

Marek Walesiak

Grażyna Dehnel

Wizualizacja wyników porządkowania liniowego z wykorzystaniem skalowania wielowymiarowego – przegląd badań

Konferencja naukowa

MET2023
Metodologia Badań
Statystycznych

3-5 lipca 2023 r., Warszawa



- 1. Cel referatu**
- 2. Charakterystyka metody hybrydowej**
- 3. Ogólna procedura metody hybrydowej**
- 4. Zastosowania metody hybrydowej – analiza porównawcza**
- 5. Wnioski końcowe**

1. Cel referatu

Referat ma charakter przeglądowny.

Jego celem jest szeroka prezentacja metody hybrydowej łączącej porządkowanie liniowe ze skalowaniem wielowymiarowym ze względu na:

- typ danych (metryczne, porządkowe, symboliczne interwałowe)
Dane klasyczne (macierz danych, kostka danych): dla każdej zmiennej obiekt opisany jest przez tylko jedną liczbę rzeczywistą (dane metryczne) lub jedną kategorię (dane porządkowe)
Dane symboliczne interwałowe (tablica danych): dla każdej zmiennej obiekt opisany jest przez przedział liczbowy
- zastosowane ujęcie: statyczne, dynamiczne
- źródło danych: pierwotne, wtórne
- rozwiązywane problemy badawcze

2. Charakterystyka metody hybrydowej (1)

- W pracy (Walesiak 2016) zaproponowano dwukrokową procedurę badawczą (metoda hybrydowa) pozwalającą na wizualizację wyników porządkowania liniowego dla danych metrycznych
- W pierwszym kroku w wyniku zastosowania skalowania wielowymiarowego otrzymuje się wizualizację obiektów w przestrzeni dwuwymiarowej. W kroku drugim przeprowadza się porządkowanie liniowe obiektów na podstawie odległości Euklidesa od wzorca rozwoju
- W metodzie hybrydowej wykorzystano koncepcję izokwant i ścieżki rozwoju (najkrótszej drogi łączącej wzorzec i antywzorzec rozwoju) zaproponowaną przez Profesora Zdzisława Hellwiga (1981)

2. Charakterystyka metody hybrydowej (2)

- Graficzna prezentacja wyników porządkowania liniowego w tej koncepcji możliwa była dla dwóch zmiennych. Zastosowanie skalowania wielowymiarowego rozszerzyło możliwości zastosowania wizualizacji wyników porządkowania liniowego dla m ($m > 2$) zmiennych
- W pracy (Walesiak 2017) przedstawiono modyfikację metody hybrydowej na dane porządkowe
- W pracy (Walesiak, Dehnel 2018) przedstawiono modyfikację metody hybrydowej na dane symboliczne interwałowe

3. Ogólna procedura metody hybrydowej (1)

1. Wybrać zjawisko złożone, które nie podlega pomiarowi bezpośredniemu (np. poziom spójności społecznej).
2. Ustalić zbiór obiektów oraz zbiór zmiennych (metrycznych, porządkowych lub symbolicznych interwałowych) merytorycznie związanych z badanym zjawiskiem złożonym.

Do zbioru obiektów dodaje się wzorzec (górną wartość rozwoju) oraz antywzorzec (dolną wartość rozwoju). Wśród zmiennych wyróżnia się zmienne preferencyjne (stymulanty, destymulanty, nominanty).

3. Zgromadzić dane i zestawić je w postaci:
 - a. Macierzy danych $\mathbf{X} = [x_{ij}]_{n \times m}$ dla zmiennych metrycznych lub porządkowych,
 - b. Tablicy danych $\mathbf{X} = [x_{ij}]_{n \times m}$ ($x_{ij} = [x_{ij}^l, x_{ij}^u]$, gdzie $x_{ij}^l \leq x_{ij}^u$) dla zmiennych symbolicznych interwałowych.

Wzorzec (górną wartość) obejmuje najkorzystniejsze wartości zmiennych, antywzorzec (dolną wartość) zaś najmniej korzystne wartości zmiennych preferencyjnych.

W przypadku zmiennych symbolicznych interwałowych współrzędne wyznacza się osobno dla dolnej i górnej wartości przedziału.

3. Ogólna procedura metody hybrydowej (2)

4. Przeprowadzić normalizację wartości zmiennych oraz zestawić dane w postaci znormalizowanej (**krok ten nie dotyczy danych porządkowych**):
 - a. Macierzy danych $\mathbf{Z} = [z_{ij}]_{n \times m}$ dla zmiennych metrycznych
 - b. Tablicy danych $\mathbf{Z} = [z_{ij}]_{n \times m}$ ($z_{ij} = [z_{ij}^l, z_{ij}^u]$, gdzie $z_{ij}^l \leq z_{ij}^u$, z_{ij} – znormalizowana obserwacja) dla zmiennych symbolicznych interwałowych.

Dane symboliczne interwałowe wymagają specjalnego podejścia do problemu normalizacji wartości zmiennych. Dolną i górną granicę interwału j -tej zmiennej dla n obiektów łączy się w jeden wektor zawierający $2n$ obserwacji. Takie podejście umożliwia zastosowanie metod normalizacji właściwych dla klasycznych danych metrycznych.

Normalizację zmiennych metrycznych przeprowadzono z wykorzystaniem funkcji **data.Normalization**, a zmiennych symbolicznych interwałowych z wykorzystaniem funkcji **interval_normalization** pakietu **clusterSim**.

3. Ogólna procedura metody hybrydowej (3)

5. Wybrać miarę odległości dla **danych metrycznych** (miejska, Euklidesa, Czebyszewa, kwadrat Euklidesa, GDM1), obliczyć odległości i zestawić w macierz odległości $\delta = [\delta_{ik}(\mathbf{Z})]_{n \times n}$ ($i, k = 1, \dots, n$).

Dla **danych porządkowych** zastosować odległość GDM2, obliczyć odległości i zestawić w macierz odległości $\delta = [\delta_{ik}(\mathbf{Z})]_{n \times n}$ ($i, k = 1, \dots, n$).

Dla **danych symbolicznych interwałowych** wybrać miarę odległości (Ichino-Yaguchi, euklidesowa Ichino-Yaguchi, Hausdorffa, euklidesowa Hausdorffa), obliczyć odległości i zestawić w macierz $\delta = [\delta_{ik}(\mathbf{Z})]_{n \times n}$.

Symbol	Nazwa	Miara odległości $\delta_{ik}(\mathbf{Z})$
U_2_q1	Ichino-Yaguchi $q = 1, \gamma = 0,5$	$\sum_{j=1}^m \varphi(z_{ij}, z_{kj})$
U_2_q2	Euklidesowa Ichino-Yaguchi $q = 2, \gamma = 0,5$	$\sqrt{\sum_{j=1}^m \varphi(z_{ij}, z_{kj})^2}$
H_q1	Hausdorffa $q = 1$	$\sum_{j=1}^m [\max(z_{ij}^l - z_{kj}^l , z_{ij}^u - z_{kj}^u)]$
H_q2	Euklidesowa Hausdorffa $q = 2$	$\left\{ \sum_{j=1}^m [\max(z_{ij}^l - z_{kj}^l , z_{ij}^u - z_{kj}^u)]^2 \right\}^{1/2}$

$z_{ij} = [z_{ij}^l, z_{ij}^u]$; $\varphi(z_{ij}, z_{kj}) = |z_{ij} \oplus z_{kj}| - |z_{ij} \otimes z_{kj}| + \gamma(2 \cdot |z_{ij} \otimes z_{kj}| - |z_{ij}| - |z_{kj}|)$; $| \quad |$ – długość interwału; $z_{ij} \oplus z_{kj} = z_{ij} \cup z_{kj}$; $z_{ij} \otimes z_{kj} = z_{ij} \cap z_{kj}$

3. Ogólna procedura metody hybrydowej (4)

6. Przeprowadzić skalowanie wielowymiarowe (MDS): $f: \delta_{ik}(\mathbf{Z}) \rightarrow d_{ik}(\mathbf{V})$ dla wszystkich par (i, k) , gdzie f oznacza odwzorowanie odległości z przestrzeni m -wymiarowej $\delta_{ik}(\mathbf{Z})$ w odpowiadające odległości $d_{ik}(\mathbf{V})$ w przestrzeni q -wymiarowej ($q < m$).

W artykule zastosowano rozwiązanie pozwalające na wybór optymalnej procedury skalowania wielowymiarowego, ze względu na zastosowane:

- metody normalizacji wartości zmiennych (**nie dotyczy danych porządkowych**)
- miary odległości,
- modele skalowania (ratio, interval mspline),

zgodnie z procedurami (dla danych metrycznych oraz symbolicznych interwałowych) dostępnymi w pakiecie `mdsOpt` (Walesiak, Dudek 2018) programu R.

3. Ogólna procedura metody hybrydowej (5)

The evaluation of tourist attractiveness of Lower Silesian counties

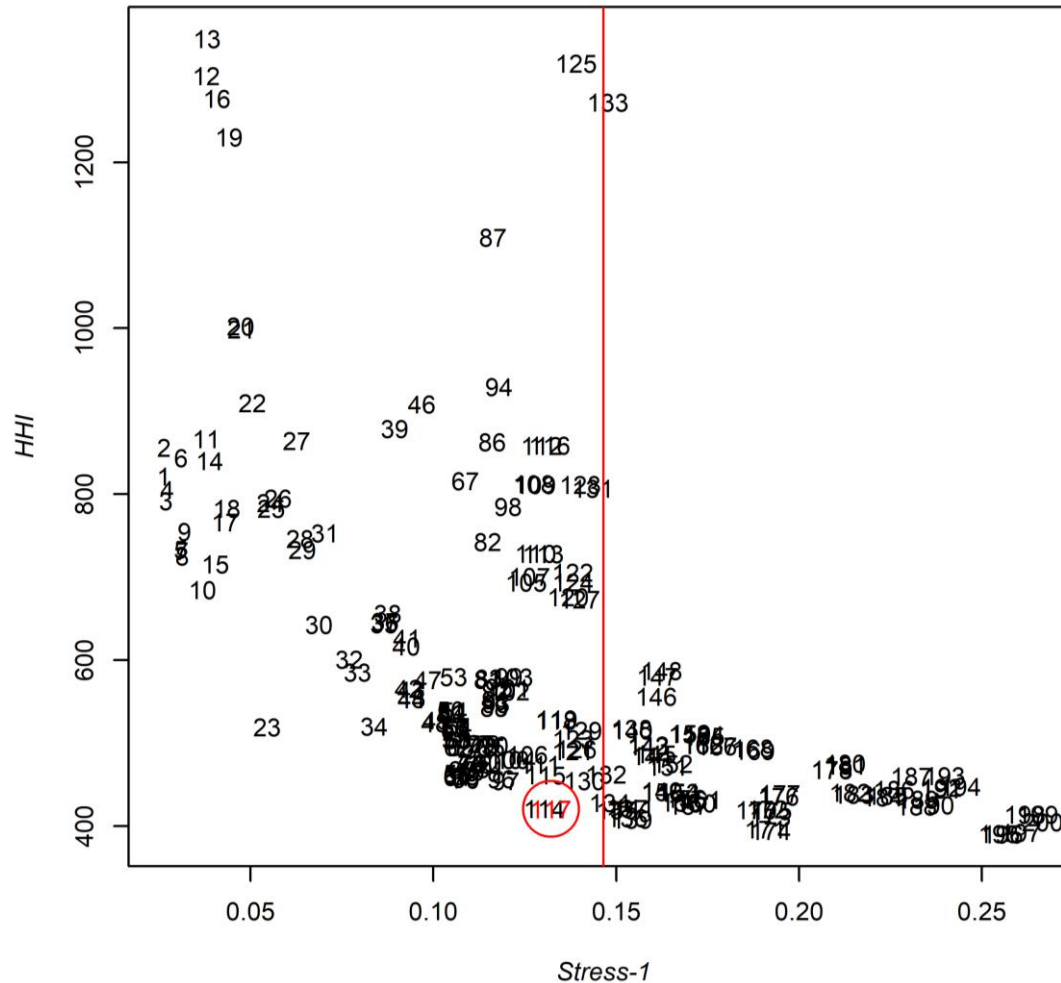


Fig. 1. The values of *Stress-1* fit measure and *HHI* index for *p* MDS procedures (10 normalization methods, 5 distance measures, 4 MDS models) – with best solution marked by red circle 10/20

3. Ogólna procedura metody hybrydowej (6)

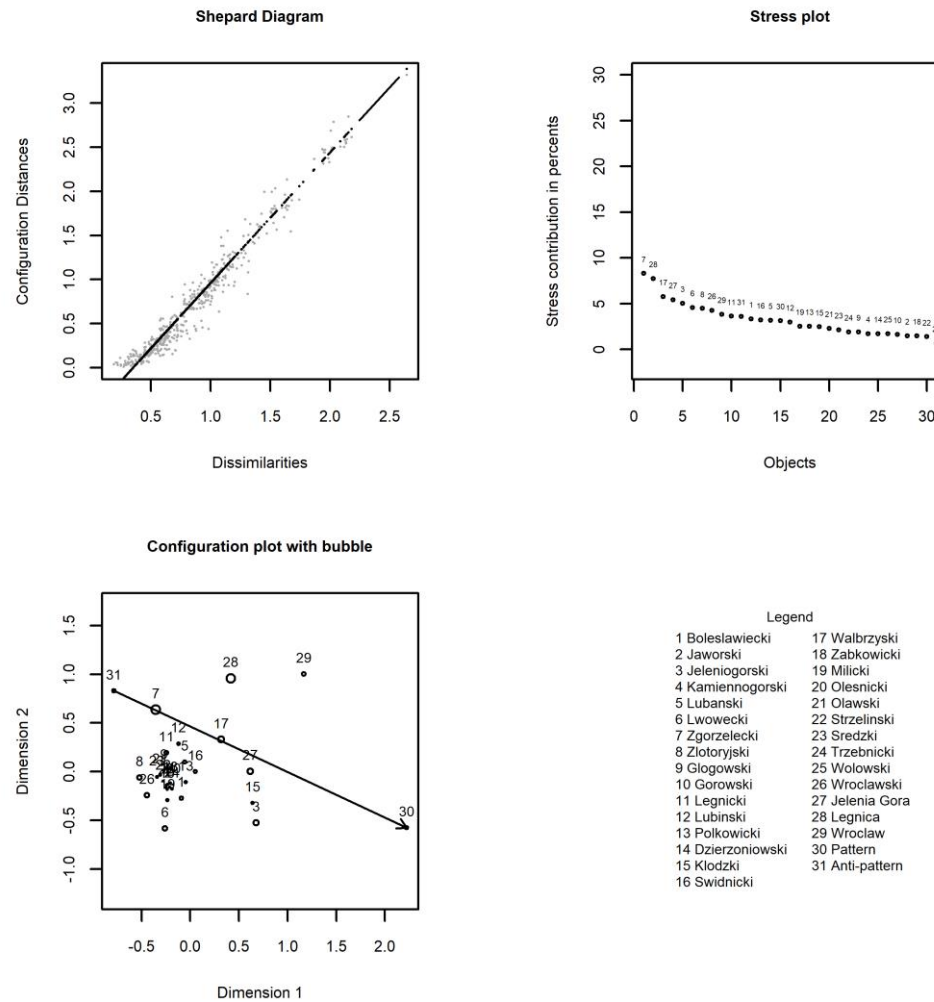


Fig. 2. The results of MDS (procedure 117 – n12a normalization method, Euclidean distance and interval MDS model) of 31 objects according to the level of tourist attractiveness

3. Ogólna procedura metody hybrydowej (7)

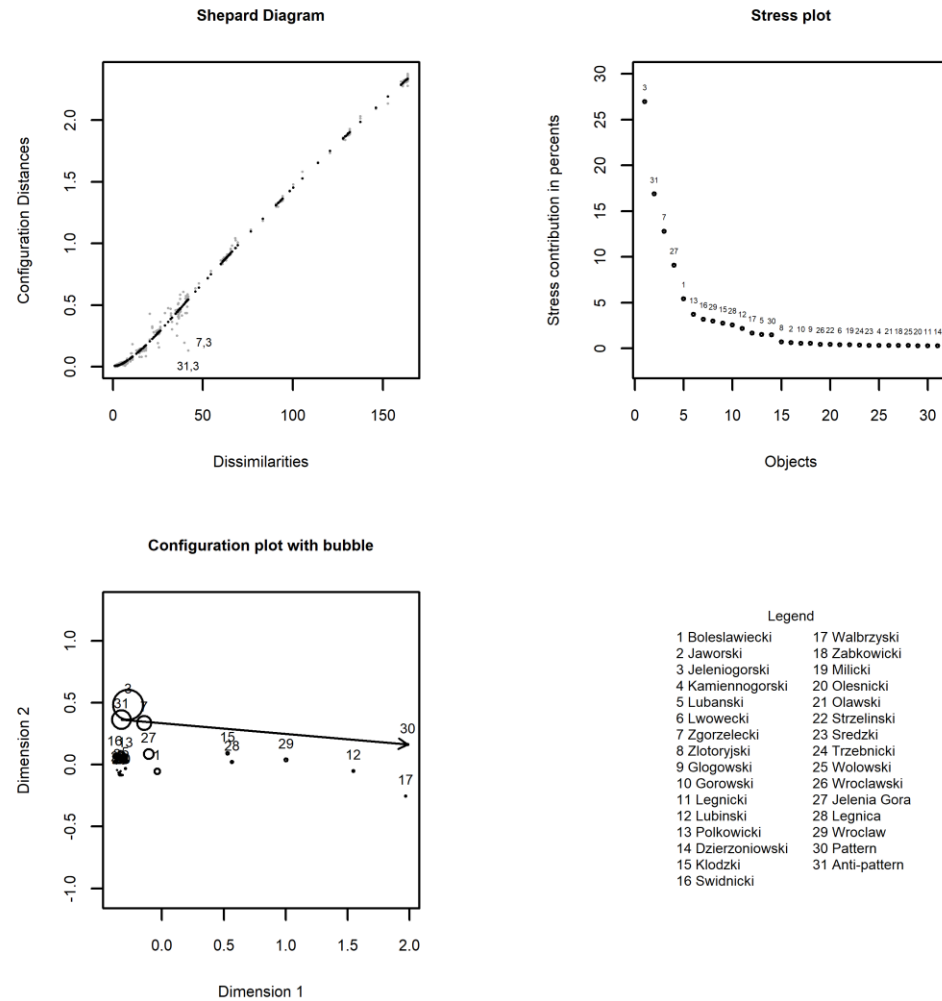


Fig. 3. The results of MDS (procedure 13 – n9a normalization method, mspline of third degree model, Chebyshev distance) of 31 objects according to the level of tourist attractiveness

3. Ogólna procedura metody hybrydowej (8)

7. Ostatecznie w wyniku zastosowania skalowania wielowymiarowego otrzymuje się macierz danych w przestrzeni dwuwymiarowej $\mathbf{V} = [v_{ij}]_{n \times q}$ ($q = 2$).

W zależności od położenia wzorca i antywzorca w dwuwymiarowej przestrzeni skalowania $\mathbf{V} = [v_{ij}]_{n \times 2}$ wymagany jest obrót układu współrzędnych o kąt φ zgodnie ze wzorem:

$$[v'_{ij}]_{n \times 2} = [v_{ij}]_{n \times 2} \times D$$

gdzie: $[v'_{ij}]_{n \times 2}$ – macierz danych w dwuwymiarowej przestrzeni skalowania po obrocie układu współrzędnych o kąt φ ,

$$D = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \text{ – macierz obrotu.}$$

Obrót ten nie zmienia rozmieszczenia obiektów względem siebie. Pozwala na umieszczenie osi zbioru, łączącej antywzorzec i wzorzec, wzdłuż głównej przekątnej, co daje korzystniejszą wizualizację wyników.

3. Ogólna procedura metody hybrydowej (9)

- Przedstawić graficznie oraz zinterpretować wyniki w przestrzeni dwuwymiarowej (wyniki skalowania wielowymiarowego).

Na rysunku w przestrzeni dwuwymiarowej łączy się linią prostą punkty oznaczające antywzorzec i wzorzec w tzw. oś zbioru. Wyznacza się od punktu wzorca izokwanty rozwoju (krzywe jednakowego rozwoju). Obiekty znajdujące się pomiędzy izokwantami prezentują zbliżony poziom rozwoju. Ten sam poziom rozwoju mogą osiągnąć obiekty znajdujące się w różnych punktach na tej samej izokwancie rozwoju (z uwagi na inną konfigurację wartości zmiennych).

- Uporządkować obiekty na podstawie wartości miary agregatywnej d_i bazującej na odległości Euklidesa od obiektu wzorca (Hellwig, 1981):

$$d_i = 1 - \frac{\sqrt{\sum_{j=1}^2 (v_{ij} - v_{+j})^2}}{\sqrt{\sum_{j=1}^2 (v_{+j} - v_{-j})^2}}$$

v_{ij} – j -ta współrzędna dla i -tego obiektu w dwuwymiarowej przestrzeni MDS,

v_{+j} (v_{-j}) – j -ta współrzędna dla obiektu wzorca (antywzorca) w dwuwymiarowej przestrzeni MDS.

Wartości miary agregatywnej d_i należą do przedziału $[0; 1]$. Im wyższa jest wartość d_i , tym wyższy jest poziom spójności społecznej badanych obiektów. Porządkuje się obiekty badania według malejących wartości miary agregatywnej d_i .

4. Zastosowania metody hybrydowej – analiza porównawcza (1)

Lp.	Publikacja	Typ danych	Ujęcie	Źródło danych	Problem badawczy
1	Walesiak M. (2016) „Ekonometria”, 2(52), 9-21	metryczne	statyczne	wtórne GUS	Ocena atrakcyjności turystycznej 29 powiatów województwa dolnośląskiego w roku 2014
2	Walesiak M. (2017) „Przegląd Statystyczny”, zeszyt 1, Tom 64, 5-19	porządkowe	statyczne	pierwotne	Przeprowadzenie porządkowania liniowego 27 nieruchomości lokalowych opisanych 6 zmiennymi porządkowymi na jeleniogórskim rynku nieruchomości ze względu na ich atrakcyjność
3	Walesiak M. (2017) „Ekonometria”, 3(57), 9-25	metryczne	dynamiczne	wtórne GUS	Pomiar i ocena zmian poziomu spójności społecznej województwa dolnośląskiego w przekroju powiatów w latach 2005-2015
4	Walesiak M., Dudek A. (2017) „STATISTICS IN TRANSITION new series”, September, Vol. 18, No. 3, 521-540	metryczne	statyczne	wtórne GUS	Ocena atrakcyjności turystycznej 29 powiatów województwa dolnośląskiego w roku 2014
5	Dehnel G., Gołata E., Obrębalski M., Walesiak M. (2018) Taksonomia 30, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 507, 39-52	metryczne	dynamiczne	wtórne Eurostat	Ocena zmian w procesie starzenia się ludności w 35 regionach państw Grupy Wyszehradzkiej w latach 2005–2016

4. Zastosowania metody hybrydowej – analiza porównawcza (2)

Lp.	Publikacja	Typ danych	Ujęcie	Źródło danych	Problem badawczy
6	Walesiak M., Dehnel G. (2018) In M. Papież and S. Śmiech (Eds.), The 12th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings. Cracow: Foundation of the Cracow University of Economics, 563-572	symboliczne interwałowe	statyczne	wtórne GUS	Ocena efektywności ekonomicznej małych przedsiębiorstw przemysłowych (o liczbie pracujących do 10 do 49 osób) w powiatach województwa wielkopolskiego na podstawie danych symbolicznych interwałowych
7	Dehnel G., Walesiak M. (2019) „STATISTICS IN TRANSITION new series”, June, Vol. 20, No. 2, 49-67	metryczne, symboliczne interwałowe	statyczne	wtórne GUS	Ocena efektywności ekonomicznej małych przedsiębiorstw przemysłowych (o liczbie pracujących do 10 do 49 osób) w powiatach województwa wielkopolskiego na podstawie danych metrycznych oraz symbolicznych interwałowych
8	Dehnel G., Walesiak M., Obrębalski M. (2019) “Argumenta Oeconomica Cracoviensia”, No 1(20), 71-85.	metryczne, symboliczne interwałowe	statyczne	wtórne GUS	Analiza porównawcza uporządkowania województw Polski ze względu na spójność społeczną na podstawie klasycznych danych metrycznych oraz symbolicznych interwałowych z roku 2016
9	Obrębalski M., Walesiak M. (2019) “Wiadomości Statystyczne. The Polish Statistician”, vol. 64, 12, 7–26	metryczne	dynamiczne	wtórne Eurostat	Pomiar zakresu i stopnia zróżnicowania sytuacji młodzieży na rynku pracy w przygranicznych regionach Polski, Czech i Niemiec w latach 2010 i 2018 z wykorzystaniem sześciu zmiennych
10	Walesiak M., Dehnel G. (2019) “Econometrics Ekonometria. Advances in Applied Data Analysis”, vol. 23, no 4, 1-15	metryczne	dynamiczne	wtórne GUS	Porównanie stopnia zaawansowania procesu starzenia się ludności województw Polski w latach 2002, 2010 i 2017 z wykorzystaniem zmiennych: mediana wieku, współczynnik starości, wskaźnik podwójnego starzenia, indeks starości, współczynnik obciążenia osobami starszymi

4. Zastosowania metody hybrydowej – analiza porównawcza (3)

Lp.	Publikacja	Typ danych	Ujęcie	Źródło danych	Problem badawczy
11	Walesiak M., Dehnel G. (2019) In M. Papież and S. Śmiech (Eds.), The 13th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Warszawa: Wydawnictwo C.H. Beck, 250-258.	metryczne, symboliczne interwałowe	statyczne	wtórne GUS	Analiza porównawcza uporządkowania województw Polski ze względu na spójność społeczną na podstawie danych metrycznych oraz symbolicznych interwałowych (3 typy danych) z roku 2016
12	Walesiak M., Dehnel G. (2020) „Sustainability”, 12(18), 7664, 1-19	metryczne, symboliczne interwałowe	statyczne	wtórne GUS	Analiza porównawcza uporządkowania województw Polski ze względu na spójność społeczną na podstawie danych metrycznych oraz symbolicznych interwałowych (3 typy danych) z roku 2018.
13	Dehnel G., Gołata E., Walesiak M. (2020) „Argumenta Oeconomica”, No 1 (44), 77-100.	metryczne	dynamiczne	wtórne Eurostat	Ocena zmian w procesie starzenia się ludności w regionach krajów Grupy Wyszehradzkiej w roku 2016 w porównaniu z 2005 rokiem
14	Walesiak M., Dudek A. (2020) [In:] K. S. Soliman (Ed.), Education Excellence and Innovation Management: A 2025 Vision to Sustain Economic Development during Global Challenges, Proceedings of the 35th IBIMA, 307-324,, 1-2 April 2020, Seville, Spain	metryczne, symboliczne interwałowe	statyczne	wtórne GUS	Ocena atrakcyjności turystycznej powiatów województwa dolnośląskiego na podstawie danych metrycznych Ocena atrakcyjności turystycznej województw Polski na podstawie danych symbolicznych interwałowych

4. Zastosowania metody hybrydowej – analiza porównawcza (4)

Charakterystyka zastosowań empirycznych metody hybrydowej

1. Publikacje

- 11 artykułów, w tym 2 artykuły z listy JCR
- 5 pozycji w bazie SCOPUS i 5 w bazie Web of Science
- 3 pozycje w materiałach konferencyjnych z konferencji zagranicznych
- 11 pozycji w języku angielskim

2. Typ danych

- 12 pozycji – dane metryczne
- 1 pozycja – dane porządkowe
- 6 pozycji – dane symboliczne interwałowe

3. Ujęcie

- statyczne – 9 pozycji
- dynamiczne – 5 pozycji

4. Źródło danych

- pierwotne – 1 pozycja
- wtórne – 13 pozycji (GUS – 10, Eurostat – 3)

5. Wnioski końcowe

- W referacie zaprezentowano przegląd rozwiązań metodycznych w zakresie metody hybrydowej będącej połączeniem skalowania wielowymiarowego i porządkowania liniowego
- Skalowanie wielowymiarowe pozwala na wizualizację wyników porządkowania liniowego obiektów w przestrzeni dwuwymiarowej
- Wizualizację wzbogaca się o izokwanty rozwoju (krzywe jednakowego rozwoju) oraz ścieżkę rozwoju (tj. najkrótszą drogę łączącą wzorzec i antywzorzec rozwoju)
- Przedstawiono przegląd zawierający analizę porównawczą zastosowań metody hybrydowej ze względu na typ danych (metryczne, porządkowe, symboliczne interwałowe), zastosowane ujęcie (statyczne, dynamiczne), źródło danych (pierwotne, wtórne) oraz rozwiązywane problemy badawcze
- Oprogramowaniem wspomagającym zastosowanie metody hybrydowej jest pakiet mdsOpt w środowisku R autorstwa M. Walesiaka i A. Dudka

Dziękujemy za
uwagę