1}(x) = 2\Pi\,\frac{N(X)-1}{N(X)}.$$

Określenie 5.2.

Stopniem (współczynnikiem) rozbieżności modelu nazywać będziemy liczbę

$$D_M(x) = \frac{\overline{D_M(x)}}{2\Pi} = 1 - \frac{M}{N(X)} \quad (5.3)$$

Mamy przy tym

$$0 \le D_M(x) \le 1,\quad M \le N(X) \quad (5.4)$$

Określenie 5.3.
Dokładnością modelu $S_M(x)$ nazywać będziemy liczbę

$$\overline{d_M}(x) = 2\Pi - \overline{D_M}(x) = 2\Pi \frac{M}{N(X)} \qquad (5.5)$$

Stopień dokładności określimy następująco:

$$d_M(x) = \frac{\overline{d_M}}{2\Pi} = \frac{M}{N(X)} \qquad (5.6)$$

Zachodzi przy tym

$$0 \le d_M(x) \le 1, \quad M \le N(X) \qquad (5.7)$$

Gdy $M \to \infty$ zachodzi oczywiście $D_\infty(x) = 0$ i $d_\infty(x) = 1$

Zagadnienie adekwatności czy też rozbieżności modelu jest oczywiście problemem bardziej złożonym. Analizując charakterystykę $p(S_n(x))$ jak również „kształt" zbioru $S_N(x)$ można podjąć się prób definiowania innych, „bardziej związanych" z modelem obiektu $x \in X$ wskaźników charakteryzujących proces modelowania i sam model. Porównując kształty zbiorów $S_N(x)$ przy wzroście liczby N uwzględnianych cech, ich powierzchnię, obwód itp. można jakość modelu zdefiniować bardziej precyzyjnie. Kluczową rolę mogą odgrywać w tym podejściu graniczne wartości $S_\infty(x)$, $p(S_\infty(x))$ przy $N \to \infty$ jako swoistego rodzaju punkty odniesienia. Poniżej zostaną przedstawione pewne rozważania dotyczące możliwej interpretacji charakterystyki $p(S_\infty(x))$ w określaniu stopnia szczegółowości modelu $S_N(x)$ w zależności od wielkości liczby N. Rozważania te nie mają charakteru ogólnego i są słuszne jedynie dla obiektów $x \in X$ spełniających pewne szczególne założenia. Charakterystyki takie bowiem jak kształt zbioru $S_N(x)$, pole powierzchni, środek ciężkości itp. służą przede wszystkim analizie jakościowej obiektów $x \in X$ [1], [2], [3], a nie samego procesu modelowania.

Załóżmy, że mamy dane zadanie modelowania pajęczynowego w przestrzeni znormalizowanej $Z = (P, X, F)$. Intuicyjnie, stopień szczegółowości modelu (opisu) obiektu $x \in X$ jak już było wspomniane powinien rosnąć wraz z wielkością liczby M cech uwzględnionych w procesie modelowania $(M \le N)$, jakość (adekwatność) modelu powinna również posiadać podobną własność.

Wielkość pola powierzchni modelu $S_M(x)$, zgodnie z zależnością (3.9) i (3.10) formalnie jest pewną funkcją liczby M uwzględnianych cech oraz wartości tych cech określonych jako ciąg liczb:

$$y = (F_1(x), \ldots, F_m(x), \ldots, F_M(x)) \in \Re^M$$

co zapiszemy

$$p(S_M(x)) = f(y, M) \qquad (5.8)$$

Ustalając konkretną wartość M, możemy badać funkcję $f_M(y)$, $y \in Y$. Przy pewnych założeniach jakościowych odnośnie interpretacji $y_m = F_m(x)$, $m \in N$ wartości $f_M(y)$ może być interpretowana jako uogólniona jakość obiektu x, takiego że $F(x) = y$ [1], [6]. Można zatem ją wykorzystać do budowy tzw. funkcji rankingowej oraz innych funkcji decyzyjnych w tym w szczególności optymalizacji.

Z kolei ustalając $y = F(x)$, możemy badać zmienność $f_y(M)$ w zależności od $M = 1, 2, \ldots, N$, co może pozwolić na ocenę zmian szczegółowości (dokładności) procesu modelowania, wynikających ze zmian liczby uwzględnionych cech. Przykładowo adekwatność modelu $S_M(x)$ możemy zdefiniować jako pewną funkcję $q(S_M(x))$ o postaci następującej:

$$q(S_M(x)) = \frac{p(S_M(x))}{p(S_N(x))}, \; x \in X, \; M = 1, \ldots N \qquad (5.9)$$

Zauważmy, że dla $M = N$, $q(S_M(x)) = 1$.

6. Modelowanie obiektów w przestrzeni nieskończenie wymiarowej

W dalszej części opracowania zakładać będziemy, że dysponujemy znormalizowaną funkcją modelowania

$$F : X \to Y \subset \underbrace{[0,1]x \ldots x[0,1]}_{\text{M-razy}} \qquad (6.1)$$

Biorąc dowolny element $y \in Y$ możemy przedstawić jego model w przestrzeni pajęczynowej P jako pewien podzbiór $S_M(y)$ zbioru S.

Głównymi charakterystykami tego modelu będą „szczegółowość" oraz „adekwatność" rozumiane następująco :

$$p(S_M(y)) \text{ oraz } q(S_M(y)) = \frac{p(S_M(y))}{p(S_N(y))} \qquad (6.2)$$

Załóżmy, że chcemy zwiększyć „stopień szczegółowości modelu" oraz jego adekwatność poprzez zwiększenie maksymalnej liczby N uwzględnianych cech.

Zwiększając liczbę N do nieskończoności $(N \to \infty)$ otrzymamy nieskończenie wymiarową przestrzeń pajęczynową $P_\infty = (S, \infty)$.

Ustalając w konkretnym zadaniu modelowania liczbę $M \leq N$ najważniejszych cech otrzymamy model M-dokładny $S_M(y) \subset S$.

Określenie adekwatności $q(S_M(y))$ tak uzyskanego modelu $S_M(y)$ wymaga wyznaczenia granicznej wartości „stopnia szczegółowości" $p(S_N(y))$

$$\lim_{N \to \infty} p(S_N(y)) = p(S_\infty(y)) \tag{6.3}$$

o ile oczywiście granica taka istnieje.

O charakterystyce $p(S_N(y))$ możemy powiedzieć, że

$$0 \leq p(S_N(y)) \leq \Pi \ , \ y \in Y, \ N = 1,2,.. \tag{6.4}$$

Dla modelu $y = (1, \ldots, 1)$, przy $N \to \infty$

$$p(S_N(y)) \to p(S_\infty(y)) = \Pi \tag{6.5}$$

$S_\infty(y)$ jest modelem obiektu $y \in Y$ w nieskończenie wymiarowej przestrzeni pajęczynowej (S_∞, ∞).

Liczba $p(S_\infty(y))$ wyraża maksymalną (idealną) szczegółowość modelu, zaś adekwatność modelu jest równa 1.

Badając model przy ustalonej liczbie $M < \infty$ uwzględnianych cech, możemy określić też dodatkową jego charakterystykę:

$$d(S_M(y)) = \frac{p(S_\infty(y)) - p(S_M(y))}{p(S_\infty(y))} = 1 - \frac{p(S_M(y))}{p(S_\infty(y))} = 1 - q(S_M(y)) \tag{6.6}$$

oczywiście przy $M \to \infty$, $d(S_M(y)) \to 0$.

7. Podsumowanie

Przedstawione powyżej rozważania stanowią próbę sformalizowania procesu modelowania obiektów, bazujące na tzw. przestrzeni pajęczynowej.

Graficzna interpretacja modeli pajęczynowych nawet dla dużych wartości liczby N jest możliwa i łatwa do implementacji komputerowej. Daje nowe przesłanki pogłębionej analizy jakościowej modelowanych obiektów. Jest typowym narzędziem analitycznym między innymi w procedurach eksploracji danych. Bardzo interesujące wyniki można uzyskać wprowadzając w przestrzeni dodatkowe relacje np. inkluzji, określające wzajemne związki poszczególnych par modeli lub ich sekwencji. Wprowadzenie takiej relacji umożliwia porządkowanie obiektów wg różnych kryteriów lub wyznaczania klas równoważności, jak też konstruowanie funkcji rankingowych.

Dodatkowe charakterystyki w postaci wielkości pól powierzchni poszczególnych modeli graficznych oraz ich „wzajemnego usytuowania" i zależności w znormalizowanej przestrzeni pajęczynowej mogą być kolejnymi przesłankami ustaleń w zakresie wiedzy o analizowanych obiektach.

Pajęczynową przestrzeń modelową można potraktować też jako przestrzeń decyzyjną. Wprowadzenie do tej przestrzeni modelu preferencji decydenta może pozwolić na formułowanie, badanie i rozwiązywanie odpowiednio sformułowanych zadań optymalizacji.

8. Bibliografia

[1] A. Ameljańczyk, *Optymalizacja wielo-kryterialna w problemach sterowania i zarządzania*, Ossolineum, Warszawa-Kraków, 1984.

[2] A. Ameljańczyk, „Matematyczny model przestrzeni życia w komputerowym systemie wspomagania decyzji medycznych", materiały I Krajowej Konferencji „Systemy komputerowe i teleinformatyczne w służbie zdrowia", Warszawa, 2009.

[3] A. Ameljańczyk, *O pewnej koncepcji modelowania repozytorium medycznego*, opracowanie wewnętrzne WAT w ramach projektu POIG.01.03.01-00-145/08/2009.

[4] J. Gutenbaum, *Modelowanie matematyczne systemów*, PWN, 1987.

[5] J. Kacprzyk, *Zbiory rozmyte w analizie systemowej*, PWN, 1986.

[6] D. Hand, H. Mannila, P. Smyth, *Eksploracja danych*, WNT, Warszawa, 2005.

[7] D.T. Larose, *Odkrywanie wiedzy z danych,* PWN, 2006.

[8] Z. Pawlak, *Systemy informacyjne – podstawy teoretyczne*, WNT, Warszawa, 1983.

[9] P. Smets, „Medical diagnosis: Fuzzy sets and degrees of belief", *Fuzzy Sets and Systems*, vol. 5, 1981.

[10] L.A. Zadeh, „Information Control", *Fuzzy Sets*, vol. 8, 1965.