



Modelowanie podobieństwa diagnostycznych wzorców medycznych w przestrzeni pajęczynowej

ANDRZEJ AMELJAŃCZYK

Wojskowa Akademia Techniczna, Wydział Cybernetyki,
00-908 Warszawa, ul. S. Kaliskiego 2, aameljanczyk@wat.edu.pl

Streszczenie. W pracy przedstawiono możliwość wykorzystania przestrzeni pajęczynowych do graficznego modelowania wzorców medycznych jednostek chorobowych oraz stanu zdrowia pacjenta. Przedstawiono ogólną koncepcję algorytmu wspomagania procesu diagnozowania stanu zdrowia pacjenta, bazującą na określaniu podobieństwa modeli graficznych stanu zdrowia pacjenta i zawartych w repozytorium wzorców jednostek chorobowych. Sformułowano wielokryterialne zadanie optymalizacji rozpoznawania wzorców medycznych.

Słowa kluczowe: przestrzeń pajęczynowa, model graficzny, wzorzec medyczny, podobieństwo modeli graficznych, przestrzeń podobieństwa, relacja podobieństwa, model Tversky'ego

1. Wprowadzenie

Pojęcie diagnostycznego wzorca medycznego wiąże się z procesem diagnozowania medycznego, rozumianego jako proces określania podobieństwa „stanu zdrowia pacjenta” do wzorców jednostek chorobowych zawartych w repozytorium [2, 8, 9]. W poniższej pracy zostaną przedstawione metody modelowania stanu zdrowia pacjenta, wzorców zdrowia, wzorców jednostek chorobowych oraz ich podobieństwa w przestrzeni pajęczynowej [5].

Stan zdrowia pacjenta $x \in X$ (gdzie X zbiór numerów pacjentów, np. według PESEL) w chwili $t \in T$ można określić, podając wartości jego parametrów zdrowotnych [10, 14, 18]. Liczba rozpatrywanych parametrów zdrowotnych zależy oczywiście od potrzeb modelowania i oczekiwanej dokładności. Może ich być od kilku do nawet kilkuset. Przykładowo mogą to być:

$f_1(x, t)$ — temperatura ciała,

- $f_2(x, t)$ — ciśnienie krwi,
- $f_3(x, t)$ — stężenie glukozy we krwi,
- $f_4(x, t)$ — stężenie cholesterolu,
- $f_5(x, t)$ — stężenie PSA,
- $f_6(x, t)$ — stężenie kreatyniny,
- $f_7(x, t)$ — stężenie HGB, itp.

Ogólnie przez $f_n(x, t)$ oznaczać będziemy wartość parametru zdrowotnego $n \in \mathbf{N}$, pacjenta $x \in X$ w chwili $t \in T$. Zbiór $\mathbf{N} = \{1, \dots, n, \dots, N\}$ jest zbiorem numerów wyróżnionych (uwzględnionych w modelu) parametrów zdrowotnych.

W przypadku parametrów opisowych („trudno mierzalnych”) można zastosować odpowiednie techniki wartościowania w ustalonej skali [4, 10]. Symbolem S_n oznaczmy zakres możliwych wartości parametru $n \in \mathbf{N}$

$$S_n = [\underline{s}_n, \bar{s}_n] \subset \mathbf{R}^1. \quad (1.1)$$

Przestrzenią stanu zdrowia pacjenta (przestrzenią życia) nazywać będziemy zbiór $S = S_1 \times S_2 \times \dots \times S_n$, będący iloczynem kartezjańskim zakresów wartości S_n , $n \in \mathbf{N}$.

Oczywiście

$$f(x, t) \in S \subset \mathbf{R}^N, \quad x \in X, \quad t \in T. \quad (1.2)$$

Modelem stanu zdrowia pacjenta x w chwili t będzie zatem wektor

$$f(x, t) = (f_1(x, t), \dots, f_n(x, t), \dots, f_N(x, t)) \in \mathbf{R}^N. \quad (1.3)$$

W zależności od doboru wielkości liczby N , która jest licznością zbioru będziemy mieli do czynienia z modelem bardziej lub mniej dokładnym [5, 10]. Zbiory S_n , $n \in \mathbf{N}$ mogą być definiowane również jako rozmyte lub przybliżone [1, 12, 13, 15, 19, 20]. W większości dalszych rozważań będziemy dla uproszczenia pomijali czynnik czasu, stosując zamiast modelu $f(x, t)$ uproszczony model $f(x) = (f_1(x), \dots, f_n(x), \dots, f_N(x))$.

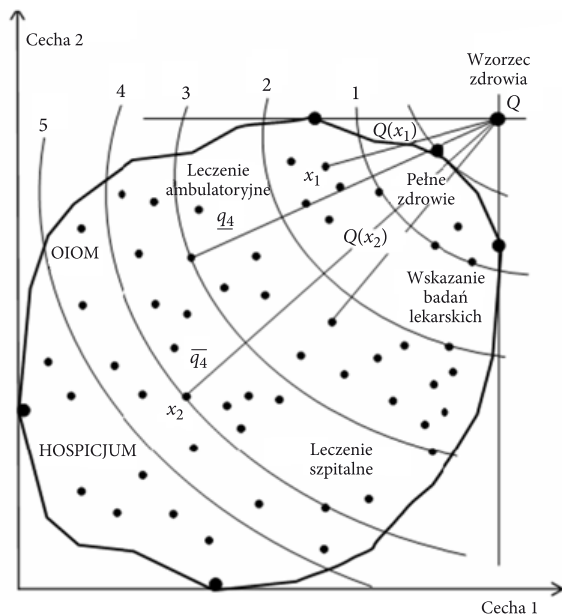
Ustalając wzorcową (idealną) wartość parametru $n \in \mathbf{N}$, świadczącą o „pełnym zdrowiu” jako $q_n \in S_n$ otrzymamy wzorzec zdrowia w postaci

$$Q = (q_1, \dots, q_n, \dots, q_N) \in S. \quad (1.4)$$

W bardziej precyzyjnych analizach oczywiście będziemy definiować zamiast wzorca Q wzorzec $Q(t)$ (wzorzec stanu zdrowia w chwili t).

Dokonując normalizacji wartości poszczególnych cech (parametrów zdrowotnych) [3, 4, 7], otrzymamy tak zwaną znormalizowaną przestrzeń życia. Na rysunku 1

została przedstawiona znormalizowana przestrzeń życia w sytuacji, gdy w modelu uwzględniono tylko dwa parametry ($N = 2$).



Rys. 1. Przestrzeń życia

Zbiór Y będący obrazem zbioru X w projekcji wartości parametrów zdrowotnych stanowi zbiór modeli stanu zdrowia pacjentów ze zbioru X

$$Y = f(X) = \{f(x) \in \mathbf{R}^N \mid x \in X\}. \tag{1.5}$$

Uogólniony wskaźnik $Q_p(x)$ stanu zdrowia pacjenta $x \in X$ możemy zdefiniować jako odległość modelu jego stanu zdrowia $f(x)$ od wzorca zdrowia Q :

$$Q_p(x) = \|Q - f(x)\|_p = \left(\sum_{n \in N} (q_n - f_n(x))^p \right)^{1/p}, \quad p \geq 1 \tag{1.6}$$

Parametr $p \geq 1$ określa konkretną postać normy $\|\cdot\|_p$ (dla wartości p całkowitych jest to tzw. norma Minkowskiego) [3]. W praktyce najczęściej stosujemy odległość euklidesową lub Czebyszewa ($p = 2, \infty$) [3]. Im wartość funkcji $Q_p(x)$ jest mniejsza, tym stan zdrowia pacjenta x jest lepszy. Często używa się miary pochodnej typu $\bar{Q}_p(x) = 1 - Q_p(x)$, $x \in X$ (w przestrzeni znormalizowanej). Zastosowanie normy z parametrem p jest oczywiście dużym przybliżeniem, szczególnie niebezpiecznym w algorytmach diagnostycznych, gdyż pozwala wyznaczać jedynie

tw. podobieństwo metryczne. Może być jednak z dużym powodzeniem wykorzystywane jako wskaźnik zmian odpowiednich charakterystyk kompleksowych.

Przykładowo stan zdrowia pacjenta x_2 jest znacznie gorszy niż stan zdrowia pacjenta x_1 , gdyż $Q_p(x_1) < Q_p(x_2)$ (patrz rys. 1).

Określając pewne wartości graniczne wskaźnika stanu zdrowia, można zbiór Y podzielić na klasy:

$\underline{q}_1 \leq Q_p(x) \leq \bar{q}_1$ — pacjent x zdrowy,

$\underline{q}_2 \leq Q_p(x) \leq \bar{q}_2$ — pacjent x wymagający badań,

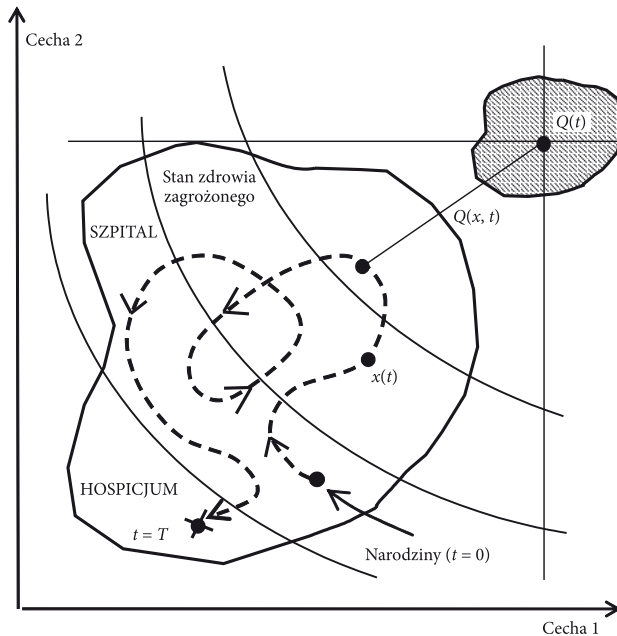
$\underline{q}_3 \leq Q_p(x) \leq \bar{q}_3$ — pacjent x powinien być leczony (ambulatoryjnie/w szpitalu).

Dla uproszczenia zapisu modelu stanu zdrowia pacjenta x oznaczymy na roboczo:

$$f_n(x, t) = x_n(t), \quad n \in \mathbf{N}.$$

Uwzględniając aspekt czasu $t \in T$ można analizować dynamikę zmian $x_n(t)$ („wędrowka” po przestrzeni życia) oraz dynamikę zmian globalnego wskaźnika $Q(x, t)$.

Przedstawiono to poglądowo na rysunku 2. Punkt $x(t)$ (uproszczony model stanu zdrowia pacjenta x), poruszając się wraz upływem czasu, „wykreśla” pewną trajektorię w zależności od zmieniających się w czasie wartości poszczególnych jego



Rys. 2. „Wędrowka” pacjenta w przestrzeni życia

cech zdrowotnych. W zależności od położenia tego punktu można podejmować określone „działania medyczne”, by dokonywać celowej korekty tego „ruchu”, np. stosując określone leki (terapię), programy medyczne. Dokonując odpowiedniego monitoringu zmian położenia tego „punktu”, można zapisywać w bazie danych pacjenta kolejne wartości jego parametrów zdrowotnych. Dane te w zestawieniu z informacją o stosowanej terapii umożliwią pełną analizę jej skutków. Dane z monitoringu zawierają będą oprócz danych o szczegółowych zmianach wszystkich parametrów zdrowotnych również dane o zmianach globalnej odległości od wzorca zdrowia. Dane te świadczą będą o ogólnym stanie zdrowia pacjenta. „Umieszczenie” w przestrzeni życia dodatkowo wzorców jednostek chorobowych oraz wzorców innych zagrożeń zdrowotnych pozwoli na jeszcze bardziej precyzyjną analizę skutków wprowadzanych leków. Określając każdorazowo odległość aktualnego stanu zdrowia pacjenta x (położenie punktu $x(t)$ w przestrzeni życia) względem wzorców wybranych jednostek chorobowych, można precyzyjnie ustalić wpływ badanego leku na zmianę stopnia zagrożenia tymi schorzeniami. W ten sposób można zaobserwować wszelkie skutki uboczne stosowanego leku. Kolejne obrazy stanu zdrowia pacjenta w postaci ciągu wektorów $x(t)$ można dodatkowo „wizualizować” w przestrzeni dwuwymiarowej, korzystając z tzw. „obrazów pajęczynowych” [5], wydobywając dodatkowe informacje diagnostyczne [5, 7].

Powyższa koncepcja może stanowić podstawę budowy odpowiedniego systemu informatycznego pozwalającego na monitoring i kompleksową, obiektywną analizę wszelakich skutków wprowadzanych programów ochrony zdrowia, terapii czy też nowych leków.

Modele jednostek chorobowych oraz ich wzorce można zdefiniować analogicznie. Niech $\mathbf{M} = \{1, \dots, m, \dots, M\}$ — zbiór numerów wyróżnionych jednostek chorobowych. Symbolem C_n^m oznaczmy zakres chorobowych wartości parametru $n \in \mathbf{N}$ w chorobie $m \in \mathbf{M}$

$$C_n^m = [c_n^m, \bar{c}_n^m]. \quad (1.7)$$

Modelem uproszczonym jednostki chorobowej w $m \in \mathbf{M}$ będzie zbiór

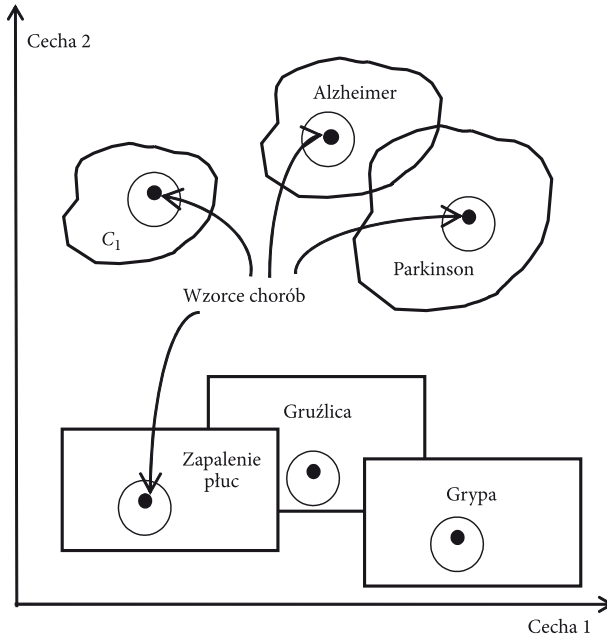
$$C^m = \{c \in \mathbf{R}^N \mid c_n^m \in C_n^m, n \in \mathbf{N}\} \quad (1.8)$$

Jest to tzw. „model przedziałowy” — najczęściej stosowany. Do modelowania wzorców jednostek chorobowych stosuje się też wiele innych technologii [1, 4, 5, 12, 15, 20].

Podobnie jak w przypadku wzorca zdrowia (1.4), za wzorec choroby $m \in \mathbf{M}$ możemy przyjąć wektor

$$w^m = (w_1^m, \dots, w_n^m, \dots, w_N^m) \in \mathbf{R}^N,$$

gdzie w_n^m — „wzorcowa wartość chorobowa” parametru n w chorobie $m \in \mathbf{M}$.



Rys. 3. Wzorce chorobowe JCH przy różnych technologiach modelowania

Bardzo często zamiast wektora w^m korzysta się ze zbioru wzorcowych wartości chorobowych [8, 9, 10]:

$$\mathbf{E}(w^m) = \left\{ x \in \mathbf{R}^N \mid \|w^m - x\|_p \leq \mathbf{E} \right\} \subset C^m, \quad (1.9)$$

wykorzystując przy tym pojęcie zbioru rozmytego lub przybliżonego [13, 20].

Na kolejnym rysunku przedstawiono zintegrowaną z obszarami zagrożeń chorobowych przestrzeń stanu zdrowia pacjenta.

Elementy zbioru Y oznaczają „położenie” stanu zdrowia pacjenta w przestrzeni życia. Oczywiście, linie oznaczające odpowiednie obszary na rysunku 4 mają charakter jedynie poglądowy. Konkretnie ich wyznaczenie możliwe jest dopiero po odpowiednim „wykalibrowaniu” modelu, np. w wyniku symulacji komputerowej.

Stopień zagrożenia pacjenta x chorobą $m \in \mathbf{M}$ możemy określić jako odległość modelu jego stanu zdrowia $x(t)$ w chwili $t \in T$ od wzorca choroby $m \in \mathbf{M}$

$$Z^m(x, t) = f \left(\|w^m - x(t)\|_p \right). \quad (1.10)$$

Jeśli $x(t) \in \mathbf{E}(w^m)$ — istnieje podejrzenie choroby numer $m \in \mathbf{M}$.

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 bloków 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. stan zdrowia pacjenta [7], liczba ta może sięgać kilkuset i więcej.

Teoretycznie możemy rozpatrywać procesy modelowania, gdzie liczba $N \rightarrow \infty$ [5]. Tego typu analizy pozwalają na badanie tzw. „modeli idealnych”.

Określenie 2.1

Modelem M -dokładnym ($M \leq N$) 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 $F: X \rightarrow \mathbf{R}^N$, przyporządkowującą każdemu modelowanemu obiektowi $x \in X$ ciąg wartości poszczególnych jego cech. Tak więc modelem obiektu $x \in X$ będzie jego obraz

$$F(x) = (F_1(x), \dots, F_n(x), \dots, F_N(x)) \in \mathbf{R}^N \quad (2.1)$$

dany funkcją modelowania $F: X \rightarrow \mathbf{R}^N$, gdzie $F_n(x)$ – wartość n -tej cechy modelowanego obiektu $x \in X$.

Założmy dalej, że dla każdego $n \in \mathbf{N}$

$$w_n \leq F_n(x) \leq w^n, \quad x \in X. \quad (2.2)$$

Dolny i górny indeks n oznaczają odpowiednio dolną i górną wartość tego parametru.

Warunek (2.2) będziemy zapisywać następująco:

$$y_n = F_n(x) \in [w_n, w^n] \subset \mathbf{R}^1, n \in \mathbf{N}, x \in X. \quad (2.3)$$

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

Niech przykładowo $N = 15$, $M = 9$. Modelem dokładnym obiektu $x \in X$ jest $F(x) = (1, 2, 1, 1, 3, 4, 8, 7, 2, 5, 4, 3, 1, 0, 7) \in \mathbf{R}^{15}$, zaś modelem M -dokładnym $F^M(x) = (1, 2, 1, 1, 3, 4, 8, 7, 2) \in \mathbf{R}^9$.

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

$$Z_0 = (\mathbf{R}^N, X, F). \quad (2.4)$$

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

$$Y = F(X) = \{y = F(x) \in \mathbf{R}^N \mid x \in X\} \subset \mathbf{R}^N. \quad (2.5)$$

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

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

Jeśli w zbiorze klas równoważności $\{X(y) \subset X \mid y \in Y\}$, 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 rozróżnialności, czyli określenia takiej liczby N (o ile istnieje), że dla każdego $y \in Y$, $|X(y)| = 1$.

Załóżmy, że dane jest zadanie modelowania opisowego: $Z_0 = (\mathbf{R}^N, X, F)$, 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 „wydobycia” dodatkowej wiedzy z danych opisujących modelowane obiekty daje tzw. modelowanie pajęczynowe (web modelling).

Określenie 2.4 [5]

N -wymiarową przestrzeń pajęczynową nazywać będziemy parą uporządkowaną $P = (S, N)$, gdzie S jest pewnym zbiorem wyznaczanym na płaszczyźnie, np. w układzie współrzędnych biegunowych w następujący sposób:

$$S = \{(d, \varphi) \mid d \geq 0, 0 \leq \varphi \leq 2\Pi\}. \quad (2.7)$$

Para (d, φ) to współrzędne punktu (elementu) zbioru S (d — odległość od tzw. bieguna, φ — kąt skierowany).

Osie współrzędnych $0x_n$, $n \in \mathbf{N}$ przestrzeni pajęczynowej możemy wyznaczyć następująco:

$$0x_n = \left\{ (d, \varphi_n) \mid d \geq 0, \varphi_n = \frac{2\Pi(n-1)}{N} \right\}, \quad n \in \mathbf{N}. \quad (2.8)$$

Dla celów praktycznej analizy najczęściej jednak korzystamy z tzw. pojęcia „ograniczonej przestrzeni pajęczynowej”.

Określenie 2.5

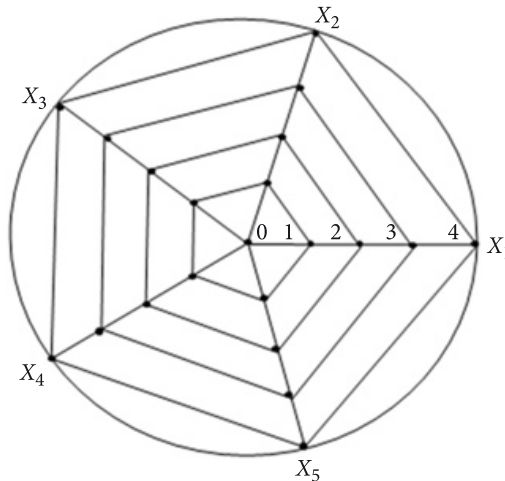
N -wymiarową, ograniczoną przestrzenią pajęczynową nazywamy parę uporządkowaną

$$P_r = (S_r, N), \quad (2.9)$$

gdzie $S_r = \{(d, \varphi) \mid 0 \leq d \leq r, 0 \leq \varphi \leq 2\Pi\}$.

Liczba $r > 0$ nazywana jest promieniem (ograniczeniem) przestrzeni pajęczynowej. Liczba ta odgrywa zasadniczą rolę w tzw. przestrzeniach znormalizowanych.

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



Rys. 5. Pięciowymiarowa przestrzeń pajęczynowa o promieniu $r = 4$

Określenie 2.6 [5]

Funkcją modelowania pajęczynowego F^P nazywać będziemy odwzorowanie typu:

$$F^P : Y \rightarrow 2^{S_r} \quad (2.10)$$

przyporządkowujące każdemu elementowi $y \in Y$ zbiór

$$F^P(y) \subset S_r. \quad (2.11)$$

Określenie 2.7

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

$$F^P(y) \subset S_r. \quad (2.12)$$

Jest to model graficzny zestawu danych y , tzw. „glyph”.

Określenie 2.8

Zadaniem modelowania pajęczynowego Z_p nazywać będziemy trójkę uporządkowaną $Z_p = (P_r, Y, F_p)$.

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

$$Y^P = \{F^P(y) \in 2^{S_r} \mid y \in Y\}. \quad (2.13)$$

Określenie 2.9

Znormalizowaną N -wymiarową przestrzenią pajęczynową nazywać będziemy parę uporządkowaną $P_r = (S_r, N)$, gdzie $r = 1$.

W praktyce, z racji różnych mian poszczególnych cech, wykorzystywane są głównie przestrzenie znormalizowane.

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

3. Modele graficzne wzorców medycznych

Niech $Y = F(X) \subset \mathbf{R}^N$ — zbiór modeli opisowych (zbiór „zestawów” diagnostycznych danych medycznych) pacjentów ze zbioru X , przy czym $F(x) = y = (y_1, \dots, y_n, \dots, y_N) \in \mathbf{R}^N$ jest zestawem wyników badań medycznych (parametrów medycznych) pacjenta $x \in X$.

W zależności od postaci (2.12) funkcji modelowania pajęczynowego:

$$F^P : Y \rightarrow 2^{S_r} \quad (3.1)$$

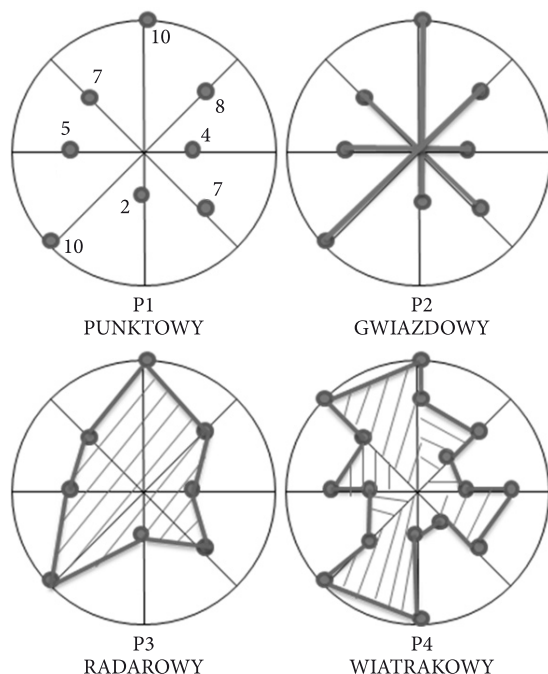
otrzymamy ich modele graficzne

$$F^P(y) \subset S_r, \quad y \in Y. \quad (3.2)$$

Symbolem Z oznaczmy zbiór modeli graficznych danych

$$Z = F^P(Y) = \{F^P(y) \subset S_r \mid y \in Y\} \quad (3.3)$$

Poniższe rysunki przedstawiają przykładowe modele graficzne („glyphy”) wybranego zestawu danych $y = (4, 8, 10, 7, 5, 10, 27)$ dla różnych funkcji modelowania pajęczynowego.

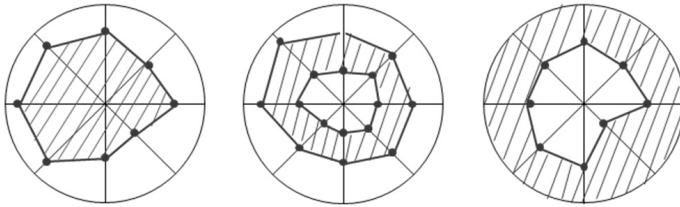


Rys. 6. Przykładowe modele graficzne zestawu danych $y \in Y$

Dla ustalonych funkcji modelowania, modelami graficznymi danych są więc odpowiednie „glyphy” (obrazki). Modele te po „nałożeniu na siebie” już na „pierwszy rzut oka” różnią się wieloma cechami: kształtem, polem powierzchni, wzajemnym usytuowaniem, położeniem „środków ciężkości” itp.

Przy określonych założeniach, proces diagnozowania choroby pacjenta może polegać na badaniu podobieństwa jego stanu zdrowia do odpowiednio zdefiniowanych wzorców jednostek chorobowych, przedstawionych w postaci modeli graficznych [6]. W podobny sposób można również badać ogólny stan zdrowia pacjenta jako „stopień podobieństwa” modelu graficznego aktualnego stanu jego zdrowia $y \in Y$, do modelu graficznego „wzorca zdrowia” odpowiadającego tej klasie pacjentów (wiek, płeć itp.). Na rysunku 7 przedstawiono typowe modele graficzne wzorców jednostek chorobowych.

W zależności od warunków, jakie muszą być spełnione przez poszczególne „parametry chorobowe” definiujące jednostkę chorobową, otrzymamy tzw. glyphy wewnętrzne, przedziałowe lub zewnętrzne.



Rys. 7. Typowe wzorce jednostek chorobowych w przestrzeni pajęczynowej

Przykład 3.1

Niech stan zdrowia pacjenta x określa zestaw wyników badań wartości parametrów medycznych $F(x) = y = (y_1, \dots, y_n, \dots, y_N) \in \mathbf{R}^N$.

Wymiar N — wynika oczywiście z potrzeby uzyskania odpowiedniej dokładności oceny.

Przykładowo niech $y = (7, 6, 7, 8, 5, 3, 6, 7) \in \mathbf{R}^8$.

Wzorcem zdrowia dla tej klasy pacjentów niech będzie wektor

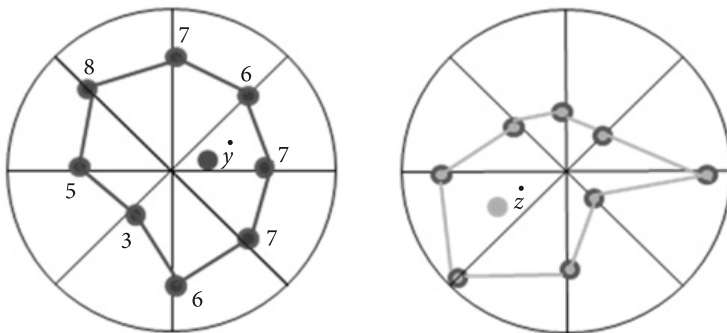
$z = (9, 4, 5, 5, 8, 10, 8, 2) \in \mathbf{R}^8$.

Stopień podobieństwa metrycznego określa liczba $\|z - y\|_p, p \geq 1$ (dla $p = 2$ — odległość geometryczna). Jest to „miara podobieństwa” bardzo przybliżona i obciążona wieloma wadami.

Analizując podobieństwo graficznych modeli $F^p(y)$ i $F^p(z)$, możemy „wydobyć” o wiele więcej informacji świadczących o stopniu podobieństwa y i z .

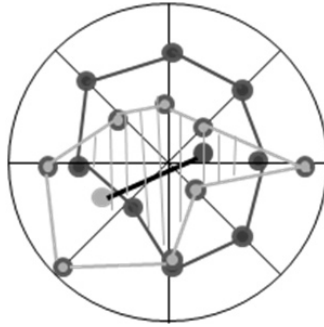
4. Charakterystyki podobieństwa modeli graficznych

Na rysunku 8 przedstawione zostały: model graficzny stanu zdrowia pacjenta $F^p(y)$ oraz model graficzny wzorca zdrowia $F^p(z)$



Rys. 8. Modele Graficzne $F^p(y)$ i $F^p(z)$

Poniższy rysunek, powstały z nałożenia na siebie graficznego modelu stanu zdrowia pacjenta i modelu wzorca zdrowia, pozwala zauważyć ich „podobieństwa” i „różnice”.



Rys. 9. Porównanie modeli graficznych

Na rysunku tym zaznaczono między innymi środki ciężkości obu modeli, część wspólną oraz ich różnice.

Im część wspólna $F^P(y) \cap F^P(z)$ „jest większa”, tym podobieństwo modeli większe. Natomiast, im różnice $(F^P(y) - F^P(z))$ oraz $(F^P(z) - F^P(y))$ są większe, tym podobieństwo modeli jest mniejsze.

Podobnie, im środki ciężkości obu modeli znajdują się bliżej siebie, tym podobieństwo jest większe.

Są to dodatkowe przesłanki oceny podobieństwa porównywanych „obiektów”.

Analiza tych wielkości pozwala bardziej precyzyjnie definiować i oceniać stopień podobieństwa rozpatrywanych modeli graficznych. Wnioskowanie o stopniu podobieństwa na podstawie powyższych charakterystyk jest zatem o wiele pełniejsze niż wnioskowanie jedynie w oparciu o charakterystykę metryczną $\|z - y\|_p$.

Niech $p(A)$ — oznacza pole powierzchni zbioru $A \subset \mathbf{R}^2$.

$y \in F^P(y)$ oraz $z \in F^P(z)$ — odpowiednio oznaczają środki ciężkości zbiorów $F^P(y)$ i $F^P(z)$.

Omawiane „charakterystyki podobieństwa” w uproszczeniu można zdefiniować następująco:

$$f_1(y, z) = p(F^P(y) \cap F^P(z)), \quad (4.1)$$

$$f_2(y, z) = p(F^P(y) - F^P(z)), \quad (4.2)$$

$$f_3(y, z) = p(F^P(z) - F^P(y)), \quad (4.3)$$

$$f_4(y, z) = \left\| z - y \right\|_p^*, \quad p \geq 1. \quad (4.4)$$

Wielokryterialną funkcją podobieństwa nazywać będziemy funkcję:

$f : Y \times Y \rightarrow \mathbf{R}^4$, gdzie

$$f(y, z) = (f_1(y, z), f_2(y, z), f_3(y, z), f_4(y, z)) \in \mathbf{R}^4, \quad (y, z) \in Y \times Y. \quad (4.5)$$

Trzy pierwsze funkcje cząstkowe można też zastąpić ich wersjami znormalizowanymi:

$$\overline{f}_1(y, z) = \frac{p(F^P(y) \cap F^P(z))}{p(F^P(y) \cup F^P(z))}, \quad (4.6)$$

$$\overline{f}_2(y, z) = \frac{p(F^P(y) - F^P(z))}{p(F^P(y))}, \quad (4.7)$$

$$\overline{f}_3(y, z) = \frac{p(F^P(z) - F^P(y))}{p(F^P(z))}. \quad (4.8)$$

5. Rozpoznawanie wzorców medycznych

Założmy, że $Y \subset \mathbf{R}^N$ to zbiór (repozytorium) wzorców medycznych (np. wzorców jednostek chorobowych). Stan zdrowia pacjenta określa zestaw parametrów medycznych $z \in \mathbf{R}^N$. Większość procesów diagnozowania medycznego można sprowadzić do „procesu rozpoznawania” właściwego wzorca jednostki chorobowej w zbiorze (repozytorium), rozumianego jako wyznaczenie „najbardziej podobnego” elementu ze zbioru Y do elementu z .

Ogólnie, korzystając ze zdefiniowanych cząstkowych charakterystyk podobieństwa, możemy ocenić podobieństwo każdego elementu $y \in Y$ do elementu z funkcją f^z zdefiniowaną następująco:

$$f^z : Y \rightarrow \mathbf{R}^K$$

taką, że

$$f^z(y) = f(y, z), \quad y \in Y, \quad (5.1)$$

gdzie K licznosci zbioru \mathbf{K} charakterystyk cząstkowych podobieństwa.

Obraz zbioru Y wyznaczony przez funkcję podobieństwa f^z nazywać będziemy przestrzenią ocen podobieństwa P .

$$P = f^z(Y). \quad (5.2)$$

Podobieństwo każdego elementu $y \in Y$ do elementu z określane jest wartościami charakterystyk:

$$f_k^z(y) \in \mathbf{R}^1, \quad k \in \mathbf{K}.$$

W ten sposób możliwe jest sformułowanie wielokryterialnego zadania optymalizacji rozpoznawania wzorców w postaci:

$$(Y, f^z, R), \quad (5.3)$$

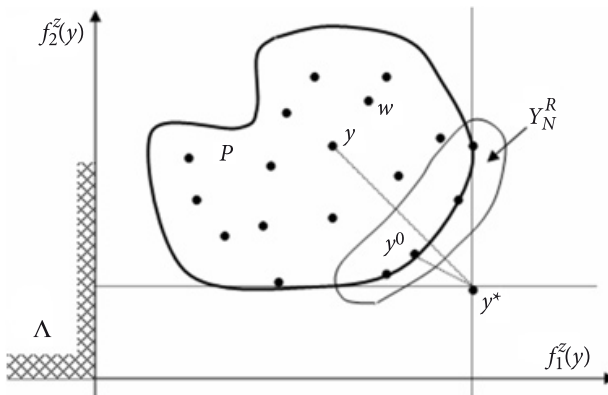
gdzie $R \subset P \times P$ — relacja podobieństwa [7].

$$R = \{(f^z(y), f^z(w)) \in P \times P \mid \text{„}y \text{ jest bardziej podobny do } z \text{ niż } w\text{”}\} \quad (5.4)$$

Rozwiązaniem zadania (5.3) jest zbiór Y_D^R wzorców najbardziej podobnych do z lub też zbiór Y_N^R wzorców, od których nie ma bardziej podobnych do z [3, 6].

Podejście takie umożliwia też wyznaczenie tzw. wzorca idealnego (utopijnego) oraz dokonanie rankingu wzorców ze zbioru Y w kontekście podobieństwa do wzorca idealnego [3], obraz tzw. idealnego podobieństwa.

Na rysunku 10 została przedstawiona przestrzeń ocen podobieństwa P , modelu z do wzorców $y \in Y$ przy założeniu, że stopień podobieństwa jest określony jedynie przez dwie charakterystyki podobieństwa (np. (4.2) i (4.5)).



Rys. 10. Przestrzeń ocen podobieństwa P oraz wzorec idealny

Relację podobieństwa w tym przypadku generuje stożek Λ , prowadzący do tzw. paretowskiej relacji podobieństwa [3]. Element y jest tak zwanym wzorcem idealnym (kresem dolnym zbioru Y), charakteryzującym kompleksowo cały zbiór Y w kontekście podobieństwa do elementu z . Zaznaczony na rysunku zbiór Y_N^R , to zbiór obrazów wzorców, od których nie ma bardziej podobnych do z spośród wzorców ze

zbioru Y . Wyznaczenie wzorca y_0^* umożliwia dokonanie rankingu wzorców ze zbioru Y od najbardziej podobnego y_0 do najmniej podobnego y_n , do badanego modelu z .

6. Podsumowanie

W pracy przedstawiono możliwość definiowania modeli graficznych danych medycznych w przestrzeni pajączynowej. Zdefiniowano charakterystyki podobieństwa graficznego, które pozwoliły sformułować zadanie wielokryterialnej optymalizacji rozpoznawania wzorców medycznych. Wśród zdefiniowanych w pracy cząstkowych charakterystyk podobieństwa znalazły się charakterystyki „mnogościowe”: (4.1), (4.2), (4.3) oraz wybrana charakterystyka metryczna (4.4). W procesie definiowania podobieństwa modeli graficznych możliwości tych jest znacznie więcej [5]. Wprowadzenie ogólnej relacji podobieństwa R w przestrzeni podobieństwa umożliwia bardzo elastyczne modelowanie podobieństwa modeli graficznych danych o dowolnych obiektach. Szczególnym przypadkiem zdefiniowanego w pracy wielokryterialnego modelu podobieństwa jest model podobieństwa Tversky'ego przedstawiony w pracach [16, 17]. Model Tversky'ego jest liniową funkcją (sumą ważoną) charakterystyk świadczących o wspólnych cechach porównywanych obiektów oraz charakterystyk świadczących o cechach różniących. Główną wadą takiego podejścia jest brak racjonalnych przesłanek pozwalających wyznaczać współczynniki wagowe tej funkcji. Model ten prowadzi bowiem do zadania jednokryterialnego, generującego rozwiązanie bardzo wrażliwe na zmiany współczynników wagowych. Stąd też rankingi podobieństwa otrzymywane na podstawie modelu Tversky'ego bardzo się różnią od siebie w zależności od subiektywnie przyjmowanych współczynników wagowych decydujących o roli cech wspólnych i odróżniających porównywanych obiektów. Model Tversky'ego nie uwzględnia również podobieństwa metrycznego obiektów.

Zdefiniowanie wielokryterialnej przestrzeni podobieństwa umożliwia badanie różnych relacji podobieństwa oraz obiektywne wyznaczenie podzbioru wzorców, od których nie ma bardziej podobnych (zbiór Pareto) w sensie przyjętej relacji podobieństwa.

Wyznaczenie wzorca idealnego, który jest kresem danym zbioru wzorców (charakterystyką obiektywną zbioru wzorców), umożliwia zdefiniowanie całej rodziny funkcji rankingowych (normy z parametrem $p \geq 1$), pozwalających dokonać rankingu wzorców wg podobieństwa.

Uzyskane wyniki mogą znaleźć zastosowanie w algorytmach rozpoznawania wzorców, które stanowią podstawowy moduł procesów diagnozowania nie tylko w zastosowaniach medycznych.

LITERATURA

- [1] M. ALBIN, *Fuzzy sets and their applications to medical diagnosis*’s, Berkeley, 1975.
- [2] M. ALLAN, *Crash Course — wywiad i badania przedmiotowe*, Elsevier Urban & Partner, Wrocław, 2005.
- [3] A. AMELJAŃCZYK, *Optymalizacja wielokryterialna w problemach sterowania i zarządzania*, Ossolineum, 1984.
- [4] A. AMELJAŃCZYK, *Matematyczny model przestrzeni życia w komputerowym systemie wspomagania decyzji medycznych*, I Krajowa Konferencja „Systemy Komputerowe i Teleinformatyczne w Służbie Zdrowia”, Warszawa, 2009.
- [5] A. AMELJAŃCZYK, *Matematyczne aspekty modelowania pajęczynowego obiektów*, Biul. ISI, 4, 2009.
- [6] A. AMELJAŃCZYK, *Wielokryterialne mechanizmy wspomagania podejmowania decyzji medycznych w modelu repozytorium w oparciu o wzorce*, Biul. ISI, 5, 2010.
- [7] A. AMELJAŃCZYK, *Modelowanie wzorców medycznych w przestrzeni pajęczynowej*, VII Konferencja Naukowa „Modelowanie Cybernetyczne Systemów Biologicznych”, MCSB2010, Kraków, 2010.
- [8] R. DOUGLAS COLLINS, *Algorytmy interpretacji objawów klinicznych*, Medipage, Warszawa, 2010.
- [9] F. KOKOT, *Diagnostyka różnicowa objawów chorobowych*, WL PZWL, Warszawa, 2007.
- [10] M. C. KUMAR, C. DODDS, *Podstawowe procedury diagnostyczno-lecznicze*, WL PZWL, Warszawa, 2007.
- [11] *Medyczne Systemy Ekspertowe*, <http://WWW.computer.privateweb.at/judith/links3.htm>
- [12] Z. PAWLAK, *Rough Sets*, *International Journal of Computer and Information Sciences*, 11, 1965, 341-356.
- [13] Z. PAWLAK, *Systemy informacyjne — podstawy teoretyczne*, WNT, Warszawa, 1983.
- [14] W. SIEGENTHALER, *Rozpoznanie różnicowe w medycynie wewnętrznej*, tom 1-2, Medipage, Warszawa, 2009.
- [15] P. SMETS, *Medical diagnosis fuzzy sets and degrees of belief*, *Fuzzy sets and Systems*, 5, 1981.
- [16] J. A. TVERSKY, *Features of Similarity*, *Psychological Review*, 84, 4, 1977.
- [17] J. A. TVERSKY, I. GATI, *Studies of Similarity*, Hebrew University, Jerusalem, 1978.
- [18] *The Merck Manual*, *Objawy kliniczne*, Elsevier Urban & Partner, Wrocław, 2010.
- [19] H. WECHSLER, *Applications of fuzzy logic to medical diagnosis*, Proc. Symp. on Multiple-Valued Logic, Logan, 1975.
- [20] PL. A. ZADEH, *Fuzzy Sets*, *Information and Control*, 8, 1965, 338-353.

A. AMELJAŃCZYK

Modelling of similarity of diagnostic medical patterns in the web space

Abstract. The paper presents the possibility of using web space for graphical modelling of disease unit patterns and status of the patient health. An algorithm provides a general idea of supporting the process diagnosing of the patient health, based on determining the similarity graphical models of the health status and contained in the repository of disease unit patterns. Multicriteria optimization task of medical pattern recognition has been formulated.

Keywords: web space, graphical model, medical pattern, similarity graphical models, similarity space, the relation of similarity, Tversky’s model