

# AUTOREFERAT

1. Imię i Nazwisko: Jan W. Owsiański

2. Posiadany stopień naukowy: dr nauk technicznych w dziedzinie informatyki, uzyskany, z wyróżnieniem, w Instytucie Badań Systemowych PAN w roku 1992 za rozprawę doktorską pt. „*Nowa metoda analizy skupień z globalną funkcją celu*”

3. Informacja o dotychczasowym zatrudnieniu w jednostkach naukowych: w całym okresie od doktoratu (tak, jak i przedtem) jestem zatrudniony w Instytucie Badań Systemowych PAN. Jednocześnie od 2004 roku jestem wykładowcą w Wyższej Szkole Informatyki Stosowanej i Zarządzania w Warszawie.

4. Wskazanie osiągnięcia naukowego:

Monografia (załączona):

**Owsiański, J.W.:** *Data Analysis in Bi-partial Perspective: Clustering and Beyond. Studies in Computational Intelligence 818*. Springer Nature Switzerland, 2020. ISBN 978-3-030-13388-7

## Zasadnicza treść pracy

Praca ta stanowi podsumowanie głównego nurtu badań wnioskodawcy w zakresie metodyki analizy danych, w szczególności zaś – analizy skupień. W innych, licznych publikacjach wnioskodawcy, związanych z tym tematem (por. Załącznik 4), rozpatrywane są zagadnienia cząstkowe, zastosowania (na ogół także mające charakter metodyczny), lub związki z innymi zagadnieniami.

W przedmiotowej pracy przedstawiono autorskie podejście do analizy skupień, które może także być stosowane do szeregu innych zagadnień z zakresu analizy danych, co w pracy również zilustrowano. Podejście to opiera się na zaproponowaniu ogólnej postaci funkcji celu analizy skupień, odzwierciedlającej wiernie samo podstawowe sformułowanie zadania analizy skupień, tj. „*dokonać podziału zbioru obiektów tak, by obiekty w tych samych podzbiorach (skupieniach) były możliwie bliskie sobie (podobne), zaś obiekty, umieszczone w różnych skupieniach – możliwie dalekie (niepodobne)*”. Zaproponowana funkcja celu (jakości podziału zbioru obiektów) składa się zatem z dwóch części: jednej, odzwierciedlającej podobieństwo (lub zróżnicowanie) obiektów wewnątrz skupień oraz drugiej, odzwierciedlającej zróżnicowanie (bądź podobieństwo) obiektów w różnych skupieniach. Pokazano szereg konkretnych realizacji tej funkcji celu dla różnych wariantów modelowanej sytuacji, w tym także spoza właściwej analizy skupień.

Należy podkreślić, że odpowiednie zaprojektowanie zaproponowanej funkcji celu pozwala w zasadzie zrezygnować z wykorzystania wielu używanych w analizie skupień kryteriów jakości podziału, w ogromnej większości przypadków nie mających nic wspólnego z algorytmami, przy pomocy których oceniane podziały zostały otrzymane.

Pokazano dalej, że dla zaproponowanej funkcji celu można pod pewnymi, dość realistycznymi, warunkami, zaprojektować algorytm suboptymalizujący, przy czym algorytm ten może także przyjmować różne konkretne postacie, w zależności od przyjętej postaci funkcji celu. Niezmiernie ważną własnością otrzymanych w ten sposób różnych wariantów algorytmu jest ich pełna analogiczność do klasycznych algorytmów agregacji hierarchicznej (najbliższego sąsiada, najdalszego sąsiada, itp.), z jednoczesnym odniesieniem do konkretnej funkcji celu, co pozwala na ustanowienie „warunku stopu” tych algorytmów.

Opracowana metodyka może być stosowana elastycznie do różnych modeli zadania analizy skupień, jak również innych zagadnień, dla których sensowne jest sformułowanie analogicznej funkcji celu. Ograniczeniem tej metodyki są własności numeryczne algorytmów, także analogiczne do własności klasycznych algorytmów agregacji hierarchicznej. Tym niemniej, możliwe jest jej wykorzystanie w metodykach hybrydowych, podobnie, jak to ma często miejsce obecnie z algorytmami agregacji hierarchicznej.

### Opis ogólnej metodyki

Mamy do czynienia z  $n$  obiektami (obserwacjami), ponumerowanymi indeksem  $i$ ,  $i \in I = \{1, \dots, n\}$ . Obiekty te są opisane wektorami  $x_i = [x_{i1}, \dots, x_{ik}, \dots, x_{im}]$ , tj. wektorami  $m$  wartości zmiennych (atrybutów, cech), ponumerowanych indeksem  $k$ . Nie zakładamy niczego co do skal poszczególnych zmiennych, zaś ich dopuszczalne wartości tworzą przestrzeń, w której znajdują się opisy obiektów, oznaczoną  $E_X$ . Dla tak określonych (opisów) obiektów dysponujemy pewną definicją odległości  $d(\cdot, \cdot)$  i możliwością wyznaczenia tej odległości dla wszystkich par rozpatrywanych obiektów i dla wszystkich par punktów w  $E_X$ . Dla uproszczenia, odległości między danymi obiektami oznaczamy  $d_{ij}$ ,  $i, j \in I$ .

Dla takich danych rozwiązujemy przytoczone uprzednio zadanie analizy skupień: „*dokonać podziału zbioru obiektów tak, by obiekty, umieszczone w tych samych podzbiorach (skupieniach) były możliwie bliskie siebie (podobne), zaś obiekty, umieszczone w różnych skupieniach – możliwie dalekie (niepodobne)*”. W tym celu posługujemy się funkcją celu (jakości podziału) o postaci:

$$Q_S^D(P) = Q_S(P) + Q^D(P),$$

gdzie  $P$  jest podziałem zbioru  $I$  na podzbiory (skupienia)  $A_q$ ,  $q = 1, \dots$ , rozłączne i wyczerpujące zbiór  $I$ ,  $Q_S(P)$  jest komponentem funkcji celu, odzwierciedlającym podobieństwo obiektów wewnątrz skupień<sup>1</sup>, zaś  $Q^D(P)$  jest komponentem, odzwierciedlającym odległości pomiędzy obiektami w różnych skupieniach.  $Q_S^D(P)$  jest maksymalizowana względem  $P$ , ale można także rozpatrywać jej postać „dualną”,  $Q_D^S(P)$ , minimalizowaną:

$$Q_D^S(P) = Q_D(P) + Q^S(P).$$

Liczba skupień w podziale wynosi  $p$  ( $q = 1, \dots, p$ ), ale nie przyjmujemy żadnego założenia odnośnie tej wartości.

---

<sup>1</sup> Co jest także związane z faktem wprowadzenia w tej metodyce jawnej funkcji „bliskości”, czy „podobieństwa”,  $s(d)$ .

Klasycznym przykładem komponentu  $Q_D(P)$  jest funkcja celu bardzo znanej metody k-means, a więc funkcja

$$Q_D(P) = \sum_q \sum_{i \in A_q} (d(x_i, x^q))^2$$

w której  $x^q$  są „reprezentantami” skupień  $A_q$ , np. średnimi w skupieniu. Ogólnie, zakłada się, że komponenty funkcji  $Q_S^D(P)$ , względnie  $Q_D^S(P)$ , spełniają wymagania racjonalności, wynikające ze sformułowania zadania analizy skupień. W celu uzyskania efektywnego algorytmu nakładane są dodatkowe warunki, przede wszystkim związane z zachowaniem tych komponentów w ramach (binarnej) hierarchii podziałów, tj. zbioru podziałów, w którym kolejny podział powstaje przez agregację dwóch skupień z poprzedniego, zaś podział początkowy jest podziałem na  $n$  skupień, odpowiadających poszczególnym obiektom (tj.  $p = n$ ). Zakładamy, mianowicie, że wzdłuż tak rozumianej hierarchii wartości wspomnianych komponentów obu dualnych funkcji celu są wzajemnie przeciwnie monotoniczne (jeśli jeden rośnie, to drugi maleje). Łatwo zauważyć, że np. przytoczona funkcja celu metody k-means ma dla  $p = n$  po prostu wartość = 0. Wartość tej funkcji rośnie dla kolejnych, coraz niższych  $p$ , nawet, jeśli rozważamy kolejne optymalne podziały dla malejącego ciągu wartości  $p$ , a nie tylko hierarchię podziałów.

Dla takich komponentów funkcji celu formułujemy jej sparаметryzowaną postać:

$$Q_S^D(P, r) = rQ_S(P) + (1-r)Q_D^S(P), \text{ lub}$$

$$Q_D^S(P, r) = (1-r)Q_D(P) + rQ^S(P),$$

w której  $r \in [0, 1]$ . To dla takiej postaci funkcji celu można sformułować ogólny algorytm, którego podstawą jest zmiana wartości parametru  $r$  (wagi dwóch komponentów, odzwierciedlających dwie „strony” zadania analizy skupień) od 0 do 1, lub odwrotnie.

Przyjrzyjmy się przykładowemu schematowi przebiegu odpowiedniej procedury. Załóżmy, że minimalizujemy  $Q_D^S(P, r)$  i zaczynamy procedurę od  $r^0 = 0$ . Łatwo zauważyć, że przy tej wartości parametru  $r$  optimum osiągnięte jest dla  $P^{\text{opt}}(r=0) = P^0$ , stanowiącego przez skupienia, będące pojedynczymi obiektami ( $p^0 = n$ ), jako, że wówczas  $Q_D^S(P, 0) = Q_D(P)$  jest minimalne (w przypadku k-means = 0).

Mając zatem podział  $P^0$  (ogólniej:  $P^t$ , gdzie  $t$  jest indeksem kolejnego podziału, otrzymanego w procedurze), szukamy najmniejszego  $r$ , większego od poprzedniego, zakładając, że wszystkie  $d_{ij} > 0$ ,  $i, j \in I$ , dla którego wartość  $Q_D^S(P, r)$  jest mniejsza, niż wartość  $Q_D^S(P^{t-1}, r)$ . Szukamy tego nowego podziału wśród podziałów, utworzonych z  $P^{t-1}$  przez agregację dwóch skupień, tworzących ten poprzedni podział. Granicznym warunkiem jest warunek

$$(1-r)Q_D(P^{t-1}) + rQ^S(P^{t-1}) = (1-r)Q^D(P^{Ht-1}(q, q')) + rQ_S(P^{Ht-1}(q, q')) \quad [*]$$

w którym  $P^{Ht}(q, q')$  jest podziałem utworzonym z  $P^t$  przez agregację skupień  $A_q$  oraz  $A_{q'}$ . Najniższe  $r$ , spełniające ten warunek, staje się nowym  $r^t$ , a odpowiadająca tej wartości para skupień jest agregowana, aby utworzyć  $P^t$ .

Zakładamy, dodatkowo, że dysponujemy definicjami odległości i podobieństwa na poziomie skupień,  $D(A_q, A_{q'})$ ,  $S(A_q, A_{q'})$ ,  $D(A_q)$ ,  $S(A_q)$ . I znów, dobry przykład jednej z takich funkcji można zaczerpnąć z metody k-means, a mianowicie:

$$D(A_q) = \sum_{i \in A_q} (d(x_i, x^q))^2.$$

Należy podkreślić, że tego rodzaju definicje są formułowane, czy to w sposób jawny, czy nie, w wielu metodykach analizy skupień, w tym, między innymi, dla algorytmów agregacji hierarchicznej.

Jest naturalnym, że komponenty funkcji celu są bezpośrednio zależne od owych odległości i bliskości na poziomie skupień. W szczególności, dotyczy to założenia o monotoniczności, w tym i monotoniczności wzdłuż hierarchii. Dla takich założeń, wartości  $Q_D^S(P^t, r)$  w funkcji  $r$  tworzą funkcję wypukłą, kawałkami liniową.

Jeśli dodatkowo założymy, że komponenty funkcji celu są po prostu addytywne względem odpowiednich  $D(A_q, A_q')$ ,  $S(A_q, A_q')$ ,  $D(A_q)$ ,  $S(A_q)$ , wówczas warunek [\*], na podstawie którego wyznaczone są kolejne  $r^t$ , a zatem i pary skupień do agregacji, przyjmuje postać „reguły minimalnej odległości”, stanowiącej podstawę działania klasycznych algorytmów agregacji hierarchicznej.

Zakładając, że komponenty funkcji celu mają równą sobie wagę (kwestia odpowiedniego, „racjonalnego”, zaprojektowania funkcji celu), zarysowaną procedurę stopujemy w momencie, gdy kolejne  $r^t$  przekroczy wartość  $\frac{1}{2}$ .

Suboptymalność otrzymanych rozwiązań wynika, oczywiście, z ograniczenia operacji do łączenia skupień. Jednakże odwołanie do jawnej postaci funkcji celu powoduje, że możliwe jest uzupełnienie procedury o dodatkowe operacje, poprawiające i weryfikujące jakość otrzymanych rozwiązań.

Powyższe rozumowanie może zostać skonkretyzowane dla różnych postaci funkcji celu<sup>2</sup>, opartych na różnych założeniach i intuicjach co do jakości podziałów i składających się na nie skupień, w tym także prowadząc do różnych konkretnych postaci algorytmu suboptymalizacji. Wybrane przykłady zostaną pokazane w następnym punkcie niniejszego dokumentu.

\*\*\*\*\*

Warto zwrócić uwagę na następujące ważne cechy opracowanego podejścia:

- (1) *ogólne sformułowanie funkcji celu o charakterze globalnym, która w istocie powoduje, że nie jest konieczne odwoływanie się do popularnych wskaźników jakości podziału (klasteringu), wskaźników najczęściej nie mających nic wspólnego z algorytmem, przy pomocy którego analizowany podział został otrzymany;*
- (2) *możliwość tworzenia dla tego ogólnego sformułowania wielu konkretnych postaci funkcji celu, w tym także dla zagadnień, wykraczających poza dziedzinę analizy skupień;*
- (3) *brak konieczności osobnego zajmowania się często podnoszoną kwestią liczby skupień, ponieważ liczba skupień stanowi integralną część otrzymywanego rozwiązania;*
- (4) *powiązanie z funkcją celu prostego algorytmu suboptymalizującego, a w zasadzie rodziny algorytmów, mających postać analogiczną do klasycznych algorytmów agregacji hierarchicznej.*

---

<sup>2</sup> Przedstawiony tutaj szkic rozumowania musi być, naturalnie, odpowiednio zmodyfikowany dla „dualnej” postaci funkcji celu, ewentualnie dla alternatywnej (odwrotnej) jej parametryzacji.

## Przykłady konkretnych realizacji ogólnej funkcji celu i algorytmu

Przytoczymy obecnie wybrane przykłady, zarówno sformułowania funkcji celu, jak i odpowiadających im algorytmów subotymalizacji.

### 1. Addytywna funkcja celu

Podstawowym przykładem realizacji zarysowanych tutaj zasad budowy funkcji celu i odpowiadającego jej algorytmu jest następująca funkcja:

$$Q^D_S(P) = C_S(P) + C^D(P) = \sum_q \sum_{i < j \in A_q} s_{ij} + \sum_q \sum_{q' > q} \sum_{i \in A_q} \sum_{j \in A_{q'}} d_{ij}.$$

która jest, naturalnie, maksymalizowana. Funkcja ta była początkiem opisywanych tutaj prac i pojawiła się w szeregu publikacji wnioskodawcy jeszcze z początku lat 80-tych XX w. (Owsiński, 1980, 1981, 1984ab, 1986). Należy także wspomnieć, że prace nad tą samą postacią funkcji, jednakże optymalizowaną w formie zadania programowania matematycznego, były równoległe prowadzone przez francuskich naukowców, F. Marcotorchino i P. Michaud.

Dla tej funkcji, jak widać, przyjęto definicje

$$S(A) = \frac{1}{2} \sum_{i,j \in A} s_{ij}, \text{ oraz } D(A,B) = \sum_{i \in A, j \in B} d_{ij},$$

a także

$$S(A,B) = \sum_{i \in A, j \in B} s_{ij} \text{ oraz } D(A) = \frac{1}{2} \sum_{i,j \in A} d_{ij},$$

zaś warunek na wartość parametru  $r$ , dla już ustalonego podziału  $P^t$ , w którym występują skupienia  $A_q$  i  $A_{q'}$ , ma postać

$$r^{t+1}(q, q') = \frac{S(A_q, A_{q'})}{S(A_q, A_{q'}) + D(A_q, A_{q'})}.$$

Jest oczywistym, że oznacza to szukanie, w każdym kolejnym kroku, pary najbliższych sobie skupień, które będą agregowane, identycznie jak w klasycznych algorytmach agregacji hierarchicznej. Należy podkreślić, że otrzymana procedura jest identyczna z klasyczną procedurą „*average link*”, jednakże wyposażona jest ona w możliwość pomiaru jakości podziału, a zatem i kryterium stopu.

### 2. Funkcja celu addytywnych podobieństw i stałego kosztu skupienia

W tym przypadku mamy do czynienia z nieco specyficzną funkcją celu – przytaczamy ją tutaj także dla zaznaczenia możliwości reprezentowania w proponowanym podejściu szerokiej gamy różnych modeli analizy skupień, w tym i całkowicie niekonwencjonalnych. Mogą one być, w szczególności, związane z takimi rodzajami zastosowań analizy skupień, niekoniecznie wynikającymi wprost z zagadnień analizy danych, jak powiązane z badaniami operacyjnymi. Ta funkcja ma postać:

$$Q^D_S(P) = C_S(P) + C^D(P) = \sum_q \sum_{i < j \in A_q} s_{ij} + pP,$$

gdzie  $pP$  jest liczbą skupień w podziale  $P$ . Postać ta spełnia warunki metody, a jej interpretacja może być, na przykład, związana z zagadnieniem lokalizacji. Dla tej funkcji, warunek na kolejną wartość parametru  $r$  dla już istniejącego podziału  $P^t$  jest następujący:

$$r^{t+1}(q, q') = \frac{S(A_q, A_{q'})}{S(A_q, A_{q'}) + 1},$$

co, biorąc pod uwagę fakt, że definicja  $S(A, B)$  pozostała taka sama jak w poprzednim przykładzie, oznacza, że, tak, jak i poprzednio, w każdym kolejnym kroku szukamy pary najbliższych sobie skupień, które zostaną następnie połączone w nowym podziale.

### 3. Funkcja celu skrajnych odległości i bliskości

Dla tego przypadku przyjmujemy następujące definicje komponentów funkcji celu:

$$Q^D(P) = \sum_{q=1}^{p-1} \sum_{q'=q+1}^p D(A_q, A_{q'}) \quad \text{oraz} \quad Q^S(P) = \sum_{q=1}^p \text{card} A_q S(A_q)$$

przy czym

$$D(A, B) = \min_{i \in A, j \in B} d_{ij} \quad \text{oraz} \quad S(A) = \max_{i, j \in A} s_{ij}.$$

Przyjęcie powyższych definicji, skądinąd całkowicie racjonalnych z punktu widzenia sformułowania zadania analizy skupień, powoduje, że wyznaczenie reguły szukania wartości parametru  $r$  staje się nieco bardziej skomplikowane niż w poprzednich dwóch przypadkach. Dlatego też najpierw wyrazimy tę wartość w nieco ogólniejszej postaci, a mianowicie

$$r(q, q') = \frac{\Delta Q^S(q, q')}{\Delta Q^S(q, q') + \Delta Q^D(q, q')}.$$

Wielkości, występujące w powyższym wzorze, wyliczane są w następujący sposób:

$$\Delta Q^S(q, q') = (\text{card} A_q + \text{card} A_{q'}) S(A_q \cup A_{q'}) - \text{card} A_q S(A_q) - \text{card} A_{q'} S(A_{q'})$$

oraz

$$\Delta Q^D(q, q') = D(A_q, A_{q'}) + \sum_{q^* \neq q, q'} \max \{D(A_q, A_{q^*}), D(A_{q'}, A_{q^*})\}.$$

W pracy, stanowiącej przedmiot wniosku, zaprezentowano cały szereg przykładów, poza tutaj przytoczonymi, konkretnej realizacji zarówno „dwustronnej” funkcji celu, jak i odpowiadających tym postaciom funkcji algorytmów agregacji. W szczególności – wskazano całą klasę algorytmów najmniejszej odległości, analogicznych do klasycznych procedur agregacji hierarchicznej. Dla tych ostatnich wskazano współczynniki znanego wzoru Lance’a-Williamsa (następnie rozszerzonego przez, np. Jambu), służącego do wyznaczania różnych formuł aktualizacji odległości między skupieniami, a więc, w istocie, samych definicji odległości.

### Zastosowania do innych zadań z zakresu analizy danych

W pracy przedstawiono wybrane przykłady możliwości zastosowania podejścia „dwustronnego” do formułowania i rozwiązywania zadań z dziedziny analizy danych, zasadniczo różnych od analizy skupień, z których większość, jednakże, może być sprowadzona do postaci zadania analogicznego do ogólnego zadania analizy skupień. I tak, przytoczono zadania kategoryzacji, wyznaczania optymalnego histogramu, podziału empirycznej skumulowanej funkcji rozkładu, blokowej diagonalizacji, itp. W niektórych przypadkach

pokazano tę analogię w szczegółach, w innych zaś ograniczono się tylko do jej zarysowania i pokazania możliwości sformułowania odpowiednich funkcji, na podstawie zasadniczych cech danego zadania.

Dość specyficznym przypadkiem takiego zastosowania poza obrębem analizy skupień jest opracowana całościowo „dwustronna” metodyka agregacji uporządkowań lub preferencji (w tym, w szczególności, wyrażonych dla par obiektów).

Wspomniane tutaj obszary zastosowań są wynikiem uogólnień, które stanowiły jeden z ostatnich etapów opracowywania podejścia „dwustronnego”. O ile w pierwszych pracach, jak to już wspomniano, koncentrowano się na jednej konkretnej postaci funkcji celu, w kolejnych pracach pojawiały się rozszerzenia, związane z innymi postaciami funkcji celu, jej postacią ogólną i własnościami tej postaci (publikacje Owsieński i Zadrożny, 1988, Owsieński, 1989, 1990a,b,c, oraz rozprawa doktorska z roku 1991). Począwszy od pracy Owsieński (1996), a zwłaszcza Owsieński (2011, 2012), uogólnienia objęły nie tylko kolejne konkretne formy funkcji celu, ale, po pierwsze, analogię i powiązanie z klasycznymi algorytmami agregacji hierarchicznej, a po drugie – wspomniane w niniejszym punkcie zastosowania poza analizą skupień (poza metodyka agregacji preferencji, która była przedmiotem badań od samego początku). Publikacja książkowa, stanowiąca przedmiot wniosku, podsumowuje wymienione tutaj wątki metodyczne.

### **Literatura – prace wnioskodawcy, dotyczące głównego osiągnięcia naukowego<sup>3</sup>**

- Owsieński J.W.: *Regionalization revisited: an explicit optimization approach*. CP-80-26. IIASA, Laxenburg 1980, str.24.
- Owsieński J.W.: Intuition vs. formalization: local and global criteria of grouping. *Control and Cybernetics*, 10, 1981, nr 1-2, 73-88.
- Owsieński J.W.: On a quasi-objective global clustering method. *Data Analysis and Informatics* 3, p.red.E.Didaya i in. North Holland, Amsterdam 1984a, 293-305.
- Owsieński J.W.: On aggregation in large scale models through a global clustering technique. W: *Large Scale Systems: Theory and Applications*, p. red. A. Straszaka. Pergamon Press for IFAC, Oxford, 1984b, 207-212.
- Owsieński J.W.: Optimization in clustering: an approach and other approaches. *Control and Cybernetics*, 18, 1986, nr 2, 107-114.
- Owsieński J.W., Zadrożny Sł.: Komputerowy system analizy skupień dla wspomagania procesów decyzyjnych. W: *Komputerowe systemy i metody wspomagające podejmowanie decyzji*, p. red. R. Kulikowskiego, IBS PAN, Warszawa, 1988, 199-210.
- Owsieński J.W.: On global optimality in cluster-wise regression. *Control and Cybernetics*, 18, 1989, nr 1, 53-67.
- Owsieński J.W.: Nowa metoda analizy skupień z globalną funkcją celu i jej algorytmy agregacji. W: *Taksonomia i jej zastosowania*. Materiały z konferencji naukowej zorganizowanej przez Akademię Ekonomiczną w Krakowie oraz Polskie Towarzystwo Statystyczne, Mogilany, 27-28 września 1989. AE Kraków, 1990a, 49-60
- Owsieński J.W.: Nowa agregacyjna metoda analizy skupień z globalną funkcją celu. W: *Badania systemowe*. Tom 1: *Podstawy teoretyczne i metody komputerowego modelowania systemów*, p. red. R. Kulikowskiego i A. Stachurskiego. Omnitech Press, Warszawa, 1990b, 234-255.

---

<sup>3</sup> Ta krótka lista nie zawiera prac, dotyczących zastosowań lub specyficznych sformułowań rozpatrywanego zadania, w pewnej mierze, przykładowo, zasygnalizowanych w dalszych punktach dokumentu.

- Owsiński J.W.: On a new naturally indexed quick clustering method with a global objective function. *Applied Stochastic Models and Data Analysis*, 6, 1990c, 157-171.
- Owsiński J.W.: Clustering, Distances and Knowledge from Data. W: *Ordinal and Symbolic Data Analysis*, p. red. E. Didaya, Y. Lechevalliera i O. Opitza. Springer, Berlin-Heidelberg, 1996, 277-287.
- Owsiński, J.W.: The bi-partial approach in clustering and ordering: the model and the algorithms. *Statistica & Applicazioni*, 2011, Special Issue, 43-59.
- Owsiński, J. W.: Clustering and ordering via the bi-partial approach: the rationale, the model and some algorithmic considerations. W: J. Pocięcha & Reinhold Decker, eds., *Data Analysis Methods and its Applications*. Wydawnictwo C.H. Beck, Warszawa, 2012, 109-124.

## 5. Omówienie pozostałych osiągnięć naukowo-badawczych

W okresie od uzyskania doktoratu, wnioskodawca opublikował 200 prac, poświęconych różnicowanej tematyce, w tym jednak przede wszystkim analizie skupień, omawianej tutaj metodyce, oraz ich zastosowaniom. W niniejszym punkcie bardzo skrótowo omówione zostaną niektóre, wybrane, z kierunków prac wnioskodawcy, wraz z przykładowymi publikacjami, które ukazały się po uzyskaniu doktoratu, a których zestawienie podano na końcu tego punktu.

### 5.1. Inne prace z dziedziny analizy skupień

Wśród prac o charakterze twórczym, oryginalnym, dotyczących analizy skupień, ale nie metodyki „dwustronnej” można wymienić, przykładowo – opracowanie podejścia hybrydowego, polegającego na połączeniu odpowiednio sparametryzowanych algorytmów k-means i agregacji hierarchicznej. Głównym celem tak zaprojektowanego podejścia była możliwość identyfikacji skupień o skomplikowanych kształtach (np. o kształtach krzywych zawartych wzajemnie w sobie), co faktycznie, w szeregu prób, okazało się możliwe (por. Owsiński i Mejza, 2007, 2008).

Innym istotnym nurtem – ostatnio prowadzonych – badań jest opracowanie i (liczne) zastosowania tzw. odwrotnego klasteringu, tj. poszukiwania procedury analizy skupień, która dla danego zbioru obiektów i danego ich podziału znajdzie podział możliwie najbliższy zadanemu (np. Owsiński i in., 2017a, b, Owsiński, Stańczak i Zadrożny, 2018). Procedura ta znajduje zastosowanie w wielu sytuacjach, różniących się między sobą znacznie co do, np., stopnia pewności zadanego podziału oraz jego związku z danymi o obiektach.

Należy wspomnieć także o pracach, dotyczących analizy skupień dla asymetrycznych odległości, tematu całkowicie, z dość oczywistych względów, zmarginalizowanego. Opracowano konkretne propozycje metodyczne i pokazano ich działanie na przykładach (Owsiński, 2009b, 2010a,b).

Osobny nurt prac związany był z podejściami, opartymi na zbiorach rozmytych i intuicjonistycznych. Nurt ten był realizowany przede wszystkim we współpracy ze zmarłym



w roku 2018 Dmirijem A. Viattcheninem (por. np. Kacprzyk, Viattchenin i Owsieński, 2014; Kacprzyk i in., 2015; Owsieński i in., 2018; Viattchenin, Owsieński, Kacprzyk, 2018).

Niezależnie od tego prowadzone były oryginalne prace, dotyczące wykrywania obiektów nietypowych (Owsieński, 2005, 2006b, Owsieński i Więclaw, 2007), czy możliwości wprowadzania opisów lingwistycznych skupień przez zastosowanie metody k-histogramów (Owsieński, 2004b; Owsieński, Pielak, 2008). Dodać to tego warto prace nad tzw. strukturami idealnymi, czyli warunkami o charakterze geometrycznym na zbiory obiektów, które to warunki mogłyby być wykorzystywane jako kryteria progowe do weryfikowania działania algorytmów analizy skupień (Owsieński, 2004a, 2006a).

Cały szereg prac był ponadto poświęcony zastosowaniom analizy skupień, na przykład w takich zagadnieniach, związanych z dziedziną badań operacyjnych, jak elastyczne systemy produkcji (np. Owsieński, 2008a, 2009a) czy zagadnienie lokalizacji (Owsieński, 2014a).

## 5.2. Inne prace z zakresu analizy danych

Przytoczymy tutaj przykłady prac z szerszej dziedziny analizy danych, z pominięciem samej analizy skupień. Zacząć należy od prac, już marginesowo wspomnianych, poświęconych agregacji preferencji (ogólniej: uporządkowań) i wspomaganii – z tego punktu widzenia – podejmowania decyzji, zwłaszcza grupowych (np. Brüggemann, Fattore, Owsieński, 2011; Owsieński, 2003a,b; 2017a,b). Prace te dotyczyły zarówno podejścia „dwustronnego” do samej agregacji opinii, jak i do jednoczesnego grupowania i agregacji opinii, a w tym i możliwości osiągnięcia konsensusu.

Nieco podobny charakter miały prace, poświęcone identyfikacji modeli z równoczesnym podziałem zbioru obserwacji na podzbiory (skupienia), dla których identyfikowane są osobne modele. Na tle szeregu istniejących metodyk, przeznaczonych do tego celu, zajmowano się głównie ogólnymi cechami takich metodyk, które powinny spełniać warunki podejścia „dwustronnego” (Owsieński, 2002a,b).

W ramach prowadzonych w Wyższej Szkole Informatyki Stosowanej i Zarządzania w Warszawie została także opracowana książka o charakterze podręcznikowym z analizy danych (Owsieński, 2003c).

## 5.3. Przykłady prac badawczych z innych dziedzin

Dość istotnym obszarem zainteresowań i prac badawczych wnioskodawcy są metody i systemy wyszukiwania informacji tekstowej. Także i w tej dziedzinie powstał tekst podręcznikowy do wykładów (Owsieński, 2014b), ale, niezależnie od tego, szereg prac przyczynkowych, np. Kowalska, Owsieński (2014), Ryczaj, Owsieński (2015), w znacznej mierze, choć nie wyłącznie, w ramach zajęć w Wyższej Szkole Informatyki Stosowanej i Zarządzania w Warszawie.

W ciągu ostatnich około 10 lat, wnioskodawca jest zaangażowany we współpracę z grupą osób, zajmujących się zastosowaniami teorii grafów w dziedzinie transportu i logistyki oraz nieco szerzej – w ogóle zagadnieniami transportowo-logistycznymi (np. Maźbic-Kulma i in., 2014, 2015; Stańczak i in., 2017).

Już od wczesnych lat 80-tych XX wieku wnioskodawca prowadził prace nad modelami gospodarczymi rolnictwa, zwłaszcza rolnictwa (regionalnego). Do zbliżonej tematyki wnioskodawca jeszcze wielokrotnie wracał, także w odniesieniu do skali kraju. Przykładami publikacji w tym obszarze z okresu po roku 2000 mogą być: tom Nahorski, Owsiniński, Szapiro (2007), w którym zamieszczono, m.in., oryginalne tekst późniejszego laureata nagrody Nobla, Leonida Hurwicza, czy raporty Gadomski, Owsiniński (2009), oraz Czapiewski, Janc, Owsiniński (2016).

Prowadzone były, w powiązaniu z badaniami z dziedziny analizy danych (analiza skupień, agregacji uporządkowani itp.), prace, dotyczące oceny jakości życia, zwłaszcza w aspekcie przestrzennym (por. Owsiniński, Tarchalski, 2008; Owsiniński, 2008b, 2009c). Prace te były także związane z szerszymi zagadnieniami, odnoszącymi się do analizy terytorialnej i przestrzennej, zwłaszcza na poziomie gmin (np. Owsiniński, Andrzejewski, 2010; Owsiniński, Milczewski, 2010), zarówno dla całego kraju, jak i dla poszczególnych województw, w tym mazowieckiego. W odniesieniu do tego ostatniego prowadzono przez kilka lat badania jakości i roli społecznej (sieciowej) stron internetowych gminnych i powiatowych (Pielak, Owsiniński, 2010; Owsiniński i in., 2010; Owsiniński, Pielak, 2011).

Innymi dziedzinami, których dotyczą publikacje, głównie samodzielne, jakie ukazały się po uzyskaniu doktoratu przez wnioskodawcę, są, m.in., pojęcie informacji i jego związek z takimi pojęciami jak wiedza czy mądrość, rozwój zrównoważony, struktury handlu międzynarodowego, szara strefa w gospodarce, czy struktury instytucjonalne społeczeństw.

#### 5.4. Przykładowe publikacje dotyczące pozostałych osiągnięć naukowo-badawczych<sup>4</sup>

- Brüggemann R., Fattore M., Owsiniński J. W.: Using poset theory to compare fuzzy multidimensional material deprivation across regions. W: Salvatore Ingrassia, Roberto Rocci, Maurizio Vichi, eds., *New Perspectives in Statistical Modeling and Data Analysis. Series: Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, 2011. ISBN: 978-3-642-11362-8, 49-56.
- Czapiewski K., Janc K., Owsiniński J. W., Śleszyński P.: A modelling Project in Poland: the social and intellectual capital aspects. *Studies in Agricultural Economics*, 118 (1), 2016, 5-15, <http://dx.doi.org/10.7896/j.1602>
- Gadomski J., Owsiniński J. W.: Model rolnictwa polskiego MODROL do analizy skutków polityki rolnej dla dochodowości gospodarstw rolnych. Koncepcja całości oraz modele wybranych rynków. *Ekonomiczne i Społeczne Uwarunkowania Rozwoju Polskiej Gospodarki Żywnościowej po Wstąpieniu Polski do Unii Europejskiej*, nr 137, IERiGŻ-PIB, Warszawa 2009, ISBN 978-83-7658-033-3, 98 str.
- Kacprzyk J., Viattchenin D. A., Owsiniński J. W.: Extracting fuzzy classification rules from three-way data. *JAMRIS*, 2014, 2, 47-57. DOI 10.14313/JAMRIS\_2-2014/19
- Kacprzyk J., Owsiniński J. W., Viattchenin D. A., Shyrai S.: A New Heuristic Algorithm of Possibilistic Clustering Based on Intuitionistic Fuzzy Relations. W: K. T. Atanassov et al., eds.: *Novel Developments in Uncertainty Representation and Processing. Advances in Intuitionistic Fuzzy Sets and Generalized Nets – Proc. of 14th International on Intuitionistic Fuzzy Sets and Generalized Nets. Advances in Intelligent Systems and Computing* **401**, Springer, 2015, 199-214.
- Kowalska M., Owsiniński J. W.: Wspomaganie lokalizacji językowej dokumentów. *Zeszyty Naukowe Wydziału Informatyki WSISiZ „Informatyka Stosowana”*. Nr 1/2014, 29-61.
- Mazbicz-Kulma B., Owsiniński J. W., Sęp K., Stańczak J.: The Kernel and Shell Structures as a Tool for Improving the Graph of Transportation Connections. *Operations Research and Decisions*, 2013, No. 2/3, vol. 23 (2014), 91-105, DOI: 10.5277/ord130208

---

<sup>4</sup> Wymienione w niniejszym dokumencie publikacje – w sumie 61 pozycji, stanowią jedynie niewielką część całości dorobku publikacyjnego wnioskodawcy od czasu uzyskania doktoratu (blisko 200 pozycji).

- Mazbicz-Kulma B., Owskiński J. W., Stańczak J., Barski A., Sęp, K.: Shaping Urban Transport System: Convenience, environment and urban shape. W: Nguyen Xuan Thinh, red.: *Modelling and Simulation Ecosystems – Workshop Kölpinsee 2014*. Rhombos Verlag, Berlin 2015, 71-82.
- Nahorski Z., Owskiński J.W., Szapiro T., red., *The Socio-Economic Transformation. Getting Closer to What?* Palgrave Macmillan, Houndmills – New York, 2007, 146+xvii str., ISBN 978-0-230-00794-9
- Owskiński J.W.: Cluster-Wise Modelling: Issues, Proposals, Methods. W: *Taksonomia 9. Klasyfikacja i analiza danych – teoria i zastosowania*, p. red. K. Jajugi i M. Walesiaka. *Prace Naukowe AE we Wrocławiu*, nr 942. Wydawnictwo AE we Wrocławiu, Wrocław 2002a, 400-410.
- Owskiński J.W.: Cluster-Wise Model Identification: What Happened During the Last Twenty Years? W: Jakub Gutenbaum, red., *Automatyka, Sterowanie, Zarządzanie. Książka jubileuszowa z okazji 70-lecia urodzin Profesora Kazimierza Mańczaka*. IBS PAN, Warszawa 2002b, 325-338.
- Owskiński J.W.: Opinion structure and agreement conditions in group decision: the cluster analysis perspective. W: Janusz Kacprzyk, Dariusz Wagner, Eds., *Group Decisions and Voting*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2003a, 195-204.
- Owskiński J.W.: Group choice: opinion structure, consensus, and cluster analysis. W: *Taksonomia 10. Klasyfikacja i analiza danych - teoria i zastosowania*, p. red. K.Jajugi i M.Walesiaka. *Prace Naukowe AE we Wrocławiu*, nr 988, Wyd. AE im. Oskara Langego we Wrocławiu, Wrocław 2003b, 332-342.
- Owskiński J.W.: *Wykład z metod analizy danych*. WSISiZ, Warszawa, 2003c, 160 str., ISBN 83-88311-62-X
- Owskiński J.W.: Group opinion structure: The ideal structures, their relevance and effective use. W: D. Baier and K.-D. Wernecke, red.: *Innovations in Classification, Data Science, and Information Systems. Proc. 27<sup>th</sup> Annual GfKI Conference, University of Cottbus, March 12-14, 2003*. Springer-Verlag, Heidelberg-Berlin, 2004a, 471-481.
- Owskiński J.W. Distances, effective clustering and k-histograms: a natural way to produce linguistically meaningful clusters? W: P. Grzegorzewski, M. Krawczak, Sł. Zadrozny, red., *Soft Computing. Tools, Techniques and Applications*. Oficyna Wydawnicza EXIT, Warszawa 2004b, 219-227.
- Owskiński J.W.: What is an outlier? Some questions and some responses. W: *Issues in the Representation and Processing of Uncertain and Imprecision Information: Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets, and Related Topics*, K.T. Atanassov, J. Kacprzyk, M. Krawczak, E. Szmidt, eds., Warsaw, EXIT, 2005, 278-289.
- Owskiński J.W.: The ideal structures in group opinion analysis. W: J. Kacprzyk, R. Budziński, red., *Badania Operacyjne i Systemowe 2006. Metody i techniki*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2006a, 207-217.
- Owskiński J.W.: Outlier detection: notions, problems, and methodological proposals. W: *Taksonomia 13. Klasyfikacja i analiza danych - teoria i zastosowania*, p. red. K.Jajugi i M.Walesiaka. *Prace Naukowe AE we Wrocławiu*, nr 1126, Wyd. AE im. Oskara Langego we Wrocławiu, Wrocław 2006b, 45-55.
- Owskiński J.W.: Machine-part grouping and cluster analysis. W: C.E. Pereira, O. Zaikin, Z. Banaszak, eds., *9th IFAC Workshop on Intelligent Manufacturing Systems (IMS'08). Preprints*. IFAC, 2008a, 399-404.
- Owskiński J.W.: How can we read out life quality assessments from the official socio-economic data on Polish municipalities? W: W.Ostasiewicz, ed., *Quality of Life Improvement Through Social Cohesion. Proc. of the 4<sup>th</sup> International Conference*. Wrocław University of Economics, Department of Statistics, 2008b, 140-155.
- Owskiński J. W.: Machine-part grouping and cluster analysis: similarities, distances and grouping criteria. *Bulletin of the Polish Academy of Sciences. Technical Sciences*, Special Issue: “*Modeling and optimization of manufacturing systems*”, Guest Editors: Z. A. Banaszak, J. Józefczyk, **57**, 3, September 2009a, 217-228.
- Owskiński J.W.: Asymmetric distances – a natural case for fuzzy clustering? W: D. A. Viattchenin, red., *Developments in Fuzzy Clustering*. Vever, Minsk (Belarus’), 2009b; ISBN 978-985-6215-77-6; 36-45.
- Owskiński J. W.: How can we read out life quality assessments from the official socio-economic data on Polish municipalities? W: W. Ostasiewicz, red., *Quality of Life Improvement through Social Cohesion. Statystyka i Ryzyko. Research Papers of Wrocław University of Economics*, No. 73. Wrocław 2009c, 50-63.
- Owskiński J. W.: Asymmetric distances: potential output structures and procedures. *Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą*, 31. Numer pt. *Dane – analiza – modelowanie – optymalizacja – podejmowanie decyzji*, 2010a, 317-325.

- Owsiński, J.W.: Asymmetric distances and fuzzy grouping. W: K. T. Atanassov et al., eds., *Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Vol. II: Applications*. SRI PAS / IBS PAN, Warsaw 2010b, 207-215, ISBN-13 9788389475305
- Owsiński J. W.: Bi-partial version of the p-median / p-center facility location problem and some algorithmic considerations. *JAMRIS*, 2014a, 3, 59-63. DOI 10.14313/JAMRIS\_3-2014/28
- Owsiński J. W.: *Wprowadzenie do wyszukiwania informacji tekstowych: modele, techniki, zasadnicze zagadnienia*. WSISiZ, Warszawa, 2014b, 150 str.,
- Owsiński J. W.: Endowing posets with flesh: if, why and how? W: R. Brüggemann, M. Fattore, eds., *Partial Order Concepts in Applied Science*. Springer, 2017a, 3-20, ISBN 978-3-319-45419-1
- Owsiński J. W.: Is There Any 'Law of Requisite Variety' in Construction of Indices for Complex Systems? *Social Indicators Research*, 2017b, **136**(3), 1125-1137, DOI: 10.1007/s11205-016-1545-5
- Owsiński J. W., Andrzejewski M.: Rural municipalities in Poland: Development types, paths and perspectives. Preliminary results – the spatial aspect. W: A. Fieldsend, red., *Linking Competitiveness with Equity and Sustainability: New Ideas for the Socio-Economic Development of Rural Areas*. W serii: *Rural Areas and Development*, vol. 7, ERDN, Warsaw, 2010, 177-196. ISBN 978-83-7658-196-5.
- Owsiński, J. W., Kacprzyk, J., Opara, K., Stańczak, J., Zadrozny, Sł.: Using a reverse engineering type paradigm in clustering: An evolutionary programming based approach. W: V. Torra, A. Dalbom and Y. Narukawa, eds., *Fuzzy Sets, Rough Sets, Multisets and Clustering. Dedicated to Prof. Sadaaki Miyamoto. Studies in Computational Intelligence* **671**, Springer, 2017a, ISBN 978-3-319-47556-1.
- Owsiński J. W., Kacprzyk J., Shyrai St., Szmidt E., Viattchenin D. A., Hernandez Hormazabal J.: A heuristic algorithm of possibilistic clustering with partial supervision for classification of the intuitionistic fuzzy data. *Journal of Multiple-Valued Logic & Soft Computing*, 2018, **31** (4).
- Owsiński J. W., Mejza M. T.: On a New Hybrid Clustering Method for General Purpose Use and Pattern Recognition. W: *Proceedings of the International Multiconference on Computer Science and Information Technology*, vol. 2, ISSN 1896-7094, <http://www.papers2007.imcsit.org/> 121-126, 2007.
- Owsiński J.W., Mejza M.T.: A new hybrid clustering method: from “subassemblies” to “shapes”. W: O. Hryniewicz, A. Straszak, J. Studziński, red., *Badania operacyjne i systemowe: środowisko naturalne, przestrzeń, optymalizacja. Badania Systemowe*, **63**. IBS PAN, Warszawa, 2008, 341-348.
- Owsiński J.W., Milczewski M.: Rekursja w problemie regionalizacji. W: Owsiński, J. W., red., *Analiza systemów przestrzennych. Wybrane zagadnienia. Badania Systemowe*, tom 67. Instytut Badań Systemowych PAN, Warszawa 2010, 47-58.
- Owsiński J. W., Opara K., Stańczak J., Kacprzyk J., Zadrozny S.: Reverse Clustering. An Outline for a Concept and Its Use. *Toxicological & Environmental Chemistry*, 2017b, DOI: 10.1080/02772248.2017.1333614
- Owsiński J. W., Pielak A. M.: Qualitative assessment of the websites of local authorities in Poland with hierarchical k-histograms. W: *Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Applications. Vol. II*, K. Atanassov et al., eds. Academic Publishing House EXIT, Warsaw 2008, 185-192.
- Owsiński J. W., Pielak A. M., Sęp K., Ponichtera R.: TIROLS: local web-based networks centred on self-governmental websites vs. local activity – a project outline. *Studies and Proceedings – Polish Association for Knowledge Management*, 34. Issue edited by J. W. Owsiński, 2010, 20-36.
- Owsiński J.W., Pielak A. M.: Local Authority Websites in Rural Areas: Measuring Quality and Functionality and Assessing the Role. W: Z. Andreopoulou, B. Manos, N. Polman, D. Viaggi, eds., *Agricultural and Environmental Informatics, Governance and Management: Emerging Research Applications*. Information Science Reference / IGI Global, Hershey, PA, 2011, 39-60.
- Owsiński J. W., Stańczak J., Zadrozny Sł.: Designing the Municipality Typology for Planning Purposes: The Use of Reverse Clustering and Evolutionary Algorithms. W: P. Daniele and L. Scrimali, eds., *New Trends in Emerging Complex Real Life Problems. ODS, Taormina, Italy, September 10-13, 2018*. AIRO Springer Series, vol. 1. Springer, Cham, 2018. ISBN/EAN 9783030004736 [https://doi.org/10.1007/978-3-030-00473-6\\_33](https://doi.org/10.1007/978-3-030-00473-6_33)
- Owsiński J. W., Tarchalski T.: Pomiar jakości życia. Uwagi na marginesie pewnego rankingu. *Zeszyty Naukowe Wydziału Informatycznych Technik Zarządzania „Współczesne Problemy Zarządzania”*, 2008, nr 1, 59-95.

- Owsiński J. W., Więclaw A.: Finding the Odd-Man-Out In Development Scoring and Classifications. W: O. Hryniewicz, J. Studziński, A. Szediw, eds., *Environmental Informatics and Systems Research*. Volume 2: Workshop and application papers. Shaker Verlag, Aachen, 2007, 195-199.
- Pielak A. M., Owsiński J. W.: Jakość i dynamika rozwoju stron internetowych samorządów lokalnych Mazowsza a przestrzeń wiejska. W: Owsiński, J. W., red., *Analiza systemów przestrzennych. Wybrane zagadnienia. Badania Systemowe*, tom 67. Instytut Badań Systemowych PAN, Warszawa 2010, 147-173.
- Ryczaj W., Owsiński J.W.: Im więcej, tym lepiej? O pewnej analizie z dziedziny wyszukiwania informacji tekstowej. *Zeszyty naukowe WITZ WSISiZ "Współczesne Problemy Zarządzania"*, 1/2015, 81-124.
- Stańczak J., Barski A., Sęp K., Owsiński J. W.: The problem of distribution of Park and Ride car parks in Warsaw. *International Journal of Information and Management Science*, 2017, **27** (2) 179-190.
- Viattchenin D. A., Owsiński J. W., Kacprzyk J.: New developments in fuzzy clustering with emphasis on special types of tasks. *Control & Cybernetics*, 2018, **47** (2), 115-130.



Jan W. Owsiński