Agata Binderman
Katedra Ekonometrii i Informatyki, SGGW
e-mail: abinderman@mors.sggw.waw.pl

KLASYFIKACJA OBIEKTÓW OPARTA NA DWÓCH WZORCACH

Streszczenie: W pracy, podano sposób porządkowania obiektów (wierszy) w macierzy danych na podstawie dwóch wzorców. Proponowana metoda, do budowy syntetycznego miernika generującego porządek w zbiorze obiektów, wykorzystuje zarówno pojęcie wzorca, jak i funkcji użyteczności. Podane w pracy postacie funkcji użyteczności mają tą własność, że dwa obiekty, które są jednakowo odległe względem metryki Euklidesa od obiektu maksymalnego oraz obiektu minimalnego, mają tą samą użyteczność. W pracy zamieszczony został przykład, który pokazuje, że obiekt uznany za najgorszy według jednego wzorca może być najlepszy według drugiego wzorca.

Słowa kluczowe: mierniki syntetyczne, metryka, funkcja użyteczności, wzorzec, normalizacja, klasyfikacja.

WSTĘP

Do analizy zjawisk złożonych takich jak np.: rozwój gospodarczy i społeczny, poziom rozwoju i potencjał rolnictwa, ocena przedsiębiorstw oraz województw i gmin, poziom i jakość życia społeczeństwa, konieczne jest rozważenie wielu czynników ekonomicznych. Czynniki, które traktowane są jako zmienne objaśniające dane zjawisko, mogą być zarówno mierzalne jak i niemierzalne. Podanie ocen na podstawie tych danych ma na ogół charakter niejednoznaczny. Do oceny sumarycznej zjawisk złożonych stosuje się zmienne syntetyczne (agregatowe). [Zeliaś A. 1997]. Zastąpienie ciągu wielu zmiennych objaśniających badanego zjawiska przez zmienną syntetyczną daje pewną ocenę (niejednoznaczną) badanego zjawiska. Zmienne syntetyczne poza swą niejednoznacznością mają jeszcze taką wadę, że nie zawsze można im nadać interpretację merytoryczną. Istnieje wiele metod tworzenia zmiennych syntetycznych. Metody te można podzielić na wzorcowe i bezwzorcowe (metoda sum standaryzowanych wartości, pierwszego czynnika wspólnego), [Pociecha i in. 1988]. Metody wzorcowe zakładają istnienie pewnego hipotetycznego obiektu wzorcowego, uporządkowanie badanych obiektów dokonuje się w zależności od osiągniętych przez nich odległości od obiektu wzorcowego.

Metody te wykorzystują odpowiednio wybrane zmienne diagnostyczne (objaśniające), charakteryzujące badane zjawisko, różnią się miedzy sobą, co do sposobu normalizacji zmiennych oraz postaci funkcji je agregujących [Hellwig 1968; Bartosiewicz 1976; Borys 1978]. Wśród zmiennych objaśniających wyróżnia się zmienne, które działają w sposób pobudzający (tzw. stymulanty), podczas gdy

inne wpływają hamująco na rozwój badanego zjawiska (tzw. destymulanty). Przyjmijmy założenie, że zmiennymi stymulantami nazywać będziemy takie zmienne, których większe wartości świadczą o wyższym poziomie rozwoju badanego zjawiska, a zmiennymi destymulantami nazywać będziemy takie zmienne, których mniejsze wartości świadczą o wyższym poziomie rozwoju [zob. Borkowski B., Dudek H., Szczesny W. 2004; Zeliaś A. 2000]. Oczywiście poza stymulantami i destymulantami występują również nominanty - zmienne o trudnym do sprecyzowania sposobie oddziaływania na poziom rozwoju badanego zjawiska, jak również zmienne jakościowe. Określenie charakteru zmiennych opiera się na przesłankach merytorycznych. Przy braku odpowiedniej teorii można się posłużyć np. metodą opinii zespołu ekspertów.

Otrzymane w pracy rezultaty autorka wykorzystała do badania przestrzennego zróżnicowania polskiego rolnictwa [Binderman A. 2006]. Spośród wielu prac, poświęconym zastosowaniu wielowymiarowych metod porównawczych do badania struktur ekonomicznych regionów wymienić można prace [Zeliaś 2000; Malina 2004, Binderman 2005].

FUNKCJA UŻYTECZNOŚCI

W dalszej części rozważań załóżmy, że dane zjawisko jest opisane przez zmienne będące stymulantami. Osiągnąć to można poprzez eliminację zmiennych neutralnych, nadanie zmiennym jakościowym wartości liczbowych, przekształcenie destymulant w stymulanty (np. odwrócenie wartości destymulant). Bez straty dla ogólności rozważań, załóżmy również, że rozważane stymulanty po dokonaniu normalizacji i zmianie układu współrzędnych poprzez przesunięcie, mają wartości nieujemne. Przy takim podejściu dany obiekt (obserwacja) badanego zjawiska jest opisany za pomocą wektora, będącego elementem przestrzeni $\Re_+^n := \{\mathbf{x} = (x_1, x_2, \ldots, x_n): x_k \geq 0,\ k=1,2,...,n\}$, gdzie $n \geq 1$ jest ilością zmiennych zakwalifikowanych do oceny zjawiska. Do klasyfikacji danych obiektów obserwowanego zjawiska, przy pomocy mierników syntetycznych, wygodne może być użycie aparatu matematycznego stworzonego w teorii ekonomii dobrobytu (popytu) [zob. Allen R. 1961; Panek E. 2000, 2003]. W teorii tej opisane jest pojęcie funkcji użyteczności i przyjmuje się, że indywidualna użyteczność badanego obiektu jest mierzalna.

Rozważmy teraz problem polegający na klasyfikacji $m \in N$ obiektów badanego zjawiska za pomocą $n \in N$ zmiennych. Zgodnie z przyjętymi wcześniej założeniami każdy taki obiekt daje się przedstawić za pomocą wektora należącego do przestrzeni $\Re^n_+$. Niech wektor $\mathbf{x}_i=(x_{i1}, x_{i2}, \ldots, x_{in})$, i=1,2,...,m, opisuje i-ty obiekt.

Jeżeli $x_{ik}>x_{jk}$ ($x_{ik}\geq x_{jk}$) dla k=1,2,...,n to pisać będziemy $\mathbf{x}_i>\mathbf{x}_j$, ($\mathbf{x}_i\geq\mathbf{x}_j$), gdzie i, $j \in [1,m]$. Nietrudno zauważyć, że jeżeli $\mathbf{x}_i>\mathbf{x}_j$ i $\mathbf{x}_i\neq\mathbf{x}_j$ to naturalnym jest nazywać obiekt $\mathbf{x}_i$ lepszym (wyżej ocenianym) od obiektu $\mathbf{x}_j$. Istotnie oznacza to, że żadna ze

składowych wektora $\mathbf{x}_i$ nie jest mniejsza od odpowiednich składowych wektora $\mathbf{x}_j$, a przynajmniej jedna z nich ma wartość większą, tj. istnieje takie $k \in [1,n]$, że $x_{ik} > x_{jk}$. Z tego względu, w celu uporządkowania rozważanych obiektów przyjmijmy następującą definicję funkcji użyteczności będącą liczbową charakterystyką naszych preferencji (porównaj z definicją funkcji użyteczności w teorii popytu w warunkach niedosytu [Panek 2000, 2003]).

DEFINICJA 1. Każdą rosnącą funkcję u: $\Re_+^n \to \Re$ nazywać będziemy *funkcją użyteczności*. Z definicji wynika, że dla dowolnej pary wektorów $\mathbf{x}, \mathbf{y} \in \Re_+^n$ spełniona jest implikacja:

$$\mathbf{x} \geq \mathbf{y} \wedge \mathbf{x} \neq \mathbf{y} \Rightarrow u(\mathbf{x}) > u(\mathbf{y}).$$

Dlatego też w dalszej części pracy obiekt **x** uważać będziemy za lepszy od obiektu **y,** jeżeli u(**x**)>u(**y**), oznacza to, że obiekt lepszy od drugiego obiektu ma większą od niego użyteczność. Fakt ten zapisywać będziemy w następujący sposób: $\mathbf{y} \prec \mathbf{x} (\mathbf{x} \succ \mathbf{y})$. Obiekty **x**, **y** uważać będziemy za jednakowo dobre (obojętne), względem przyjętej funkcji użyteczności u, jeżeli u(**x**) = u(**y**). Fakt ten zapisywać będziemy w następujący sposób: **y**~**x** (**x**~**y**) W pierwszym przypadku mówić będziemy, że obiekt **x** jest *silnie preferowany* nad **y**, w drugim, że obiekty **y** i **x** są *indyferentne*. Jeżeli obiekty **y** i **x** są indyferentne lub obiekt **x** jest silnie preferowany nad **y** to mówić będziemy, że obiekt **x** jest *słabo preferowany* nad **y.** Symbol $\mathbf{x} \succsim \mathbf{y}$ lub $\mathbf{y} \precsim \mathbf{x}$ oznaczać będzie alternatywę: **x** $\succ$**y** lub **x**~**y**.

Zdefiniowane powyżej związki miedzy obiektami wyznaczają odpowiednio relację silnej preferencji, relację indyferencji oraz relację preferencji (słabej) [zob. Panek 2000, 2003]. Oczywiście w teorii popytu relacja preferencji konsumenta może indukować funkcję użyteczności. Przyjmując jakąkolwiek postać funkcji użyteczności przesądzamy istnienie relacji preferencji, którą ta funkcja opisuje.

WYKORZYSTANIE FUNKCJI UŻYTECZNOŚCI DO KLASYFIKACJI DANYCH

Przyjmijmy dla obiektów wzorcowych następujące oznaczenia:

$\mathbf{x}_0 := (x_{0,1}, x_{0,2}, ..., x_{0,n})$, $\mathbf{x}_{m+1} := (x_{m+1,1}, x_{m+1,2}, ..., x_{m+1,n})$, gdzie

$$x_{0,k} := \min_{1 \leq i \leq m} x_{ik}, \quad x_{m+1,k} := \max_{1 \leq i \leq m} x_{ik}, \quad k = 1, 2, ..., n.$$

Oczywistym jest, że tak określone obiekty $\mathbf{x}_0$, $\mathbf{x}_{m+1}$ (być może fikcyjne) są odpowiednio, niegorsze, nielepsze od pozostałych $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$, tj. $\mathbf{x}_{m+1} \geq \mathbf{x}_i$ oraz $\mathbf{x}_i \geq \mathbf{x}_0$, dla każdego i: $m \geq i \geq 1$.

W przypadku, gdy $\mathbf{x}_0$ i $\mathbf{x}_{m+1}$ są różne od rozważanych obiektów $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$, to spełniają one następującą rolę: $\mathbf{x}_{m+1}$ **obiektu najlepszego**, $\mathbf{x}_0$ **obiektu najgorszego**. Obiekty te traktowane będą, jako **wzorce**. W tym przypadku, jeżeli znana jest

funkcja użyteczności u określona na przestrzeni wektorów $\Re_+^n$ to dla każdego $i \in [1,m]$: $u(\mathbf{x}_{m+1}) > u(\mathbf{x}_i)$ oraz $u(\mathbf{x}_0) < u(\mathbf{x}_i)$. Można przyjąć różne kryteria wyboru funkcji użyteczności, aby przy jej pomocy ustalić relację porządku liniowego [zob. Panek 2000, 2003, Malawski 1999] określoną na iloczynie kartezjańskim $W \times W \subset \Re_+^n \times \Re_+^n$, gdzie $W:=\{\mathbf{x}_o,\mathbf{x}_1,...,\mathbf{x}_{m+1}\}$.

Kryterium przyjęte w pracy opiera się na pojęciu odległości pomiędzy rozważnymi obiektami. Pojęcie odległości między dwoma obiektami wiąże się bezpośrednio z koniecznością normalizowania zmiennych, które wyrażone są w różnych jednostkach fizycznych. Normalizacja ta polega na przekształceniu wartości zmiennych wyrażonych w różnych jednostkach w celu doprowadzenia ich do wzajemnej porównywalności. W literaturze przedmiotu wyróżnia się normalizację cech poprzez przekształcenie **ilorazowe**, **standaryzację** i **unitaryzację** [zob. Kukuła. 2000; Gatnar 1998; Zeliaś 2000; Strahl, Walesiak 1996, 1997]. Normalizowanie jest konieczne do konstrukcji mierników syntetycznych. Pojęcie **metryki** (**odległości**) odgrywa fundamentalną rolę w badaniach ekonomicznych, szczególnie przy porównywaniu struktury ekonomicznej regionów [Stone 1970, Zeliaś 2002; Malina 2004]. Niech $X=\Re^n$, to *metryką* nazywamy każdą funkcję $d: X \times X \to \Re_+=[0,+\infty)$ spełniającą następujące trzy warunki:

1) $d(\mathbf{x},\mathbf{y}) = 0 \Leftrightarrow \mathbf{x}=\mathbf{y}$,
2) $d(\mathbf{x},\mathbf{y})=d(\mathbf{y},\mathbf{x})$,
3) $d(\mathbf{x},\mathbf{y}) \le d(\mathbf{x},\mathbf{z})+d(\mathbf{z},\mathbf{y})$, dla wszystkich **x**, **y**, **z**, należących do X.

Metryka każdej parze wektorów przyporządkowuje liczbę nieujemną, zwaną odległością między nimi, liczbę $d(\mathbf{x},\mathbf{y})$ nazywa się *odległością wektora* **x** od wektora **y**. W ekonomii stosuje się bardzo wiele różnych metryk, przykłady ich można znaleźć w pracach [Rolewicz 1985; Zeliaś 2002; Kukuła 2000] i innych, są nimi na przykład:

$$d_k(\mathbf{x},\mathbf{y}):=(|x_1-y_1|^k+...+|x_n-y_n|^k)^{1/k},\ k\ge 1,\ d_\infty(\mathbf{x},\mathbf{y}):=\max\{|x_1-y_1|,...,|x_n-y_n|\},$$

gdzie: $\mathbf{x}=(x_1,x_2,....,x_n)$, $\mathbf{y}=(y_1,y_2,...y_n) \in X$.

Metryki te są często nazywane metrykami Minkowskiego. Dla k=1 jest to tzw. metryka liniowa Hamminga (miejska, uliczna, metropolitarna), dla k=2 to metryka Euklidesa, natomiast dla $k=\infty$ metryka Czebyszewa. Autorka proponuje w pracy przyjąć naturalne kryterium, według którego dwa obiekty o identycznych odległościach od obiektu najlepszego i najgorszego byłyby względem siebie obojętne, tj. miały tę samą użyteczność. Jeżeli zatem $d(\mathbf{x}_i,\mathbf{x}_j)$ $(i,j \in [1,m])$ oznacza odległość między obiektami $\mathbf{x}_i$ i $\mathbf{x}j$ to $\mathbf{x}_i \sim \mathbf{x}_j \Leftrightarrow u(\mathbf{x}_i) = u(\mathbf{x}_j) \Leftrightarrow d(\mathbf{x}_i, \mathbf{x}_o) = d(\mathbf{x}_j, \mathbf{x}_o) \wedge d(\mathbf{x}_i, \mathbf{x}_{m+1}) = d(\mathbf{x}_j, \mathbf{x}_{m+1})$. Funkcją użyteczności, która spełnia powyższy warunek, przy wyborze odległości na podstawie metryki Euklidesa jest np. funkcja liniowa będąca iloczynem skalarnym wektorów $\mathbf{x}_{m+1}-\mathbf{x}_o$ i $\mathbf{x}_i$ tj.

$$u(x_i) := \langle x_{m+1} - x_0, x_i \rangle = \sum_{k=1}^{n} (x_{m+1,k} - x_{0,k}) x_{ik}; \qquad (1)$$

gdzie i = 0,1,2,...,m+1. Dla tak określonej funkcji użyteczności słuszne jest następujące twierdzenie.

TWIERDZENIE 1. Dwa obiekty $\mathbf{x}_i$, $\mathbf{x}_j$ mają tą samą użyteczność wtedy i tylko wtedy, gdy różnica kwadratów ich odległości (według metryki Euklidesa) od obiektu najlepszego jest równa różnicy kwadratów ich odległości od obiektu najgorszego tj.

$$u(\mathbf{x}_i)=u(\mathbf{x}_j)\Leftrightarrow d^2(\mathbf{x}_i,\mathbf{x}_{m+1})-d^2(\mathbf{x}_j,\mathbf{x}_{m+1})=d^2(\mathbf{x}_i,\mathbf{x}_o)-d^2(\mathbf{x}_j,\ \mathbf{x}_o),$$

gdzie: $d^2(\mathbf{x}_i,\mathbf{x}_j)=\sum_{k=1}^{n}(\mathbf{x}_{ik}-\mathbf{x}_{jk})^2$; $i,j\in[0,m+1]$.

DOWÓD 1. Warunek wystarczający.
Niech $u(\mathbf{x}_i)= u(\mathbf{x}_j)$, wówczas $\langle x_{m+1} - x_0, x_i\rangle = \langle x_{m+1} - x_0, x_j\rangle$ czyli:

$$\sum_{k=1}^{n}(x_{m+1,k} - x_{0,k})x_{ik} = \sum_{k=1}^{n}(x_{m+1,k} - x_{0,k})x_{jk} \text{ oraz } \sum_{k=1}^{n}(x_{m+1,k} - x_{0,k})(x_{ik} - x_{jk}) = 0.$$

Stąd: $\sum_{k=1}^{n}(x_{ik} - x_{jk})(x_{ik} + x_{jk} - 2x_{m+1,k} - x_{ik} - x_{jk} + 2x_{0,k}) = 0$,

$$\sum_{k=1}^{n}(x_{ik} - x_{jk})(x_{ik} + x_{jk} - 2x_{m+1,k}) = \sum_{k=1}^{n}(x_{ik} - x_{jk})(x_{ik} + x_{jk} - 2x_{0,k}),$$

$$\sum_{k=1}^{n}\{(x_{ik})^2 - 2x_{m+1,k}(x_{ik} - x_{jk}) - (x_{jk})^2\} = \sum_{k=1}^{n}\{(x_{ik})^2 - 2x_{0,k}(x_{ik} - x_{jk}) - (x_{jk})^2\},$$

$$\sum_{k=1}^{n}\{(x_{ik} - x_{m+1,k})^2 - (x_{jk} - x_{m+1,k})^2\} = \sum_{k=1}^{n}\{(x_{ik} - x_{0,k})^2 - (x_{jk} - x_{0,k})^2\},$$

czyli $d^2(x_i, x_{m+1}) - d^2(x_j, x_{m+1}) = d^2(x_i, x_o) - d^2(x_j, x_o)$. W podobny sposób dowodzi się warunku koniecznego. Z powyższego twierdzenia wynika ważny wniosek.

WNIOSEK Jeżeli obiekty $\mathbf{x}_i$, $\mathbf{x}_j$ są jednakowo odległe od obiektu maksymalnego $\mathbf{x}_{m+1}$ oraz obiektu minimalnego $\mathbf{x}_0$, tj. $d(\mathbf{x}_i, \mathbf{x}_{m+1}) = d(\mathbf{x}_j, \mathbf{x}_{m+1})$, $d(\mathbf{x}_i, \mathbf{x}_o) = d(\mathbf{x}_j, \mathbf{x}_o)$ to obiekty te mają tą samą użyteczność. Istotnie, jeżeli:

$$d(\mathbf{x}_i, \mathbf{x}_{m+1}) = d(\mathbf{x}_j, \mathbf{x}_{m+1}) \text{ oraz } d(\mathbf{x}_i, \mathbf{x}_o) = d(\mathbf{x}_j, \mathbf{x}_o) \text{ to}$$

$$d^2(\mathbf{x}_i, \mathbf{x}_{m+1}) - d^2(\mathbf{x}_j, \mathbf{x}_{m+1}) = d^2(\mathbf{x}_i, \mathbf{x}_o) - d^2(\mathbf{x}_j, \mathbf{x}_o) = 0, \text{ stąd } u(\mathbf{x}_i) = u(\mathbf{x}_j).$$

Jeżeli zatem dwa obiekty są jednakowo oddalone względem metryki Euklidesa od obiektu najlepszego $\mathbf{x}_{m+1}$ i obiektu najgorszego $\mathbf{x}_0$ to ich użyteczności są identyczne.

UWAGA 1. Powierzchniami obojętności (u(x)=constans) generowanymi przez funkcję użyteczności u określoną za pomocą wzoru (1), są hiperpłaszczyzny: proste dla n=2, płaszczyzny dla n=3.

DEFINICJA 2. Układ wektorów $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m, \mathbf{x}_{m+1}$ nazywać będziemy znormalizowanym jeżeli obiekty $\mathbf{x}_0, \mathbf{x}_{m+1}$ są reprezentowane przez wektor zerowy i jednostkowy tj.

$$\mathbf{x}_0 = \mathbf{0} = (0,0,...,0),\ \mathbf{x}_{m+1} = \mathbf{1} = (1,1,...,1).$$

Oczywiście, jeżeli układ wektorów $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m, \mathbf{x}_{m+1}$ jest znormalizowany to $0 \le \mathbf{x}_{i,k} \le 1$ dla każdego i = 0,1,...,m+1; k=1,2,...,n.

Ze wzoru (1) wynika:

UWAGA 2. Jeżeli układ wektorów $x_0, x_1, x_2, ..., x_m, x_{m+1}$ jest znormalizowany to:

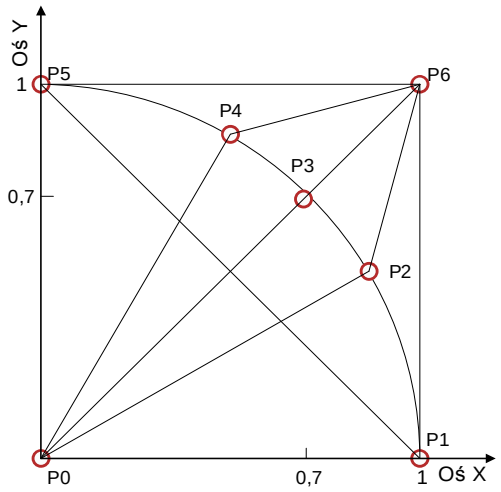
$$u(\mathbf{x}_0)=0,\ u(\mathbf{x}_{m+1})=n,\ u(\mathbf{x}_i)=\sum_{k=1}^{n} x_{i,k} \qquad (2)$$

PRZYKŁAD 1. Rozważmy na płaszczyźnie zmiennych rzeczywistych $0xy$, siedem punktów o współrzędnych:

$$P0=(0;0), P1=(1;0),\ P2=\left(\frac{\sqrt{3}}{2};\frac{1}{2}\right), P3=(0,7;0,7), P4=\left(\frac{1}{2};\frac{\sqrt{3}}{2}\right), P5=(0;1), P6=(1;1).$$

Położenie punktów na płaszczyźnie ilustruje poniższy rysunek.

Rys 1 Geometryczna interpretacja przykładu



Źródło: Opracowanie własne

Załóżmy, że punkty te opisują odpowiednio pewne badane obiekty $\mathbf{w}_0$, $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{w}_4$, $\mathbf{w}_5$, $\mathbf{w}_6$, charakteryzowane za pomocą pary cech będących stymulantami. Obiekt $\mathbf{w}_0$ opisywany za pomocą punktu $\mathbf{P}_0$ jest uznany za najgorszy, obiekt zaś $\mathbf{w}_6$ jest najlepszy. Ocena pozostałych obiektów jest dokonywana według odległości (obliczonej przy pomocy metryki Euklidesa) punktów $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$, $\mathbf{P}_5$ najpierw od punktu $\mathbf{P}_0$ potem od punktu $\mathbf{P}_6$. Przyjmijmy, że obiekt jest uważany za lepszy od

drugiego obiektu, gdy jest: położony dalej od najgorszego - według pierwszego kryterium, położony bliżej najlepszego obiektu- według drugiego kryterium. Łatwo zauważyć, że przy takich kryteriach obiekt $\mathbf{w}_3$ jest najgorszy według pierwszego kryterium, lecz jest najlepszy według drugiego kryterium, spośród obiektów $\mathbf{w}_5$, $\mathbf{w}_4$, $\mathbf{w}_3$, $\mathbf{w}_2$, $\mathbf{w}_1$. Jeżeli na rozważanym zbiorze obiektów określimy funkcję użyteczności u, określoną za pomocą wzoru (1), to ich użyteczności przedstawiają się następująco: $u(\mathbf{w}_0)=0$; $u(\mathbf{w}_1)=u(\mathbf{w}_5)=1$; $u(\mathbf{w}_2)=u(\mathbf{w}_4)=1{,}38$; $u(\mathbf{w}_3)=1{,}4$; $u(\mathbf{w}_6)=2$. Według kryterium większej użyteczności obiektu, mamy następujące uporządkowanie: $\mathbf{w}_0 \prec \mathbf{w}_1 \sim \mathbf{w}_5 \prec \mathbf{w}_2 \sim \mathbf{w}_4 \prec \mathbf{w}_3 \prec \mathbf{w}_6$.

Na odcinku wyznaczonym przez punkty $\mathbf{P}_1$ i $\mathbf{P}_5$ leżą obiekty, które mają tę samą użyteczność, równą 1. Odcinek ten jest tak zwaną krzywą obojętności generowaną przez funkcje użyteczności u (zob. Panek E. 2000). Nietrudno zauważyć, że w tym przypadku krzywe obojętności są odcinkami prostych o równaniach $x+y=c$, $0< c <2$. Istotnie, jeżeli punkty $\mathbf{P}(s,t)$, $\mathbf{Q}(v,z)$, $0 < s,t,v,z < 1$ leżą na prostej o równaniu $x + y = c$, to: $u(P)=\langle(1,1),(t,s)\rangle = t+s = c$, $u(Q)=\langle(1,1),(v,z)\rangle = v+z = c$. Oznacza to, że $\mathbf{P} \sim \mathbf{Q}$.

Powyższe rozważania pokazują, że *wybór wzorca odgrywa istotną rolę dla rankingów*, jak również przy grupowaniu obiektów. Podany w przykładzie sposób porządkowania obiektów za pomocą funkcji użyteczności *opiera się na dwóch wzorcach*.

Jeżeli dana funkcja użyteczności u indukuje relacje preferencji obiektów zbioru W to funkcja złożona g(u(x)), gdzie $g: \Re \rightarrow \Re$ jest funkcją rosnącą, jest również funkcją użyteczności, generującą tą samą relację preferencji w zbiorze obiektów W co funkcja u.

Wykorzystując powyższą własność celowe jest unormowanie funkcji użyteczności polegające na wybraniu takiej funkcji g, aby jej wartość dla obiektu najgorszego wynosiła 0, wartość zaś dla obiektu najlepszego wynosiła 1, to jest by:

$$g(u(\mathbf{x}_0))=0,\ g(u(\mathbf{x}_{m+1}))=1.$$

Funkcją o tej własności może być na przykład funkcja liniowa:

$$g(t)=(t-t_0)/(t_1-t_0),\ t\in[t_0, t_1], \qquad (3)$$

gdzie $t_1= u(\mathbf{x}_{m+1})$, $t_0=. u(\mathbf{x}_0)$. Przy oczywistym założeniu, że $t_1 > t_0$, gdyż dopuszczenie przypadku $t_1 = t_0$ oznaczałoby, że wszystkie rozważane obiekty mają tą samą użyteczność, otrzymaną za pomocą funkcji u. W tym przypadku funkcja:

$$f(\mathbf{x}_i):=g((u(\mathbf{x}_i)) = (u(\mathbf{x}_i) - u(\mathbf{x}_0) / (u(\mathbf{x}_{m+1}) - u(\mathbf{x}_0)),\ i=0,1,\ldots,m+1, \qquad (4)$$

określona na zbiorze obiektów W, jest funkcją użyteczności mającą tą własność, że $f(\mathbf{x}_0)=0$, $f(\mathbf{x}_{m+1})=1$. Jeżeli funkcja użyteczności u określona jest za pomocą wzoru (1) to przy pomocy wzoru (4) otrzymujemy postać nowej funkcji użyteczności:

$$f(\mathbf{x}_i) = \left[\langle \mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{x}_i\rangle - \langle \mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{x}_0\rangle - \langle \mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{x}_{m+1}\rangle - \langle \mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{x}_0\rangle\right]^{-1} =$$

$$= \| \mathbf{x}_{m+1} - \mathbf{x}_0 \|^{-2} \left\langle \mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{x}_i - \mathbf{x}_0 \right\rangle, \; i=0,1,\ldots,m+1, \qquad (5)$$

gdzie norma wektora $\| \mathbf{x}_i \| := (\langle \mathbf{x}_i, \mathbf{x}_i \rangle)^2$ (zob. Panek E., 2000, 2003). Funkcja f indukuje tą samą relację preferencji, co funkcja użyteczności u, oczywiście $f(\mathbf{x}_0)=0$, $f(\mathbf{x}_{m+1})=1$.

Jeżeli układ wektorów $x_0, x_1, x_2, \ldots, x_m, x_{m+1}$ jest znormalizowany to:

$$f(\mathbf{x}_i) = \frac{1}{n} \sum_{k=1}^{n} x_{i,k}, \quad i=0,1,\ldots,m+1. \qquad (6)$$

Wzór (2) jak i wzór (6) należą do najczęściej stosownych addytywnych formuł agregacyjnych [por. Cieślak M. 1993, Kukuła K. 2000].

Zauważmy ponadto, że jeżeli wektor $s:=(\mathbf{x}_{m+1}+\mathbf{x}_0)/2$ reprezentuje obiekt „pośredni" pomiędzy najlepszym $\mathbf{x}_{m+1}$ a najgorszym $\mathbf{x}_0$ to $f(\mathbf{s})=1/2$. Istotnie

$$f(\mathbf{s}) = \| \mathbf{x}_{m+1} - \mathbf{x}_0 \|^{-2} \left\langle \mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{s} - \mathbf{x}_0 \right\rangle = \| \mathbf{x}_{m+1} - \mathbf{x}_0 \|^{-2} \left\langle \mathbf{x}_{m+1} - \mathbf{x}_0, (\mathbf{x}_{m+1} - \mathbf{x}_0)/2 \right\rangle$$

$$= 0{,}5 \| \mathbf{x}_{m+1} - \mathbf{x}_0 \|^{-2} \left\langle \mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{x}_{m+1} - \mathbf{x}_0 \right\rangle = 0{,}5.$$

Istnieje skończenie wiele rosnących funkcji g jednej zmiennej, spełniających warunki:

$$g(t_0)=0,\ g(t_1)=1,\ g((t_1+t_0)/2)=1/2;\ \ t_0, t_1 \in \Re,\ \ t_0 \neq t_1 .$$

Oczywiście, istnieje również nieskończenie wiele funkcji, które będąc funkcjami rosnącymi mają te same wartości dla obiektów, które są jednakowo odległe od obiektu maksymalnego $\mathbf{x}_{m+1}$ oraz obiektu minimalnego $\mathbf{x}_0$. Przykładem jest funkcja określona w poniższym twierdzeniu.

TWIERDZENIE 2. Funkcja:

$$U(\mathbf{x}_i) = \frac{d(\mathbf{x}_0, \mathbf{x}_i) + d(\mathbf{x}_0, \mathbf{x}_{m+1}) - d(\mathbf{x}_i, \mathbf{x}_{m+1})}{2d(\mathbf{x}_0, \mathbf{x}_{m+1})}, \; i = 0,1,\ldots,m,m+1, \qquad (7)$$

jest funkcją użyteczności przyjmującą wartości z przedziału [0,1], przy czym $U(\mathbf{x}_0)=0$, $U(\mathbf{x}_{m+1})=1$. Jeżeli $d(\mathbf{x}_i,\mathbf{x}_{m+1})=d(\mathbf{x}_j,\mathbf{x}_{m+1})$ i $d(\mathbf{x}_i,\mathbf{x}_o) = d(\mathbf{x}_j,\mathbf{x}_o)$ to $U(\mathbf{x}_i)=U(\mathbf{x}_j)$.

Oczywistym jest, że powierzchnie obojętności generowane przez równanie $U(\mathbf{x})$=constans, nie będą, jak w przypadku funkcji liniowej u hiperpłaszczyznami. Warto jednak zauważyć, że w przypadku n=2 krzywa obojętności wyznaczona przez równanie $U(\mathbf{x})=c$, $c \in [0,1]$ jest odcinkiem prostej o równaniu $x_2=1-x_1$ (przekątną kwadratu) dla c=0,5 oraz hiperbolą dla pozostałych c. Wynika to z faktu, że hiperbola jest miejscem geometrycznym punktów, których różnica odległości od dwóch stałych punktów zwanych ogniskami jest stała.

PRZYKŁAD 2 Rozważmy zbiór siedmiu obiektów $\mathbf{w}_0$, $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{w}_4$, $\mathbf{w}_5$, $\mathbf{w}_6$, z poprzedniego przykładu, charakteryzowanych za pomocą pary cech będących

stymulantami. Obiekty te traktowane, jako wektory, tworzą znormalizowany układ wektorów (Definicja 2). Poniższa tabela podaje ich użyteczności obliczone za pomocą dwóch znormalizowanych funkcji użyteczności.

Tabela 1. Użyteczności obiektów

Funkcja użyteczności	w_0	w_1	w_2	w_3	w_4	w_5	w_6
$f(w_i)$-wzór (5)	0	0,5	0,683	0,7	0,683	0,5	1
$U(w_i)$-wzór (7)	0	0,5	0,671	0,7	0,671	0,5	1

Źródło: Opracowanie własne

Łatwo zauważyć, że funkcje te zachowują preferencje z przykładu pierwszego.

WNIOSKI

Przedstawione w pracy rozważania pokazują, że przy klasyfikacji obiektów wybór wzorca odgrywa istotną rolę. Zaproponowany sposób porządkowania liniowego obiektów w równym stopniu wykorzystuje obiekt wzorcowy najgorszy jak i obiekt wzorcowy najlepszy. Zaprezentowane w pracy podejście do problemu klasyfikacji obiektów nie wyczerpuje badań w tym zakresie a przydatność metody zweryfikować mogą tylko badania oparte na rzeczywistych danych.

LITERATURA

Allen R. G. D. (1964) Ekonomia matematyczna, PWN, Warszawa.
Bartosiewicz S. (1976) Propozycja metody tworzenia zmiennych syntetycznych, Prace Naukowe AE we Wrocławiu, nr 84, Wrocław.
Binderman A. (2005) O problemie wyboru wzorca przy badaniu przestrzennego zróżnicowania potencjału rolnictwa w Polsce, Metody ilościowe w badaniach ekonomicznych – V, Warszawa, str. 46.
Binderman A. (2006) Wykorzystanie funkcji użyteczności do badania przestrzennego zróżnicowania rolnictwa-praca złożona do Roczników Naukowych Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu.
Borkowski B, Dudek H., Szczesny W. (2004) Ekonometria. Wybrane zagadnienia, PWN, Warszawa.
Borys T. (1978) Propozycja agregatowej miary rozwoju obiektów, „Przegląd Statystyczny", z. 3.
Cieślak M. (1993) Ekonomiczne zastosowanie mierników syntetycznych ze zmiennym wzorcem, [w:] Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych, AE, Kraków.
Gantar E. (1998) Symboliczne metody klasyfikacji danych, PWN, Warszawa.
Hellwig Z. (1968) Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę kwalifikowanych kadr, „Przegląd Statystyczny", z. 4.
Kukuła K. (2000) Metoda unitaryzacji zerowanej, PWN, Warszawa.
Malawski A. (1999) Wprowadzenie do ekonomii matematycznej, AE, Kraków.

Malina A. (2004) Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw, AE, Seria Monografie nr 162, Kraków.
Panek E. (2000) Ekonomia matematyczna, Akademia Ekonomiczna, Poznań.
Panek E. (red.) (2003) Podstawy ekonomii matematycznej, AE, Poznań.
Pociecha J., Podolec B., Sokołowski A., Zając K. (1988) Metody taksonomiczne w badaniach społeczno-ekonomicznych, PWN, Warszawa.
Rolewicz S. (1985) Metric linear spaces, PWN-Polish Scientific Publishers and D. Reidel, Warszawa-Dordrecht.
Stone R., (1970) Matematyka w naukach społecznych, PWE, Warszawa.
Strahl D., Walesiak M. (1996) Normalizacja zmiennych w skali przedziałowej i ilorazowej w referencyjnym systemie granicznym, Seria: Taksonomia, z. 3, Sekcja Klasyfikacji i Analizy Danych, Wrocław – Kraków - Jelenia Góra.
Strahl D., Walesiak M. (1997) Normalizacja zmiennych w granicznym systemie referencyjnym, „Przegląd Statystyczny", z. 1.
Zegar J. (2003) Zróżnicowanie regionalne rolnictwa, GUS, Warszawa.
Zeliaś A. (1997) Teoria prognozy, PWE, Warszawa.
Zeliaś A. (2000) Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym, Kraków.

On a classification of objects basing on two models

Summary: In the present paper, a manner of classification of objects which is based on two model objects is given. The applied method uses comparative multidimensional analysis and conception of models, normalization, preference relations and utility functions as the preference indicators. The given utility functions have such property that two considered objects have an identical utility if their distances from two different fixed model objects are equal.

Key words: synthetic measures, metrics, utility function, model, normalization, classification.