Biuletyn WAT
Vol. LXIII, Nr 2, 2014

Metryki Minkowskiego w tworzeniu uniwersalnych algorytmów rankingowych

ANDRZEJ AMELJAŃCZYK

Wojskowa Akademia Techniczna, Wydział Cybernetyki, Instytut Systemów Informatycznych,
00-980 Warszawa, ul. gen. S. Kaliskiego 2, aameljanczyk@wat.edu.pl

Streszczenie. W pracy przedstawiono ogólną procedurę tworzenia rankingów zbioru obiektów przy relacji preferencji bazującej na dowolnej funkcji rankingowej. Analizę możliwych do wykorzystania funkcji rankingowych rozpoczęto od pokazania fundamentalnych wad powszechnie stosowanych funkcji w postaci sumy ważonej. Jako szczególny przypadek procedury rankingowej w przestrzeni z relacją zaproponowano procedurę bazującą na pojęciu elementu idealnego i uogólnionej odległości Minkowskiego od tego elementu. Procedura ta przedstawiona w postaci uniwersalnego algorytmu rankingowego eliminuje większość wad funkcji rankingowej w postaci sumy ważonej.
Słowa kluczowe: funkcje rankingowe, relacja preferencji, klastry rankingowe, punkt idealny, uniwersalny algorytm rankingowy

1. Wprowadzenie

Zagadnienie tworzenia rankingów zbioru obiektów ocenianych wielokryterialnie jest bardzo powszechne w wielu dziedzinach życia. Różni się od klasycznego zagadnienia optymalizacji tym, że określa „jakość" wszystkich obiektów od „najlepszego do najgorszego", a nie wyłącznie element „najlepszy", co jest domeną optymalizacji [1, 2, 7]. Typowe rankingi, konstruowane na potrzeby komercyjne czy też społeczne, tworzone są na ogół w oparciu o tzw. funkcje rankingowe lub też heurystycznie, a nawet intuicyjnie, bazując na bardzo prostych metodach i narzędziach. Nie mają one zatem wystarczającego waloru obiektywizmu i często służą jedynie celom marketingowym, a nawet swego rodzaju manipulacjom. W wielu obszarach życia jest to jednak niedopuszczalne i bardzo szkodliwe. Takimi przykładami są np. procedury ustalania rankingu ofert w zamówieniach publicznych (przetargach) i wszelkich innych podobnych przedsięwzięciach, jak konkursy projektów, grantów itp., w których wynik rankingu ma konsekwencje finansowe, prestiżowe lub handlowe. Powszechnie stosowane metody polegające na tworzeniu tzw. „ocen ważonych" czy też „ważonych decyzji ekspertów" są obarczone subiektywizmem i możliwością kamuflowanej manipulacji. Łatwo można pokazać, że dobierając odpowiednie wartości współczynników wagowych lub wartości progów, można istotnie zmieniać ranking wynikowy. Celem niniejszej pracy jest pokazanie metody tworzenia rankingów odpornych na możliwość manipulacji, obiektywnie wynikających z przyjętego modelu preferencji zdefiniowanego w przestrzeni kryteriów oceny poszczególnych obiektów.

2. Podstawowe wady funkcji rankingowych typu „suma ważona"

Do sformułowania zadania rankingu zbioru można podchodzić na wiele sposobów [3, 4, 5, 8]. Bardzo powszechną i szeroko stosowaną klasę metod rankingowych stanowią metody oparte na funkcji rankingowej [5].

Niech X — zbiór obiektów podlegających rankingowi

$F(x) = y \in \mathbf{R}^N$ — wielokryterialna ocena (model „jakościowy") obiektu, mająca postać:

$$F(x) = \big(F_1(x),\ldots,F_n(x),\ldots,F_N(x)\big) \in \mathbf{R}^N,$$

gdzie: $F_n(x)$ — ocena obiektu $x \in X$ w sensie wskaźnika nr $n \in N$;
$F(X) = \left\{F(x) = y \in \mathbf{R}^N \middle| x \in X\right\} = Y$ — obraz jakościowy zbioru obiektów X.

Ogólnie funkcją rankingową nazywamy dowolną funkcję postaci: $f : Y = F\big(X\big) \to \mathbf{R}^1$, taką, że jeśli dla dowolnych $y^1, y^2 \in Y$ zachodzi $f(y^1) \geq f(y^2)$, to uznajemy, że y^1 jest lepszy (co najmniej tak dobry) od y^2.

Relacja $R(f)$ preferencji rankingowych, zdefiniowana na podstawie funkcji rankingowej f, ma wobec tego postać następującą:

$$R\big(f\big) = \left\{\big(y,z\big) \in Y \times Y \middle| f\big(y\big) \geq f\big(z\big)\right\}. \tag{2.1}$$

Liczba $f(y)$ jest traktowana jako kompleksowa, skalarna ocena elementu y (a tym samym elementu x). Relacja rankingowa (2.1) w ogólnym przypadku jest jedynie relacją zwrotną i przechodnią, a więc nie gwarantuje rankingu uniwersalnego (liniowego). Do najczęściej stosowanych w praktyce metod rankingowych zaliczana jest metoda sumy ważonej. Polega ona na porządkowaniu zbioru Y, korzystając z funkcji skalarnej f, zdefiniowanej jako suma ważona kryteriów cząstkowych $F_n(x),\ \ n \in \mathrm{N}$

$$f(y) = f\big(F(x)\big) = \sum_{n \in N} \alpha_n F_n(x), \tag{2.2}$$

przy założeniu normalizacyjnym, że

a) $\sum_{n \in \mathrm{N}} \alpha_n = 1,$

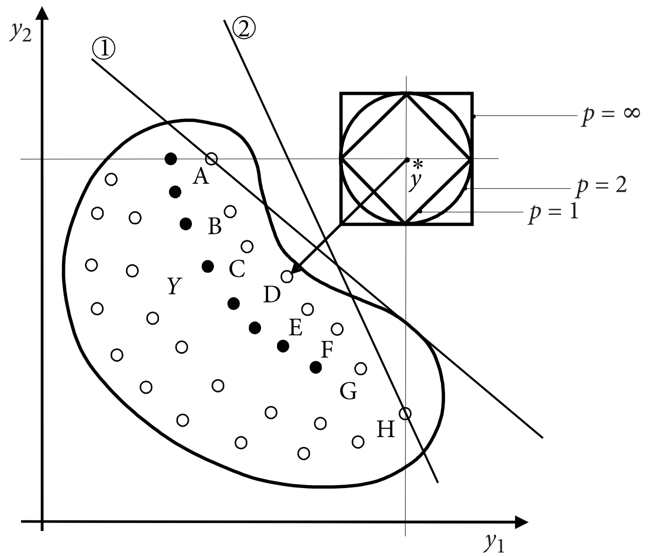
b) $\alpha_n \geq 0, \ \ n \in N.$

Metoda ta posiada bardzo istotną wadę wynikającą głównie z faktu braku możliwości „obiektywnego" wyznaczenia współczynników wagowych α_n, $n \in N$. Postać końcowego rankingu jest na ogół też bardzo „wrażliwa" na zmiany wartości tych współczynników, a ponadto nie zawsze jest to tzw. ranking jednoznaczny. Dodatkową, bardzo dużą wadą jest „bezradność" funkcji ważonej w przypadku braku wypukłości zbioru podlegającego rankingowi, co często ma miejsce w praktyce. Niestety z uwagi na prostotę metody jest ona bardzo chętnie stosowana w praktyce, co nierzadko bywa brzemienne w skutkach ze względu na łatwą możliwość manipulacji współczynnikami wagowymi. Poniżej zostaną przedstawione najważniejsze wady rankingów tworzonych w oparciu o funkcję rankingową z wagami. Analiza będzie przeprowadzona na podstawie pewnej typowej sytuacji wyboru najlepszej oferty z ustalonego zbioru ofert. Na rysunku 1 przedstawiono zbiór *Y* elementów reprezentujących oferty, z których należy wybrać najlepszą, dokonując rankingu korzystając z ważonej funkcji rankingowej typu $f\big(y\big) = \alpha_1 y_1 + \alpha_2 y_2$. Każda oferta jest przedstawiona jako punkt o dwóch współrzędnych, które reprezentują wartości dwóch cech jakościowych oferty. Zależy nam na maksymalizacji wartości obu cech. Na rysunku 1 zaznaczono tak zwany punkt idealny $\overset{*}{y}$ (utopijna oferta idealna) [1, 2], który jest obiektywnym punktem odniesienia (kres dolny zbioru *Y* przy przyjętej w przykładzie relacji preferencji jakościowych Pareto [9, 10]). Relacja jakości Pareto będąca podstawą rankingu paretowskiego jest najczęściej stosowaną w praktyce relacją preferencji [1, 4, 5]. Relację paretowską oznaczać będziemy symbolem „≥". Prowadzi ona do uzyskania rozwiązań niezdominowanych (optymalnych w sensie Pareto [2, 9]) w zadaniu wielokryterialnej maksymalizacji.

$$\geq \ \ = \Big\{(y,z) \in Y \times Y \Big| y_n \geq z_n \ \text{ dla każdego } \ n \in N\Big\}. \tag{2.3}$$

Literami od A do H oznaczone zostały te oferty, które stanowią zbiór Pareto (zbiór ofert, od których nie ma lepszych w zbiorze *Y*).

Na rysunku tym zaznaczono również kule o środku w punkcie idealnym dla trzech różnych wartości parametru *p* w metryce Minkowskiego [2, 9]. Strzałką wskazano element D leżący najbliżej punktu idealnego w sensie odległości Euklidesa (i Czebyszewa również). Proste o numerach 1 i 2, reprezentujące funkcje rankingowe z odpowiednimi wagami, wyznaczają najlepsze oferty reprezentowane przez punkty A oraz H.



Rys. 1. Wady funkcji wagowych przy wyborze optymalnej oferty

A oto najważniejsze wnioski, które wynikają z analizy powyższego przykładu:

1) Zbiór elementów oznaczonych literami od A do H stanowi zbiór Pareto [2, 9, 10] czyli zbiór ofert, od których nie ma lepszych w zbiorze *Y*.
2) „Najlepszym elementem" w sensie potocznych oczekiwań „maksymalizacyjnych" jest element D, bo jest optymalny w sensie Pareto i jednocześnie znajduje się najbliżej oferty idealnej $\overset{*}{y}$ [9, 10].
3) Funkcje rankingowe z wagami „pozwalają wybrać" spośród wielu elementów optymalnych w sensie Pareto jedynie dwa punkty: A lub H. Są to na domiar złego „punkty krańcowe" zbioru Pareto. Punkt H jest ponadto gorszy (w sensie odległości od oferty idealnej) od wszystkich pozostałych punktów Pareto, natomiast punkt A jest gorszy od zdecydowanej większości punktów Pareto (B, C, D, E, F).
4) Żaden z pozostałych punktów Pareto (mimo że są lepsze od A lub H) „nie ma szans" być wybranym przez funkcję wagową (niezależnie od wartości wag).
5) Elementy zaczernione, mimo że nie są elementami Pareto (są zatem od nich gorsze), są i tak lepsze (w sensie odległości od oferty idealnej) od wybranego przez wagową funkcję rankingową najlepszego elementu H.
6) „Optymalne oferty" otrzymane za pomocą wagowej funkcji rankingowej (A oraz H) diametralnie różnią się między sobą współrzędnymi przy stosunkowo niewielkich zmianach współczynników wagowych — punkt H niewspółmiernie faworyzuje pierwsze kryterium, zaś A faworyzuje drugie kryterium.

7) Najlepsza oferta D — optymalna w sensie Pareto i jednocześnie leżąca najbliżej oferty idealnej [2, 9, 10] — jest niemożliwa do „wydobycia" ze zbioru *Y* przy żadnym zestawie wag.

Podsumowując, metoda rankingu oparta na wagowych funkcjach rankingowych jest bardzo daleka od obiektywizmu (głównie z racji subiektywnych wag), a ponadto jest „ułomna" w sensie możliwości pełnej penetracji zbioru możliwych rozwiązań, co może prowadzić do pomijania „rzeczywiście optymalnych" rozwiązań. W praktyce na ogół jednak nikt nie robi powyższej analizy. Tym bardziej że już przy trzech cechach opisujących oferty jest ona niemożliwa do graficznej realizacji. Zatem autor takiego rankingu najczęściej nie ma nawet świadomości, że w zbiorze dopuszczalnych rozwiązań były rozwiązania znacznie lepsze niż te uzyskane w wyniku rankingu. Stosowanie tej metody do ważnych rankingów skutkujących znacznymi konsekwencjami finansowymi lub innymi jest zatem niedopuszczalne i bardzo szkodliwe. W dalszej części pracy zostanie przedstawiona ogólna procedura rankingowa, która eliminuje większość wad ważonych funkcji rankingowych. Własności proponowanej procedury będą ilustrowane obszernym przykładem liczbowym. Efektem końcowym będzie uniwersalny algorytm rankingowy wykorzystujący funkcje Minkowskiego [1, 9].

3. Przestrzeń jakości — ogólne rankingi jakościowe

Procedury rankingowe w zastosowaniach praktycznych bazują najczęściej na modelach jakościowych (rozumianych w niniejszej pracy jako modele opisujące wartości poszczególnych cech jakościowych obiektu) „rankingowanych obiektów". Stąd też pierwszym etapem jest zbudowanie modelu jakościowego obiektów. Sprowadza się to do zdefiniowania odpowiednich wskaźników jakości obiektów (cech jakościowych obiektów) [3]. Ich liczba i konkretna postać zależy oczywiście od potrzeb (celu) rankingu. W ten sposób obiektami porównań rankingowych stają się nie bezpośrednio obiekty, lecz ich modele jakościowe w przestrzeni jakości [1, 2]. Kolejnym etapem jest zazwyczaj ustalenie modelu preferencji jakościowych decydenta, najczęściej w postaci odpowiedniej relacji zdefiniowanej w przestrzeni jakości. Jest to podejście o wiele bardziej ogólne, bezpieczniejsze i „bardziej obiektywne" niż podejście polegające na bezpośrednim wykorzystaniu skalarnej funkcji oceny (rankingu), np. w postaci sumy ważonej [5]. Definiowanie modelu preferencji w postaci relacji bywa najczęściej o wiele prostsze i nie wymaga podawania liczbowych wartości współczynników, progów liczbowych itp.

Ogólne zadanie tworzenia rankingów jakościowych można przedstawić następująco:

$X \subset A$ — zbiór rozważanych obiektów (elementów) w pewnej przestrzeni obiektów A;

B — przestrzeń ocen jakościowych obiektów (najczęściej $B = \mathbf{R}^N$);
$F : A \to \mathbf{R}^N$ — funkcja oceny obiektów;
$\mathrm{N} = \{1,...,n,...,N\}$ — zbiór numerów wskaźników cząstkowych oceny obiektów;
$F(x) = y \in \mathbf{R}^N$ — ocena (model jakościowy) obiektu $x \in A$, mająca postać: $F(x) = (F_1(x),...,F_n(x),...,F_N(x)) \in \mathbf{R}^N$.

Model preferencji jakościowych rankingu możemy przedstawić w postaci relacji preferencji jakościowych R

$$R = \{(y,z) \in \mathbf{R}^N \times \mathbf{R}^N \,|\, \text{„}y \text{ jest lepszy od } z\text{''}\}. \tag{3.1}$$

Funkcję zdaniową $\varphi(y,z) \equiv$ „y jest lepszy od z" w powyższej definicji można definiować na bardzo wiele sposobów [2, 3].

W procedurach rankingowych do wyznaczania kolejnych „kategorii" (klastrów) elementów [1, 4, 5] wykorzystywane są następujące pojęcia elementów ekstremalnych [6]:

$Y^R_{\inf}$ — zbiór elementów najmniejszych (dominujących [1, 2]),
$Y^R_{\min}$ — zbiór elementów minimalnych (niezdominowanych [1, 2, 9, 10]),
które w ogólności (bez żądania własności porządkowych od relacji R) można zdefiniować następująco [2]:

$$Y^R_{\inf} = \{y \in Y \,|\, (y,z) \in R \text{ dla każdego } z \in Y - \{y\}\} \tag{3.2}$$

$$Y^R_{\min} = \{y \in Y \,|\, \text{ nie istnieje } z \in Y - \{y\}, \text{ że } (z,y) \in R\}. \tag{3.3}$$

Procedura rankingowa podziału zbioru na klastry (np. wg formuły (3.2)) polega na wyznaczeniu pierwszego klastra jako zbioru elementów najmniejszych całego zbioru Y, następnie na wyznaczeniu spośród elementów pozostałych zbioru Y zbioru elementów najmniejszych (wg tej samej formuły (3.2)) i tak dalej, aż zbiór „elementów pozostałych" stanie się pusty (szczegóły w pracy [1]).

W dalszej części pracy rozważania związane z procedurami rankingowymi będą ilustrowane obszernym przykładem dotyczącym rankingu zbioru dwudziestu obiektów. Jakość tych obiektów jest oceniana wartościami dwóch cech.

Poniższa tabela przedstawia zbiór obiektów (ponumerowanych indeksem $k = 1,\ldots, 20$) oraz odpowiadające im wartości cech.

Tabela 1

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
y_1^k	2	4	5	6	6	6	5	3	2	1	0	1	2	4	5	4	3	2	3	3
y_2^k	7	7	6	5	3	2	1	1	1	2	4	6	6	5	4	3	2	4	5	4

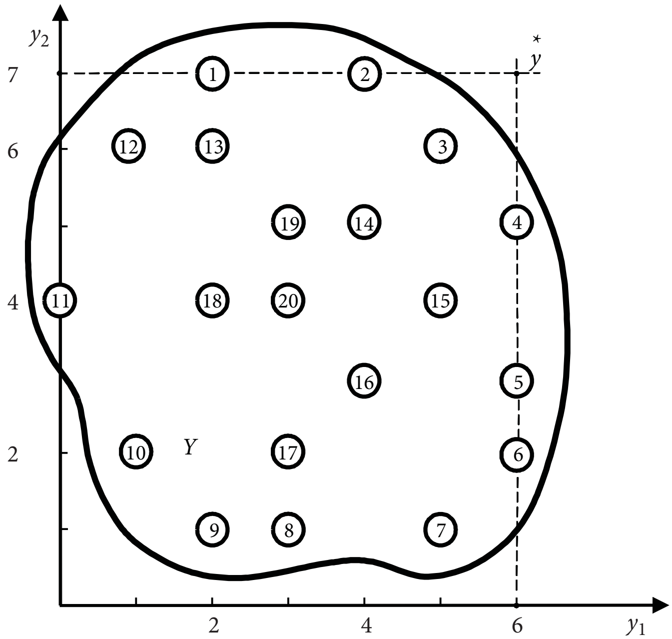
W przykładzie przyjęto założenie, że im wartości rozpatrywanych cech są większe, tym „ogólna jakość obiektu" jest wyższa. Zatem modelem jakościowym obiektu nr k będzie element:

$$y^k = \left(y_1^k, y_2^k\right) \in R^2, \quad k \in K = \{1,\ldots,20\}. \tag{3.4}$$

Zbiorem modeli jakościowych rozpatrywanych obiektów jest zbiór:

$$Y = \left\{y^k \in R^2 \,\middle|\, k \in K\right\}. \tag{3.5}$$

Na rysunku 2 przedstawiony został zbiór Y elementów podlegających procedurze rankingowania. Na rysunku tym dodatkowo został zaznaczony element $\overset{*}{y}$. Nazywany jest on punktem idealnym (utopijnym) [9, 10] zbioru Y. Jest to kres dolny zbioru



Rys. 2. Zbiór obiektów Y w procedurze rankingowej oraz punkt odniesienia $\overset{*}{y}$

Y przy relacji preferencji (dominowania) „$\geq$" zwanej relacją Pareto [2]. Element idealny $\overset{*}{y}$ odgrywa bardzo ważną rolę w wielu procedurach rankingowych jako tak zwany „obiektywny punkt odniesienia". Na rysunku 2 jest to punkt $\overset{*}{y} = (6,7)$ o największych możliwych wartościach obu cech rozpatrywanych obiektów.

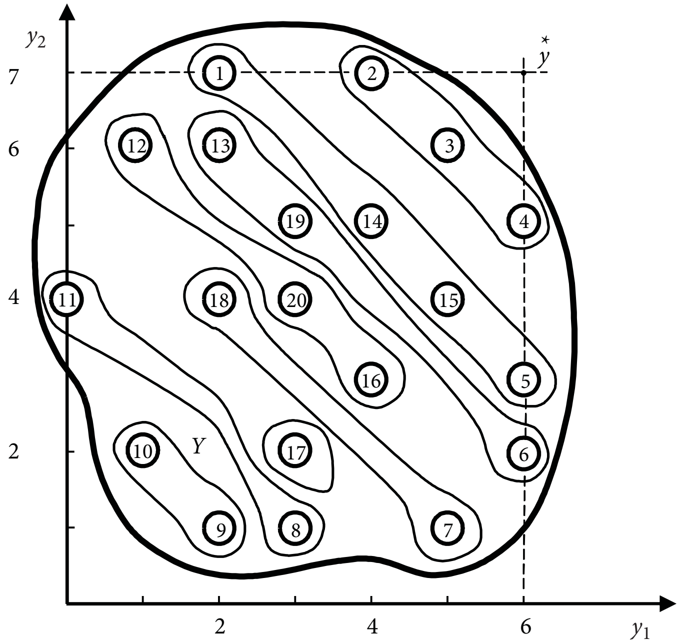
Rankingi jakościowe zbioru Y mogą być tworzone w oparciu o definicje elementów ekstremalnych (najmniejszych lub minimalnych (3.2) i (3.3)). O jednoznaczności (liniowości) rankingów decydują głównie własności modelu preferencji R. Może się jednak zdarzyć, że przeciwobraz jednoznacznego rankingu zbioru Y nie będzie jednoznaczny w zbiorze X. Oznaczać to będzie, że co najmniej dwa różne obiekty mają identyczny obraz w przestrzeni jakości (te same wartości wszystkich cech uwzględnianych w procesie modelowania — co w praktyce oznacza konieczność zwiększenia dokładności modelowania [3]).

4. Funkcje rankingowe Minkowskiego bazujące na pojęciu odległości od punktu idealnego

O wiele więcej zalet w porównaniu z funkcją ważoną mają funkcje rankingowe oparte na określaniu „odległości" poszczególnych obrazów jakościowych obiektów od tzw. obiektu idealnego [2, 9]. W pracy [9] przedstawiona jest bardzo szczegółowa analiza własności takich funkcji, potwierdzająca słuszność ich stosowania. Obiekt (punkt) idealny ma walory „obiektywizmu jakościowego", jako że jest wyłącznie „funkcją" przyjętej relacji preferencji i własności samego zbioru Y. Obiekt ten jest po prostu kresem dolnym tego zbioru [2], a więc naturalnym „punktem odniesienia" nie wymagającym żadnych wag ani „progów satysfakcji" [4, 5]. Na kolejnych rysunkach pokazano możliwość tworzenia rankingu elementów zbioru Y w oparciu o uogólnioną postać normy (metryki) $R_p^{\overset{*}{y}}(y)$ Minkowskiego [1, 2, 9] w postaci przystosowanej do specyfiki określania odległości od punktu idealnego:

$$R_{\mathrm{p}}^{\overset{*}{y}}(y) = \left\| \overset{*}{y} - y \right\|_p = \left(\sum_{n \in \mathrm{N}} \left(\overset{*}{y}_n - y_n \right)^p \right)^{1/p}, \quad y \in Y, \quad p \geq 1. \tag{4.1}$$

Postać tej funkcji ze względu na bardzo istotne „własności decyzyjne" jest powszechnie akceptowana [9]. Dla $p = 1$, funkcja (4.1) staje się tak zwaną odległością miejską typu Manhattan (w praktyce jednak rzadko stosowaną). Podział zbioru na odpowiednie klastry według tej funkcji pokazany jest na rysunku 3. Odpowiada on też wagowej funkcji rankingowej dla $\alpha_1 = \alpha_2$. Dla $p = 2$ funkcja ta jest zwykłą odległością euklidesową, zaś dla $p = \infty$ staje się odległością Czebyszewa [1, 9, 10]. Elementy tak utworzonych klastrów znajdują się w tej samej odległości od punktu



Rys. 3. Podział zbioru Y na klastry „miejskie" typu Manhattan ($p = 1$)

idealnego. W tym sensie są tak samo dobre (równoważne w aspekcie jakości). Punkt idealny (inaczej utopijny [9]), leżący poza zbiorem „obiektów rzeczywistych" Y, jest lepszy w sensie przyjętej relacji preferencji od wszystkich elementów ze zbioru Y. Jego położenie jest ściśle uwarunkowane własnościami („kształtem") rozważanego zbioru Y i przyjętej relacji preferencji [2, 9, 10]. Elementy ze zbioru Y są tym lepsze, im się znajdują bliżej punktu odniesienia. Funkcję rankingową (respektując zależność (2.1)) można zdefiniować następująco:

$$f(y) = \left\| \overset{*}{y} \right\|_p - \left\| \overset{*}{y} - y \right\|_p, \quad p \geq 1. \tag{4.2}$$

Obszerną analizę własności możliwych do zastosowania funkcji odległości Minkowskiego w aspekcie jakości decyzji można znaleźć w [2, 9].

Relację preferencji rankingowych (3.1) względem punktu idealnego $\overset{*}{y}$, korzystając z (4.2), możemy zdefiniować następująco:

$$
\begin{aligned}
R_p^{\overset{*}{y}} &= \left\{ (y,z) \in Y \times Y \,\middle|\, f(y) \geq f(z) \right\} = \\
&= \left\{ (y,z) \in Y \times Y \,\middle\|\, \left\| \overset{*}{y} \right\|_p - \left\| \overset{*}{y} - y \right\|_p \geq \left\| \overset{*}{y} \right\|_p - \left\| \overset{*}{y} - z \right\|_p \right\} = \qquad (4.3) \\
&= \left\{ (y,z) \in Y \times Y \,\middle\|\, \left\| \overset{*}{y} - y \right\|_p \leq \left\| \overset{*}{y} - z \right\|_p \right\}.
\end{aligned}
$$

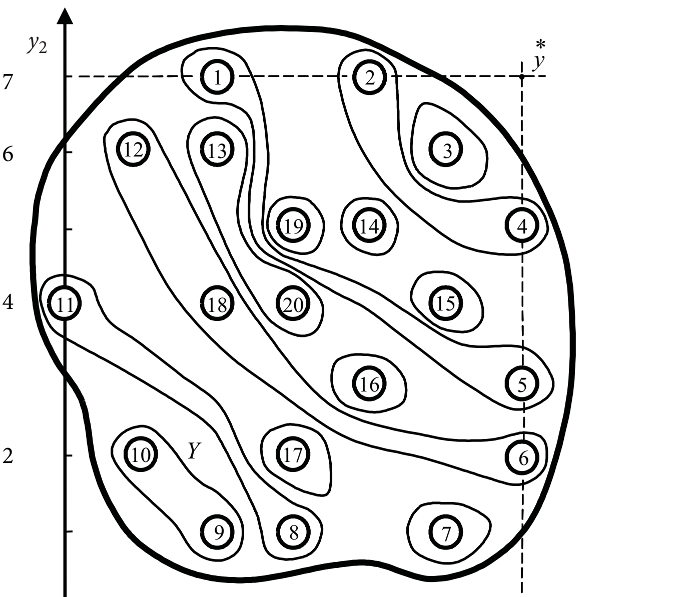
Jest to relacja quasi-porządku, stąd korzystając z definicji elementu najmniejszego (3.2), dla $p = 2$ (rys. 4), otrzymamy następujące klastry (kategorie) wyznaczone rekurencyjnie [1] zgodnie z opisem w paragrafie 3.

$Y_{\inf}^{R_p^{\overset{*}{y}}}(1) = \{3\}$,

$Y_{\inf}^{R_p^{\overset{*}{y}}}(2) = \{2, 4\}$,

$Y_{\inf}^{R_p^{\overset{*}{y}}}(3) = \{14\}$,

$Y_{\inf}^{R_p^{\overset{*}{y}}}(4) = \{15\}$ itd.



Rys. 4. Podział zbioru Y na klastry Euklidesa, $p = 2$

W sytuacji gdy niektóre klastry nie są jednoelementowe, potrzebna jest dodatkowa „informacja preferencyjna" dotycząca poprzedzania w sensie jakości w ramach elementów tego samego klastra. W takim przypadku mogą być stosowane różne sposoby wewnętrznego rankingu (np. dodatkowe funkcje rankingowe). Bardzo często jednak decydent jest pytany o hierarchię ważności (leksykografię) poszczególnych kryteriów cząstkowych (cech jakościowych analizowanych obiektów). Praktyczne eksperymenty decyzyjne na ogół potwierdzają dużą podatność decydentów na takie „uzupełnienie preferencji" rankingowych. Tak ustalone preferencje leksykograficzne w ramach wewnętrznych rankingów oznaczać będziemy $L_{\geq}$ (bez strat ogólności rozważań założymy, że leksykografia ta odpowiada porządkowi naturalnemu).

W rozpatrywanym przykładzie (odległość euklidesowa), po zastosowaniu leksykografii naturalnej (1, 2) otrzymamy w konsekwencji ostateczny ranking liniowy:

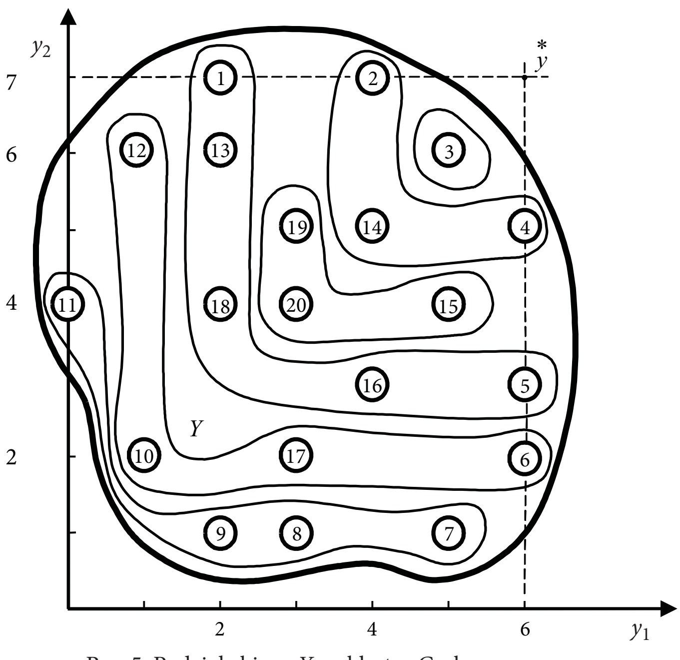
$$(3, 4, 2, 14, 15, 19, 5, 1, 20, 13, 16, 6, 18, 12, 17, 7, 8, 11, 9, 10). \tag{4.4}$$

Dla $p = \infty$ otrzymalibyśmy następujące klastry (patrz rys. 5):

$$\{3\}, \{2, 4, 14\}, \{15, 19, 20\}, \{1, 5, 13, 16, 18\}, \{6, 10, 12, 17\}, \{7, 8, 9, 11\},$$

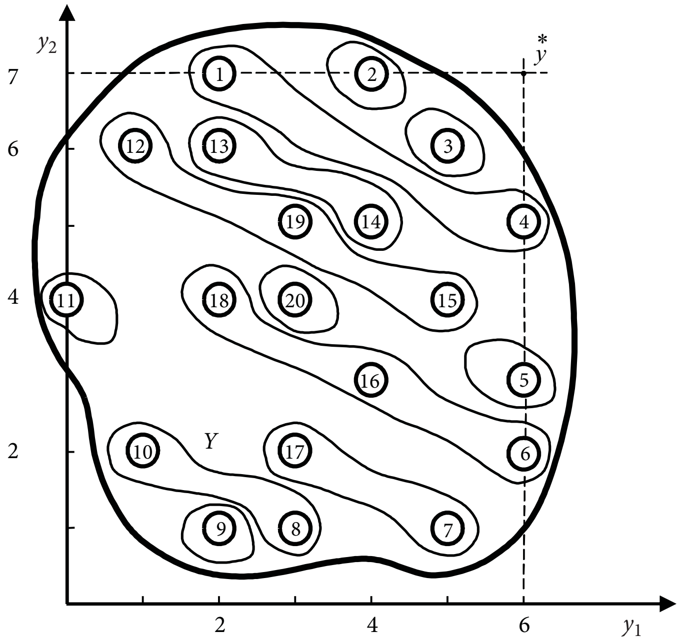
a w konsekwencji ranking ostateczny:

$$(3, 4, 2, 14, 15, 19, 20, 5, 16, 1, 13, 18, 6, 17, 12, 10, 7, 8, 9, 11). \tag{4.5}$$



Rys. 5. Podział zbioru Y na klastry Czebyszewa, $p = \infty$

Jak widać, zastosowanie funkcji rankingowej wykorzystującej odległość Minkowskiego również nie musi prowadzić do uzyskania jednoznacznego rankingu. Niektóre z wyznaczonych klastrów nie są jednoelementowe. Stosując do tych klastrów relację leksykograficzną $L \geq [1, 2]$, otrzymamy ostateczny ranking liniowy. Na rysunku 6 przedstawiony został podział zbioru na klastry rankingowe z wykorzystaniem relacji preferencji (2.1) zdefiniowanej na bazie funkcji ważonej (2.2) ze współczynnikami wagowymi odpowiednio równymi 1/3 i 2/3.



Rys. 6. Podział zbioru Y na klastry według sumy ważonej (1/3, 2/3)

Dla przypadku na rysunku 6 otrzymamy ranking ostateczny w postaci:

$$(2, 3, 4, 1, 14, 13, 15, 19, 12, 5, 20, 6, 16, 18, 11, 7, 17, 8, 10, 9). \tag{4.6}$$

Dla porównania, gdy współczynniki wagowe funkcji ważonej są równe (1/2,1/2), otrzymamy następujący ranking:

$$(4, 3, 2, 5, 15, 14, 1, 6, 19, 13, 16, 20, 12, 7, 18, 17, 8, 11, 9, 10). \tag{4.7}$$

Jak łatwo zauważyć, oba rankingi zbudowane z wykorzystaniem funkcji rankingowej w postaci sumy ważonej przy nieco innych współczynnikach wagowych różnią się bardzo istotnie. Na 20 pozycji rankingowych tylko na jednej pozycji wystąpiła zgodność. Również „wygrywająca oferta” jest inna. Porównując z kolei rankingi otrzymane z wykorzystaniem funkcji odległościowych Euklidesa i Czebyszewa względem punktu idealnego (jak niżej), można zauważyć bardzo dużą zbieżność wyników:

— dla funkcji Euklidesa: (3, 4, 2, 14, 15, 19, 5, 1, 20, 13, 16, 6, 18, 12, 17, 7, 8, 11, 9, 10).

— dla funkcji Czebyszewa: (3, 4, 2, 14, 15, 19, 20, 5, 16, 1, 13, 18, 6, 17, 12, 10, 7, 8, 9, 11).

Zbieżność wystąpiła na wielu pozycjach, w tym, co najważniejsze, aż na sześciu pierwszych pozycjach wygrywających. Wracając do przykładu opisanego w paragrafie 2 warto zauważyć, że w przeciwieństwie do metody funkcji wagowych, zastosowanie dowolnej funkcji odległości od punktu idealnego (dla $p > 1$) prowadzi do uzyskania oferty najlepszej oznaczonej literą D, czego nie umożliwiają funkcje wagowe. Wynika to z faktu, że funkcje Minkowskiego dla $p > 1$ mają znacznie większą zdolność „penetracji" zbioru Y niż funkcje wagowe (odpowiedni „kształt kuli" o środku w punkcie idealnym $\overset{*}{y}$ — patrz rys. 1).

5. Uniwersalny algorytm rankingowy

Podsumowując, każdą procedurę rankingową, z dowolną funkcją rankingową (metafunkcją [4, 5, 7, 8]) zdefiniowaną w N-wymiarowej przestrzeni jakości, można sprowadzić do uniwersalnego leksykograficznego algorytmu rankingowego w przestrzeni $(N + 1)$ wymiarowej w następujący sposób:

Niech $F : X \to \mathbf{R}^N$ wyjściowa (początkowa) N-wymiarowa funkcja oceny obiektów, taka że

$$F(x) = \big(F_1(x), \ldots, F_n(x), \ldots, F_N(x)\big) \in \mathbf{R}^N, \quad x \in X,$$

gdzie: $F_n(x)$ — cząstkowa funkcja oceny obiektu $x \in X$ w sensie kryterium $n \in N$;
$F_0(x) \to \mathbf{R}^1$ — dowolna, kompleksowa funkcja rankingowa („metafunkcja") na przykład uogólniona funkcja Minkowskiego.

Ogólne leksykograficzne zadanie rankingowe będzie miało postać [1, 2]:

$$\big(X, F, L_{\geq}\big), \tag{5.1}$$

gdzie: $F : X \to \mathbf{R}^{N+1}$, taka że
$F(x) = \big(F_0(x), F_1(x), \ldots, F_n(x), \ldots, F_N(x)\big) \in \mathbf{R}^{N+1}, \quad x \in X;$
$L_{\geq}$ — relacja leksykograficzna (hierarchia ważności kryteriów $(0,1,\ldots,n,\ldots,N)$).

Zadanie to rozwiązujemy rekurencyjnie klasycznym algorytmem leksykograficznym [2], otrzymując „jednoelementowe klastry", aż do otrzymania liniowego rankingu wynikowego całego zbioru Y [1].

Pierwsze rozwiązanie rankingowe otrzymamy, wyznaczając w sposób rekurencyjny zbiór $X_1 = X_N$.

$$X_n = \left\{ \overset{*}{x} \in X_{n-1} \middle| F_n\left(\overset{*}{x}\right) = \max_{x \in X_{n-1}} F_n(x) \right\},\ n \in \{0,1,\ldots,N\}, \tag{5.2}$$

przy czym $X_{-1} = X$.

Wyznaczając zbiory (5.2), otrzymamy odpowiednio:

$$X_0 = \left\{ \overset{*}{x} \in X \middle| F_0\left(\overset{*}{x}\right) = \max_{x \in X} F_0(x) \right\},$$

$$X_1 = \left\{ \overset{*}{x} \in X_0 \middle| F_1\left(\overset{*}{x}\right) = \max_{x \in X_o} F_1(x) \right\},\ \text{ itd. aż do } X_N.$$

Kolejne rozwiązania rankingowe otrzymujemy analogicznie, korzystając z zależności (5.2), redukując zbiór X o wyznaczone wcześniej rozwiązania rankingowe. W pracy [1] podano wersję uproszczoną tego algorytmu, polegającą na wcześniejszym wyznaczaniu kolejnych klastrów równoważności (w sensie odległości od punktu idealnego), a następnie na ich leksykograficznym porządkowaniu.

6. Podsumowanie

W pracy zostało przedstawione ogólne podejście do problematyki rankingu elementów zbioru Y bazujące na wykorzystaniu funkcji rankingowej. W wyniku obszernej analizy wskazano na wiele wad i niebezpieczeństw w stosowaniu tak zwanej ważonej funkcji rankingowej. Jako alternatywę dla funkcji ważonych przedstawiono możliwość stosowania metryk Minkowskiego, które nie mają większości wad funkcji ważonych. Funkcje rankingowe mogą być stosowane do definiowania relacji preferencji rankingowych $R \subset Y \times Y$. „Skuteczność procedury rankingowej" zależy oczywiście od własności relacji preferencji. W pracy przedstawiono między innymi relacje definiowane z wykorzystaniem funkcji rankingowych typu odległość Minkowskiego [2, 9].

Ogólna procedura rankingowa oparta na relacji preferencji pozwala dokonać podziału zbioru na „klastry jakościowe" (kategorie), tworzące pewien ciąg rankingowy klastrów. Klastry te można opatrzyć odpowiadającą im „liczbą rankingową" wynikającą z odległości klastra od punktu idealnego [1]. Liczność klastrów zależy oczywiście od własności relacji preferencji i „natury" zbioru Y. W pracy przedstawiono uniwersalny algorytm rankingowy pozwalający otrzymać ranking liniowy niezależnie od własności funkcji rankingowej. W algorytmie tym wykorzystywana jest pomocniczo relacja leksykograficzna w zakresie klastrów o liczności większej niż 1. Praktyka decyzyjna pokazuje, że decydenci bardzo łatwo akceptują wykorzystywanie pomocniczej relacji leksykograficznej w sytuacji pojawiania się klastrów o liczności większej niż 1. Podejście takie eliminuje konieczność określania jakichkolwiek wag czy też progów liczbowych. Większość stosowanych w praktyce funkcji rankingowych prowadzi do relacji quasi-porządku lub porządku. Relacje te pozwalają tworzyć bardzo interesujące rankingi klastrów jakościowych. Znalazły one między innymi zastosowanie w procedurach przetargowych i konkursowych oraz analizach scjentometrycznych dotyczących wpływu poszczególnych źródeł bibliograficznych, naukowców, instytucji naukowych czy też państw na rozwój nauki w różnych dyscyplinach. Mogą być też z powodzeniem wykorzystywane w wielu rankingach dotyczących jakości „złożonych obiektów", w tym jakości wyrobów na rynku. Zaproponowany w pracy ogólny leksykograficzny algorytm rankingowy, wykorzystujący dowolną klasę funkcji rankingowych, daje możliwość łatwego tworzenia liniowych rankingów dowolnych zbiorów w oparciu o powszechnie używane funkcje rankingowe.

Funkcje rankingowe bazujące na uogólnionej odległości Minkowskiego od tak zwanego punktu idealnego (kresu dolnego zbioru Y) mają dodatkowo bardzo wiele innych interesujących „własności decyzyjnych" [1, 9, 10]. Własności te powodują, że powinny być one najczęściej stosowanymi w praktyce funkcjami rankingowymi, również w algorytmach podziału zbiorów na klastry.

Artykuł wpłynął do redakcji 18.11.2013 r. Zweryfikowaną wersję po recenzji otrzymano 7.04.2014 r.

LITERATURA

[1] Ameljańczyk A., *Metoda podziału zbioru obiektów na wielokryterialne klastry jakościowe*, Biul. ISI, 11, 2013.

[2] Ameljańczyk A., *Optymalizacja wielokryterialna w problemach sterowania i zarządzania*, Ossolineum, 1984.

[3] Ameljańczyk A., *Matematyczne aspekty modelowania pajęczynowego obiektów*, Biul. ISI, 4, 2009.

[4] Bouyssou D., Marchant T., *An axiomatic approach to noncompensatory sorting methods in MCDM*, I: *The case of two categories*, EJOR, 2007.

[5] Brans J., Vincke P., Ph., *A preference ranking organization method: The PROMETHEE method*, Management Sci., 1985.

[6] Rasiowa H., *Wstęp do matematyki współczesnej*, PWN, Warszawa, 2005.

[7] Saaty T.L., *Rank from comparisons and from ratings in the analytic hierarchy/networkprocesses*, EJOR, 2006.

[8] Seo F., Sakawa M., *Multiple Criteria Decision Analysis in Regional Planning*, D. Reidel-Kluwer, Dordrecht-Boston-Lancaster-Tokyo, 1988.

[9] Yu P.L., Leitmann G., *Compromise solutions, domination structures and Salukwadze's solution*, JOTA, 13, 1974.

[10] Yu P.L., Leitmann G., *Nondominated decision and cone convexity in dynamic multicriteria decision problems*, JOTA, 14, 1974.

A. AMELJAŃCZYK

Minkowski metrics in creating universal ranking algorithms

Abstract. The paper presents a general procedure for creating the rankings of a set of objects, while the relation of preference based on any ranking function. The analysis was possible to use the ranking functions began by showing the fundamental drawbacks of commonly used functions in the form of a weighted sum. As a special case of the ranking procedure in the space of a relation, the procedure based on the notion of an ideal element and generalized Minkowski distance from the element was proposed. This procedure, presented as universal ranking algorithm, eliminates most of the disadvantages of ranking functions in the form of a weighted sum.

Keywords: ranking functions, preference relation, ranking clusters, categories, ideal point, universal ranking algorithm