Listy Biometryczne
Nr 49-50 (1975)

Daniela C z e r w i ń s k a
i Hubert G e m b a r z e w s k i (Wrocław)

# O WSPÓŁCZYNNIKU RENKONENA PODOBIEŃSTWA ZBIORÓW

Istnieje wiele metod obliczania podobieństwa zbiorów, a zastosowanie znajdują one tak w ekologii do badań zoologicznych lub botanicznych, jak i w archeologii, socjologii, geografii, psychologii, ekonomice, lingwistyce. Lerman [4] podaje 12 znanych mu wzorów na obliczanie podobieństwa zbiorów. W Polsce poprawny matematycznie wzór podali E. Marczewski i H. Steinhaus [5], [6] (zobacz także [2]).

Przy badaniu własności podobieństwa zbiorów wygodnie jest posługiwać się odległością zbiorów zdefiniowaną następująco:

$$(1) \qquad \rho_s(A, B) = 1-s,$$

gdzie s oznacza podobieństwo zbiorów A i B. Wzory Marczewskiego i Steinhausa mają postać:

$$\rho_s(A, B) = \frac{a+b-2w}{a+b-w} \text{ dla zbiorów,}$$

$$\rho_s(f, g) = \frac{\int|f-g|\,dm}{\int\max(f,g)\,dm} \text{ dla funkcji,}$$

gdzie a, b oznaczają odpowiednio liczbę elementów (np. gatunków) w zbiorach A, B oraz w - liczbę elementów wspólnych obu zbiorów, natomiast funkcje f i g określają skład ilościowy poszczególnych gatunków na danych dwu obszarach.

W praktyce porównywania biocenoz (szczególnie zoocenoz) rozpowszechniła się w ostatnich latach tzw. metoda Renkonena. Sposób obliczania podobieństwa tą metodą opisany jest w podręczniku [1]. Podany tam sposób obliczania spotyka się w wielu pracach zoologicznych. Identyczny z metodą Renkonena sposób obliczeń zastosował w zagadnieniach botaniczno-rolniczych S. Grzyb [3].

Metoda Renkonena uwzględnia frakcję poszczególnych gatunków. Porównując dwa zbiory dla każdego gatunku wybiera się mniejszą z obu frakcji. Suma tych frakcji po wszystkich gatunkach, jakie występują w obydwu rozważanych zbiorach jest wskaźnikiem Renkonena podobieństwa tych zbiorów.

W dostępnej literaturze autorzy niniejszej pracy nie znaleźli wzoru na podobieństwo zbiorów Renkonena.

W. Romaniszyn w pracy [7] twierdzi, że metoda Renkonena jest szczególnym przypadkiem metody Kulczyńskiego. Wiadomo, że wzór na odległość zbiorów oparty na metodzie Kulczyńskiego nie spełnia warunku trójkąta. W. Romaniszyn wnioskuje stąd, że również metoda Renkonena z tego punktu widzenia nie jest poprawna. W niniejszej pracy sformułujemy wzór na podobieństwo Renkonena oparty na znanej metodzie Renkonena oraz wykażemy, że odległość związana z tym podobieństwem spełnia warunek trójkąta. Zaproponujemy również inny wskaźnik podobieństwa zbiorów (i jego analogon dla funkcji), udowodnimy, że odpowiadająca temu wskaźnikowi odległość zbiorów spełnia warunek trójkąta oraz porównamy go ze wskaźnikiem podobieństwa Marczewskiego-Steinkausa i ze wskaźnikiem Renkonena.

Niech X będzie przestrzenią wszystkich gatunków, jakie mogą występować na obszarze A lub B, zaś m miarą określoną na podzbiorach X w następujący sposób: dla danego zbioru $E \subset X$, $m(E) = \overline{\overline{E}}$ (moc zbioru E, liczba elementów zbioru E). Niech nieujemne funkcje f i g charakteryzują w pewien sposób odpowiednio obszar A i B. Jeżeli f i g oznaczają poszczególne gatunki odpowiednio na obszarze A i B, to wówczas podobieństwo Renkonena tych obszarów wyraża się wzorem

$$r = \int \min\left(\frac{f}{\int f dm}, \frac{g}{\int g dm}\right) dm.$$

Wobec (1) mamy

$$\varrho_r(f, g) = 1 - \int \min\left(\frac{f}{\int f dm}, \frac{g}{\int g dm}\right) dm. \tag{2}$$

Nie trudno sprawdzić, że spełnione są dwa warunki:

1. $\varrho_r(f, g) \geqslant 0;\quad \varrho_r(f, g) = 0 \iff \dfrac{f(x)}{\int f dm} = \dfrac{g(x)}{\int f dm}$

2. $\varrho_r(f, g) = \varrho_r(g, f).$

Sprawdzimy, że spełniony jest warunek trójkąta:

3. $\varrho_r(f, g) + \varrho_r(g, h) \geqslant \varrho_r(f, h).$

Dla odległości (2) sprowadza się on do postaci

$$1 + \int \min\left(\frac{f}{\int f dm}, \frac{h}{\int h dm}\right) dm \geqslant$$

$$\int \min\left(\frac{f}{\int f dm}, \frac{g}{\int g dm}\right) dm + \int \min\left(\frac{g}{\int g dm}, \frac{h}{\int h dm}\right) dm.$$

Ponieważ $1 = \int \frac{g}{\int g dm} dm$, więc dla sprawdzenia warunku trójkąta z odległością (2) wystarczy pokazać, że

$$\frac{g(x)}{\int g dm} + \min\left(\frac{f(x)}{\int f dm}, \frac{h(x)}{\int h dm}\right) \geqslant$$

$$\min\left(\frac{f(x)}{\int f dm}, \frac{g(x)}{\int g dm}\right) + \min\left(\frac{g(x)}{\int g dm}, \frac{h(x)}{\int h dm}\right)$$

Ostatnią nierówność sprawdza się przez rozważenie sześciu przypadków możliwych wartości $\min\left(\frac{f}{\int f dm}, \frac{g}{\int g dm}\right)$, $\min\left(\frac{g}{\int g dm}, \frac{h}{\int h dm}\right)$, $\min\left(\frac{f}{\int f dm}, \frac{h}{\int h dm}\right)$.

Metodę Renkonena ilustruje przykład podany w tab. 1.

Niezależnie od omówionego wyżej wskaźnika podobieństwa funkcji zaproponujemy inny wskaźnik podobieństwa zbiorów i analogiczny wskaźnik podobieństwa funkcji. Mianowicie podobieństwo p zbioru A o liczbie elementów a i zbioru B o liczbie elementów b jest równe

(3) $$p = \frac{w}{\max(a, b)},$$

gdzie w oznacza liczbę elementów wspólnych zbiorów A i B. Odpowiadająca temu podobieństwu odległość zbiorów A, B określona będzie wzorem

(4) $$\varrho_p(A, B) = 1 - p.$$

Odległość ta spełnia warunek trójkąta.

Oznaczmy przez m(E) liczbę elementów zbioru E. Wówczas (3) przyjmuje postać

$$p = \frac{m(A \cap B)}{\max(m(A), m(B))},$$

gdzie $\cap$ jest symbolem przekroju (iloczynu) zbiorów.

Tabela 1

**Przykład na obliczanie podobieństwa Renkonena składu gatunkowego siana z dwu różnych stanowisk łąkowych w Sudetach: A - w Górach Stołowych, B - w Paśmie Krowiarek, po "orce chemicznej" i podsiewie taką samą mieszanką nasion**

| Nazwa gatunku | A f kg/pol. dośw. | $\frac{f}{\int f dm}$ | B g kg/pol. dośw. | $\frac{g}{\int g dm}$ | $\min\left(\frac{f}{\int f dm}, \frac{g}{\int g dm}\right)$ |
|---|---|---|---|---|---|
| Babka lancetowata | 2 | 2/160 | 1 | 1/144 | 1/144 |
| Brodawnik jesienny | 2 | 2/160 | 1 | 1/144 | 1/144 |
| Grzebienica pospolita | 2 | 2/160 | 0 | 0 | 0 |
| Gwiazdnica trawiasta | 0 | 0 | 9 | 9/144 | 0 |
| Kupkówka pospolita | 31 | 31/160 | 41 | 41/144 | 31/160 |
| Kostrzewa czerwona | 28 | 28/160 | 5 | 5/144 | 5/144 |
| Konietlica łąkowa | 2 | 2/160 | 0 | 0 | 0 |
| Koniczyna biała | 0 | 0 | 3 | 3/144 | 0 |
| Koniczyna łąkowa | 2 | 2/160 | 0 | 0 | 0 |
| Kłosówka miękka | 0 | 0 | 18 | 18/144 | 0 |
| Kłosówka wełnista | 0 | 0 | 3 | 3/144 | 0 |
| Komonica rożkowa | 0 | 0 | 1 | 1/144 | 0 |
| Mietlica pospolita | 20 | 20/160 | 3 | 3/144 | 3/144 |
| Mniszek lekarski | 2 | 2/160 | 0 | 0 | 0 |
| Perz rozłogowy | 2 | 2/160 | 5 | 5/144 | 2/160 |
| Podagrycznik łąkowy | 5 | 5/160 | 6 | 6/144 | 5/160 |
| Rumianek pospolity | 0 | 0 | 6 | 6/144 | 0 |
| Tomka wonna | 2 | 2/160 | 1 | 1/144 | 1/144 |
| Tymotka łąkowa | 0 | 0 | 5 | 5/144 | 0 |
| Wiechlina łąkowa | 0 | 0 | 3 | 3/144 | 0 |
| Zerwa kłosowa | 2 | 2/160 | 0 | 0 | 0 |
| Żywica trwała | 58 | 58/160 | 33 | 33/144 | 33/144 |
| Ogółem | 160 | | 144 | | 44/144 + + 38/160 = = 0,543 = 54,3 % |

r = 54,3%

**(Uwaga: r dla odpowiednich poletek kontrolnych A i B wynosiło 12%)**

Na podstawie (4) odległość zbiorów A, B jest równa

$$\varrho_p(A, B) = \frac{\max(m(A), m(B)) - m(A \cap B)}{\max(m(A), m(B))},$$

lub inaczej

$$\varrho_p(A,B) = \frac{\max(m(A \setminus B), m(B \setminus A))}{\max(m(A), m(B))},$$

gdzie $\setminus$ jest symbolem różnicy zbiorów.

Warunek trójkąta z odległością $\varrho_p$ jest postaci

$$\frac{\max(m(A \setminus B), m(B \setminus A))}{\max(m(A), m(B))} + \frac{\max(m(B \setminus C), m(C \setminus B))}{\max(m(B), m(C))} \geqslant \frac{\max(m(A \setminus C), m(C \setminus A))}{\max(m(A), m(C))}.$$

Dowód ostatniej nierówności nie jest trudny i sprowadza się do sprawdzenia sześciu nierówności przy różnych możliwych układach wartości $\max(m(A), m(B))$, $\max(m(B), m(C))$, $\max(m(A), m(C))$. Wzór (3) zawiera ideę ze wzoru na podobieństwo zbiorów Kulczyńskiego. Wskaźnik podobieństwa Kulczyńskiego jest średnią ze stosunku liczby gatunków wspólnych do ogółu gatunków każdego ze zbiorów (por. [4]), natomiast podobieństwo ze wzoru (3) jest mniejszym z tych dwu ułamków. Odpowiedni wzór na podobieństwo funkcji jest postaci

$$p(f, g) = \frac{\int \min(f,g)\,dm}{\max(\int f\,dm, \int g\,dm)}. \tag{5}$$

W przykładzie w tab. 1 podobieństwo to można łatwo obliczyć:

$$p = \frac{1+1+31+5+3+2+5+1+33}{160} = 0{,}51 = 51\%.$$

Interesujące jest porównanie wskaźników podobieństwa zbiorów Renkonena, Marczewskiego-Steinhausa oraz wskaźnika (5). Okazuje się, że jest spełniona następująca nierówność:

$$s \leqslant p \leqslant r,$$

gdzie

s - wskaźnik podobieństwa zbiorów Marczewskiego-Steinhausa,

r - wskaźnik podobieństwa zbiorów Renkonena,

p - wskaźnik podobieństwa zbiorów wyrażony wzorem (5),

gdyż

$$\frac{\min(f,g)}{\int \max(f,g)\,dm} \leqslant \frac{\min(f,g)}{\max(\int f\,dm, \int g\,dm)} \leqslant \min\left(\frac{f}{\int f\,dm}, \frac{g}{\int g\,dm}\right).$$

Podobieństwo zbiorów Renkonena porównuje struktury dwu zbiorów traktowanych jako odrębne całości, a podobieństwo Marczewskiego-Steinhausa zestawia udział gatunków wspólnych na tle wszystkich gatunków występujących w którymś z obu zbiorów. Podobieństwo zbiorów p ma własności pośrednie między podobieństwem Marczewskiego-Steinhausa, a podobieństwem Renkonena: zestawia ono udział gatunków wspólnych na tle zbioru liczniejszego (ogólniej: mającego większe wartości cech gatunków występujących w nim).

## Literatura cytowana


[1] Balogh, J., Lebensgemeinschaften der Landtiere. Ihre Erforschung unter besonderer Berücksichtigung der Zoozönologischen Arbeitsmethoden, Budapest-Berlin 1958.

[2] Czerwińska, D., On the similarity of sets, Zastosowania Matematyki, XIII, 2 (1972), str. 153-158.

[3] Grzyb, S., Łąki w dorzeczu rzeki Liwiec. Zagadnienia geobotaniczne i fizjograficzno-typologiczne, Roczn. Nauk Roln., seria D, 109 (1964).

[4] Lerman, I.C., Les bases de la classification automatique, Paris 1971.

[5] Marczewski, E. i Steinhaus, H., On a certain distance of sets and the corresponding distance of functions, Colloquium Mathematicum 6 (1958), str. 319-327.

[6] Marczewski, E. i Steinhaus, H., O odległości systematycznej biotopów, Zastosowania Matematyki IV, 3 (1959), str. 195-203.

[7] Romaniszyn, W., Uwagi krytyczne o definicji Sorensena i metodzie Renkonena obliczania współczynników podobieństwa zbiorów, Wiadomości Ekologiczne, XVIII, 4 (1972), str 375-380.


## Streszczenie


W pracy podano wzór na podobieństwo zbiorów obliczane dotychczas tzw. metodą Renkonena oraz dowód spełnienia warunku trójkąta odpowiedniej odległości zbiorów.

Podano również inny wzór (4) dla odległości zbiorów (oraz dla funkcji) i udowodniono, że spełnia on warunek trójkąta.


WDN Zam. 2002/76 - 365