

Metody porządkowania liniowego

Marcin Szymkowiak

Zaawansowane techniki analityczne w biznesie
Studia podyplomowe pod patronatem SAS Institute

Metody porządkowania liniowego

Metody porządkowania liniowego

- Metody porządkowania liniowego – ujęcie teoretyczne
- Bezwzorcowe metody porządkowania liniowego
- Wzorcowe metody porządkowania liniowego

Metody porządkowania liniowego – ujęcie teoretyczne

Rola metod porządkowania liniowego obiektów

- Zadaniem metod porządkowania liniowego zbioru obiektów jest uszeregowanie, czyli ustalenie kolejności obiektów lub zbiorów według określonego kryterium.
- Metody porządkowania liniowego stosowane są wówczas gdy można przyjąć pewne nadrzędne kryterium, ze względu na które będzie można uszeregować obiekty od „najlepszego” do „najgorszego”.
- Narzędziem metod porządkowania liniowego jest syntetyczny miernik rozwoju (SMR), będący pewną funkcją agregującą informacje cząstkowe zawarte w poszczególnych zmiennych i wyznaczoną dla każdego obiektu ze zbioru obiektów A .

Metody porządkowania liniowego – ujęcie teoretyczne

Metody porządkowania liniowego – założenia

- dany jest co najmniej dwuelementowy i skończony zbiór obiektów $A = \{A_i\}_{i=1}^n = \{A_1, \dots, A_n\}$;
- istnieje pewne nadrzędne syntetyczne kryterium porządkowania elementów zbioru A , które nie podlega pomiarowi bezpośredniemu (na przykład poziom rozwoju badanego produktu na tle produktów konkurencyjnych, ocena wstępnie wyselekcjonowanych koncepcji produktu);
- dany jest skończony zbiór zmiennych, merytorycznie związanych z syntetycznym kryterium porządkowania; zmienne mają charakter preferencyjny, tzn. wyróżnia się wśród nich stymulanty, destymulanty i nominanty;
- zmienne służące do opisu obiektów są mierzone przynajmniej na skali porządkowej. Jeżeli zmienne opisujące obiekty mierzone są na skali przedziałowej lub ilorazowej, należy sprowadzić je do porównywalności poprzez normalizację;
- relacją porządkującą elementy zbioru A jest relacja większości lub mniejszości, dotycząca ilczbowych wartości syntetycznego miernika rozwoju.

Metody porządkowania liniowego – ujęcie teoretyczne

Metody porządkowania liniowego

- **metody diagramowe** – w metodach tego typu stosuje się graficzną prezentację macierzy odległości zwaną diagramem;
- **metody bezwzorcowe** – w metodach tych zmienna syntetyczna jest funkcją znormalizowanych wartości zmiennych wejściowych;
- **metody wzorcowe** – w metodach tych wykorzystywane jest pojęcie obiektu wzorcowego tzw. wzorca czyli obiektu modelowego o pożądanych wartościach zmiennych diagnostycznych. Syntetyczny miernik rozwoju konstruowany jest w oparciu o odległości pomiędzy obserwowanym obiektem a obiektem wzorcowym;
- **metody iteracyjne** – w metodach tych przyjmuje się funkcję kryterium dobroci grupowania i w kolejnych iteracjach szuka się takiego uporządkowania liniowego obiektów, które optymalizują wartość tej funkcji aż do osiągnięcia przez nią wartości optymalnej (maksymalnej bądź minimalnej).

Metody porządkowania liniowego – ujęcie teoretyczne

Etapy porządkowania liniowego

- Sformułowanie celu analizy oraz wstępnych hipotez badawczych.
- Dobór cech.
- Zebranie materiału statystycznego.
- Wybór metody porządkowania liniowego.
- Utworzenie rankingu obiektów za pomocą wybranej metody.
- Weryfikacja wyników porządkowania.
- Interpretacja wyników porządkowania.

Metody porządkowania liniowego – ujęcie teoretyczne

Rodzaje cech diagnostycznych

- Cecha diagnostyczna X_k jest stymulantą (S), gdy dla każdych dwóch jej wartości x_{ik}^S, x_{jk}^S odnoszących się do obiektów A_i, A_j , spełniony jest warunek $x_{ik}^S > x_{jk}^S \Rightarrow A_i \succ A_j$ (\succ oznacza dominację obiektu A_i nad obiektem A_j) Innymi słowy cecha X_k jest stymulantą, gdy jej wyższe wartości są pożądane.
- Cecha diagnostyczna X_k jest destymulantą (D), gdy dla każdych dwóch jej wartości x_{ik}^D, x_{jk}^D odnoszących się do obiektów A_i, A_j , spełniony jest warunek $x_{ik}^D > x_{jk}^D \Rightarrow A_i \prec A_j$ (\prec oznacza dominację obiektu A_j nad obiektem A_i) Innymi słowy cecha X_k jest destymulantą, gdy jej niższe wartości są pożądane.
- Cecha diagnostyczna X_k jest nominantą (N), gdy dla każdych dwóch jej wartości x_{ik}^N, x_{jk}^N odnoszących się do obiektów A_i, A_j :
 - jeżeli $x_{ik}^N, x_{jk}^N \leq nom_k$ to $x_{ik}^N > x_{jk}^N \Rightarrow A_i \succ A_j$
 - jeżeli $x_{ik}^N, x_{jk}^N > nom_k$ to $x_{ik}^N > x_{jk}^N \Rightarrow A_i \prec A_j$

Najbardziej pożądaną z punktu widzenia rozpatrywanego obiektu lub zjawiska jest wartość nominalna, natomiast wartości mniejsze bądź też większe od wartości nominalnej są mniej korzystne.

Metody porządkowania liniowego – ujęcie teoretyczne

Sposoby normalizacji cech diagnostycznych

Konkretna technika normalizacji zmiennych zależy od ich charakteru, tj. czy są one stymulantami, destymulantami czy nominantami. Rzeczą niezmiennie istotną, jest również to, aby stosowane procedury normalizacyjne spełniały następujące wymagania:

- pozabawienie mian (jednostek), w których wyrażone są cechy diagnostyczne,
- sprowadzenie rzędu wielkości cech diagnostycznych do porównywalności, co oznacza wyrównanie zakresów zmienności cech, a w konsekwencji możliwość ich dodawania,
- równość długości przedziałów zmienności wartości wszystkich cech unormowanych,
- możliwość normowania cech diagnostycznych przyjmujących wartości zarówno dodatnie jak i ujemne,
- możliwość normowania cech przyjmujących wartość równą zero,
- nieujemność wartości cech unormowanych,
- istnienie prostych formuł – w ramach danej procedury normalizacyjnej – ujednolicających charakter zmiennych.

Metody porządkowania liniowego – ujęcie teoretyczne

Metoda unitaryzacji zerowanej

- Stymulanty

$$z_{ik} = \frac{x_{ik} - \min \{x_{ik}\}}{\max \{x_{ik}\} - \min \{x_{ik}\}} \quad (1)$$

- Destymulanty

$$z_{ik} = \frac{\max \{x_{ik}\} - x_{ik}}{\max \{x_{ik}\} - \min \{x_{ik}\}} \quad (2)$$

- Nominanty – jedna wartość nominalna $\text{nom}\{x_{ik}\}$

$$z_{ik} = \begin{cases} \frac{x_{ik} - \min \{x_{ik}\}}{\text{nom}\{x_{ik}\} - \min \{x_{ik}\}}, & x_{ik} \leq \text{nom}\{x_{ik}\} \\ \frac{x_{ik} - \max \{x_{ik}\}}{\text{nom}\{x_{ik}\} - \max \{x_{ik}\}}, & x_{ik} > \text{nom}\{x_{ik}\} \end{cases} \quad (3)$$

- Nominanty – przedział wartości nominalnych $\langle \text{nom}_1 \{x_{ik}\}, \text{nom}_2 \{x_{ik}\} \rangle$

$$z_{ik} = \begin{cases} \frac{x_{ik} - \min \{x_{ik}\}}{\text{nom}_1 \{x_{ik}\} - \min \{x_{ik}\}}, & x_{ik} < \text{nom}_1 \{x_{ik}\} \\ 1, & \text{nom}_1 \{x_{ik}\} \leq x_{ik} \leq \text{nom}_2 \{x_{ik}\} \\ \frac{x_{ik} - \max \{x_{ik}\}}{\text{nom}_2 \{x_{ik}\} - \max \{x_{ik}\}}, & x_{ik} > \text{nom}_2 \{x_{ik}\} \end{cases} \quad (4)$$

Metody bezwzorcowe i wzorcowe

Metody bezwzorcowe i wzorcowe

- W pierwszym etapie dokonujemy wyboru cech opisujących obiekty poddane porządkowaniu oraz ustalamy ich wartości. Ustalone wartości cech dla wszystkich obiektów zestawia się w macierz danych, która stanowi punkt wyjścia w konstrukcji cechy agregatowej.

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nK} \end{pmatrix}, \quad (5)$$

x_{ik} – wartość k – tej cechy w i – tym obiekcie badania ($i = 1, \dots, n$; $k = 1, \dots, K$).

- Drugi etap w konstrukcji cechy syntetycznej polega na normalizacji cech diagnostycznych poprzez ujednolicenie ich charakteru (przekształcenie destymulant i nominant w stymulanty) oraz sprowadzenie ich wartości do porównywalności, co polega na pozbawieniu ich mian oraz ujednoliceniu rzędów wielkości – unitaryzacja zerowana.
- W tym etapie wyznaczamy wartość cechy syntetycznej, która bezpośrednio wykorzystywana jest na potrzeby uszeregowania obiektów. Obiekty posiadające wyższą wartość zmiennej syntetycznej znajdują się na wyższym miejscu w sporządzanym rankingu. Metody służące do wyznaczania wartości cechy syntetycznej można najogólniej podzielić na dwie grupy: **metody bezwzorcowe** i **metody wzorcowe**.

Metody bezwzorcowe

Metody bezwzorcowe

Metody bezwzorcowe polegają na uśrednieniu znormalizowanych wartości cech wyjściowych.

$$\bar{q}_i^{(1)} = \frac{\sum_{k=1}^K z_{ik}}{K}, \quad i = 1, \dots, n \quad (6)$$

Wartości tak wyznaczonej cechy syntetycznej $\bar{q}_i^{(1)}$ należą do przedziału $(0, 1)$. Czym wyższa jest wartość tego miernika tym obiekt znajdzie się na wyższym miejscu w rankingu.

Metody wzorcowe

Metody wzorcowe

Metody wzorcowe polegają na obliczaniu odległości poszczególnych jednostek od tzw. wzorca:

$$q_i^{(2)} = \sqrt{\frac{\sum_{i=1}^K (z_{ik} - z_{0k})^2}{K}}, \quad i = 1, \dots, n \quad (7)$$

gdzie z_{0k} jest znormalizowaną wartością k -tej cechy dla jednostki wzorcowej. Wzorec można przedstawić w postaci wektora $\mathbf{z} = (z_{01}, \dots, z_{0K})$, przy czym w większości praktycznych zastosowań przyjmuje się, że $z_{0k} = \max \{z_{ik}\}$ dla k -tej cechy diagnostycznej będącej lub przekształconej w stymulantę. Uzyskane w ten sposób wartości $\hat{q}_i^{(2)}$ mogą następnie posłużyć do obliczenia syntetycznego miernika rozwoju Hellwiga postaci:

$$\hat{q}_i = 1 - \frac{q_i^{(2)}}{q_0} \quad (8)$$

$$q_0 = \bar{q}_0 + 2s_0; \quad \bar{q}_0 = \frac{\sum_{i=1}^n q_i^{(2)}}{n}; \quad s_0 = \sqrt{\frac{\sum_{i=1}^n (q_i^{(2)} - \bar{q}_0)^2}{n}}. \quad (9)$$

Miernik \hat{q}_i przyjmuje najczęściej wartości z przedziału (0,1) a większe jego wartości są bardziej pożądane.

Literatura

Literatura

- Gatnar E., Walesiak M. (red.) (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2004.
- Grabiński T. (1992), *Metody taksonometrii*, Kraków.
- Nowak E. (1990), *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.
- Walesiak M. (1996), *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa.

Dziękuję za uwagę