

Małgorzata Machowska-Szewczyk

Politechnika Szczecińska

PORÓWNANIE WYNIKÓW KLASYFIKACJI NAUCZYCIELI AKADEMICKICH POD WZGLĘDEM WPLYWU OSOBOWOŚCI NA POPULARNOŚĆ WYKŁADÓW ZE STATYSTYKI

1. Wstęp

W wykładach ze statystyki matematycznej na wyższych uczelniach najchętniej uczestniczą osoby o wrodzonej zdolności do abstrakcyjnego myślenia, gdyż nie mają one większych kłopotów ze stosowaniem pojęć statystycznych. Na chęć uczestnictwa w wykładach, oprócz predyspozycji słuchaczy, niewątpliwy wpływ mają również zdolności dydaktyczne osoby prowadzącej przedmiot. W celu zbadania wpływu osobowości nauczyciela na popularność wykładów ze statystyki wykorzystano elementy teorii zbiorów przybliżonych oraz przeprowadzono klasyfikację obiektów charakteryzowanych przez zmienne symboliczne [Machowska-Szewczyk 2006]. Badanie przeprowadzono na podstawie ankiet wypełnianych przez studentów kilku szczecińskich uczelni. W artykule porównano efekty klasyfikacji nauczycieli akademickich otrzymane w wyniku wykorzystania zmiennych symbolicznych z wagami z rezultatami klasyfikacji opartej na elementach zbiorów przybliżonych.

2. Źródło danych

Obiektami podlegającymi klasyfikacji byli nauczyciele akademicy z różnych szczecińskich uczelni, którzy prowadzili wykłady ze statystyki matematycznej i wyrazili zgodę na przeprowadzenie badania ankietowego z ich udziałem. Źródło danych w badaniu stanowiłyankiety wypełniane przez studentów. Na podstawie ankiet przyjęto następujące czynniki wpływające na frekwencję na wykładach [Machowska-Szewczyk 2006; Machowska-Szewczyk, Sompolska-Rzechuła 2007]:
 X_1 – stopień trudności zajęć (zbyt wysoki, wysoki, umiarkowany, raczej niski, niski);

- X_2 – czy uprzednio prowadzone zajęcia stanowią odpowiednią bazę dla ocenianego przedmiotu (tak, nie, nie mam zdania);
- X_3 – wymagania (zbyt wysokie, wysokie, właściwe, przeciętne, niewielkie);
- X_4 – czy zajęcia były przydatne do opanowania materiału (tak, przy dodatkowej pracy, niewystarczające);
- X_5 – czy zastosowana forma zaliczeń, egzaminów pozwoliła na rzetelną i sprawiedliwą ocenę opanowania materiału (tak, raczej tak, nie);
- X_6 – proszę ocenić swoją aktywność na zajęciach (bardzo wysoka, średnia, niewielka, żadna);
- X_7 – czy zajęcia przebiegały prawidłowo pod względem organizacyjnym (tak, raczej tak, nie);
- X_8 – ocena przygotowania prowadzącego do zajęć (zawsze odpowiednie, zawsze przeciętne, nie zawsze odpowiednie, zazwyczaj złe);
- X_9 – ocena komunikatywności prowadzącego zajęcia (wybitna, bardzo dobra, bez zastrzeżeń, czasami niewystarczająca, słaba);
- X_{10} – ocena zawartości informacyjnej zajęć (bardzo dobra, wystarczająca, słaba);
- X_{11} – ocena umiejętności organizowania pracy w grupie na ćwiczeniach (wysoka, wystarczająco dobra, słaba, niewystarczająca);
- X_{12} – kultura osobista prowadzącego zajęcia (wysoka, bez zastrzeżeń, często wątpliwa);
- X_{13} – możliwość kontaktu z prowadzącym zajęcia w przypadku chęci uzyskania konsultacji poza ustalonymi godzinami (codziennie, w określonym czasie, brak możliwości, nie mam zdania);
- X_{14} – stosunek prowadzącego do studentów (właściwy, raczej właściwy, zbyt poufały, raczej niewłaściwy, zły);

Przyjęto jeszcze jedną zmienną X_{15} – zdolności dydaktyczne nauczyciela, o których informacja nie jest znana bezpośrednio na podstawie ankiet. Jest to czynnik ludzki, oceniany na podstawie badania, który również istotnie wpływa na chęć uczestnictwa studentów w wykładach. Przyjęto również zmienną Y – uczestnictwo w wykładach (75-100%, 50-75%, 25-50%, 0-25%).

3. Klasyfikacja nauczycieli

W niniejszym badaniu nauczyciel jest obiektem symbolicznym. Obiekt symboliczny może być opisany przez zmienne symboliczne z wagami. Są to zmienne, które przyjmują różne warianty, a poszczególnym wariantom przypisane są wagi, prawdopodobieństwa lub częstości występowania. Zmienna symboliczna $X_i, i \in \{1, 2, \dots, K\}$ jest odwzorowaniem przyporządkowującym poszczególnym obiektom kolejne warianty zmiennej z ustalonymi wagami, czyli $X_i : O_k \mapsto ((x_{i1}, p_k(x_{i1})), \dots, (x_{im_i}, p_k(x_{im_i})))$, przy czym $\sum_{j=1}^{m_i} p_k(x_{ij}) = 1$.

Do oceny podobieństwa między obiektami O_k, O_l względem zmiennej symbolicznej X_i , można wykorzystać adaptacje miar sformułowanych dla rozkładów prawdopodobieństwa [Wilk 2006, s. 227-228]. Najczęściej wykorzystuje się odległość Minkowskiego, spełniającą warunki metryki i zdefiniowaną następująco:

$$d_p^i(O_k, O_l) = \left(\sum_{j=1}^{m_i} |p_k(x_{ij}) - p_l(x_{ij})|^p \right)^{\frac{1}{p}}, \text{ gdzie } p \geq 1.$$

Tabela 1. Zmienne symboliczne dla obiektu O_1

Nr zmiennej	Zmienna	Warianty zmiennej z udziałem procentowym	Typ wariantów
X_1	Stopień trudności zajęć	zbyt wysoki 4,3%; wysoki 82,6%, umiarkowany 13,1%, raczej niski 0%, niski 0%	lista kategorii porządkowych
X_2	Uprzednio prowadzone zajęcia stanowią odpowiednią bazę dla ocenianego przedmiotu	tak 60,9%; nie 13%; nie mam zdania 26,1%	lista kategorii nominalnych
X_3	Wymagania stawiane studentom	zbyt wysokie 8,7%; wysokie 60,9%; właściwe 30,4%; przeciętne 0%; niewielkie 0%	lista kategorii porządkowych
X_4	Przydatność zajęć do opanowania materiału określonego w programie	tak 26,1%; przy dodatkowej pracy 69,6%; niewystarczająco 4,3%	lista kategorii porządkowych
X_5	Czy forma zaliczeń, egzaminów pozwoliła na rzetelną i sprawiedliwą ocenę opanowania materiału	tak 21,7%; raczej tak 21,7%; nie 56,5%	lista kategorii porządkowych
X_6	Aktywność studentów na zajęciach	bardzo wysoka 4,3%; średnia 52,2%; niewielka 39,2%; żadna 4,3%	lista kategorii porządkowych
X_7	Pravidłowość przebiegu zajęć pod względem organizacyjnym	tak 73,9%; raczej tak 21,7%; nie 4,4%	lista kategorii porządkowych
X_8	Ocena przygotowania prowadzącego do zajęć	zawsze odpowiednie 78,3%; zawsze przeciętne 21,7%; nie zawsze odpowiednie 0%; zazwyczaj złe 0%	lista kategorii porządkowych
X_9	Ocena komunikatywności prowadzącego zajęcia	wybitna 4,3%; bardzo dobra 39,1%; bez zastrzeżeń 21,7%; czasami niewystarczająca 26,1%; słaba 8,8%	lista kategorii porządkowych
X_{10}	Ocena zawartości informacyjnej zajęć	bardzo dobra 30,4%; wystarczająca 52,2%; słaba 17,4%	lista kategorii porządkowych
X_{11}	Ocenić umiejętność organizowania pracy w grupie na ćwiczeniach	wysoka 8,7%; wystarczająco dobra 52,2%; słaba 21,7%; niewystarczająca 17,4%	lista kategorii porządkowych
X_{12}	Kultura osobista prowadzącego zajęcia	wysoka 60,9%; bez zastrzeżeń 34,8%; często wątpliwa 4,3%	lista kategorii porządkowych
X_{13}	Możliwość kontaktu z prowadzącym zajęcia w celu konsultacji poza ustalonymi godzinami	codziennie 4,4%; w określonym czasie 47,8%; brak możliwości 0%; nie mam zdania 47,8%	lista kategorii nominalnych
X_{14}	Stosunek prowadzącego do studentów	właściwy 43,5%; raczej właściwy 43,5%; zbyt poufały 8,7%; raczej niewłaściwy 4,3%; zły 0%	lista kategorii porządkowych

Źródło: [Machowska-Szewczyk 2006, s. 286].

Odległość dowolnych dwóch obiektów O_k, O_l , opisanych za pomocą zmiennych symbolicznych z wagami, może być mierzona jako zagregowana wartość odległości względem kolejnych zmiennych, np. odległość Minkowskiego [Wilk 2006, s. 230]:

$$d(O_k, O_l) = \left[\sum_{i=1}^K (\omega_i \cdot d^i(O_k, O_l))^p \right]^{\frac{1}{p}}$$
, gdzie $\omega_i \in [0,1]$ jest wagą odległości obiektów względem zmiennej X_i , $i \in \{1, \dots, K\}$, przy czym $\sum_{i=1}^K \omega_i = 1$.

Ze względu na to, że poszczególnym nauczycielom przypisanych jest wiele wariantów odpowiedzi na kolejne pytania zawarte w ankiecie, wygodnie jest przyjąć do opisu obiektów zmienne symboliczne z wagami. Pojedynczemu obiektowi przypisano ciąg zmiennych symbolicznych z wagami oznaczającymi procentowy udział osób wybierających dany wariant. W przestrzeni zmiennych, przyjętych jako czynniki wpływające na chęć uczestnictwa w wykładach (bez czynnika X_{15}), dokonano klasyfikacji obiektów symbolicznych. Ocenę wpływu osobowości nauczyciela na popularność jego wykładów przeprowadzono, porównując rzeczywistą przeciętną frekwencję na wykładach danej osoby ze średnią frekwencją w klasie obiektów podobnych. Jeżeli rzeczywista frekwencja na wykładach danej osoby wyraźnie przekracza średnią frekwencję w grupie nauczycieli o podobnych wartościach zmiennych symbolicznych, to można wnioskować, że jest to pozytywny wpływ czynnika ludzkiego. Jeżeli wykłady danej osoby cieszą się niższą popularnością od przeciętnej w klasie obiektów podobnych, to osobowość nauczyciela zniechęca studentów do uczestniczenia w wykładach.

Tabela 2. Ocena predyspozycji nauczyciela akademickiego do pracy wykładowcy

Obiekt	Średnia frekwencja rzeczywista	Średnia frekwencja w grupie	Różnica	Ocena
O_1	60,3	71,3	-11,0	N
O_2	53,4	68,6	-15,2	N
O_3	85,5	68,6	16,9	W
O_4	64,9	68,6	-3,7	S
O_5	83,2	71,3	11,9	W
O_6	75,9	68,6	7,3	W
O_7	75,0	71,2	3,8	S
O_8	70,5	71,3	-0,8	S
O_9	60,4	68,6	-8,2	N
O_{10}	80,5	68,6	11,9	W
O_{11}	70,9	70,9	0,0	*
O_{12}	73,5	71,2	2,3	S
O_{13}	69,9	71,2	-1,3	S
O_{14}	75,0	71,2	3,8	S
O_{15}	65,6	71,2	-5,6	N
O_{16}	48,4	52,3	-3,9	S
O_{17}	89,3	71,2	18,1	W
O_{18}	53,9	52,3	1,6	S

Źródło: obliczenia własne [Machowska-Szewczyk 2006].

Mając dla każdego obiektu informacje o zmiennych symbolicznych, analogiczne jak w tab. 1, wyznaczono macierz odległości Minkowskiego ze współczynnikiem $p = 2$ między poszczególnymi obiektami. Następnie wyznaczono liczbę klas jako liczbę znalezionych punktów skupienia [Jajuga 1984, s. 27-28] i przeprowadzono klasyfikację nauczycieli akademickich za pomocą hierarchicznej metody aglomeracyjnej Warda [Nowak 1990, s. 80-82]. W wyniku aglomeracji otrzymano pięć grup nauczycieli o zbliżonych wartościach zmiennych symbolicznych:

klasa 1: $O_1, O_5, O_8,$

klasa 4: $O_{11},$

klasa 2: $O_2, O_3, O_4, O_6, O_9, O_{10},$

klasa 5: $O_{16}, O_{18}.$

klasa 3: $O_7, O_{12}, O_{13}, O_{14}, O_{15}, O_{17},$

Do oceny wpływu osobowości nauczyciela na chęć uczestnictwa studentów w wykładach wykorzystano różnicę między frekwencją rzeczywistą (średnią ważoną wyznaczoną na podstawie ankiet) a średnią frekwencją w grupie, do której przydzielono obiekt. Wyniki tej oceny przedstawiono w tab. 2.

Interpretacja ocen predyspozycji nauczyciela do pracy w charakterze wykładowcy jest następująca: W – ocena wysoka (>5), S – średnia [$-5; 5$], N – niska (<-5), * – brak oceny ze względu na jednoelementowy zbiór obiektów podobnych.

4. Klasyfikacja nauczycieli za pomocą elementów teorii zbiorów przybliżonych

W pracy [Machowska-Szewczyk, Sompolska-Rzechuła 2007] przeprowadzono aproksymację przybliżonego modelu frekwencji w niepełnej przestrzeni zmiennych (bez czynnika X_{15}). Do opisu systemu informacyjnego przyjęto jako atrybuty warunkowe zmienne X_1 – X_{14} oraz atrybut decyzyjny Y . W celu eliminacji zbędnych danych w kolejnym etapie wyznaczono redukty względne ze względu na atrybut decyzyjny. Dla wszystkich reduktów obliczono jakość przybliżenia rodziny conceptów decyzyjnych oraz znormalizowany współczynnik istotności [Mrózek, Płonka 1999, s. 15, 24]. Na podstawie atrybutów $X_1, X_2, X_3, X_{12}, X_{13}$ określono reguły decyzyjne przybliżonego modelu uczestnictwa studentów w wykładach ze statystyki matematycznej. Zbadano siłę, wsparcie oraz pewność każdej reguły. W wyniku badania otrzymano reguły z konkluzjami w postaci symboli oznaczających numer przedziału, do którego należy frekwencja studentów na wykładach poszczególnych osób prowadzących (0-25%, 25-50%, 50-75%, 75-100%). Znając jednak obiekty (nauczycieli), które wspierają, daną regułę, można wyznaczyć średnią ważoną Y_{mod} i otrzymać wartość liczbową oznaczającą przeciętną frekwencję na wykładach danej osoby. W tab. 3 przedstawiono ocenę „zdolności dydaktycznych” osoby prowadzącej przedmiot za pomocą różnicy między frekwencją rzeczywistą a frekwencją wynikającą z przybliżonego modelowania.

Tabela 3. Ocena predyspozycji nauczyciela akademickiego do pracy wykładowcy według metody modelowania przybliżonego

Obiekty	Y	Y_{mod}	Różnica	Ocena
O_1	60,3	70,4	-10,1	N
O_2	53,4	53,4	0	*
O_3	85,5	76,6	8,9	W
O_4	64,9	69,1	-4,2	S
O_5	83,3	77,3	4,9	S
O_6	75,9	74,1	1,8	S
O_7	75,0	74,9	0,1	S
O_8	70,5	74,1	-3,6	S
O_9	60,4	73,6	-13,2	N
O_{10}	80,5	70,4	10,1	W
O_{11}	70,9	69,1	1,8	S
O_{12}	73,5	77,3	-4,9	S
O_{13}	69,9	69,1	0,8	S
O_{14}	75,0	74,1	0,9	S
O_{15}	65,6	69,1	-3,5	S
O_{16}	48,4	48,4	0	*
O_{17}	89,3	69,1	20,2	W
O_{18}	53,9	69,1	-15,2	N

Źródło: [Machowska-Szewczyk, Sompolska-Rzechuła 2007].

5. Porównanie wyników klasyfikacji

Na podstawie przeprowadzonych badań można zauważyć, że 11 obiektów zostało ocenionych jednakowo za pomocą metody klasyfikacji obiektów symbolicznych oraz metody wykorzystującej elementy teorii zbiorów przybliżonych. Wyniki zgodne ocen zdolności dydaktycznych są następujące:

- wybitnie uzdolnione (W): O_3, O_{10}, O_{17} ,
- przeciętnie uzdolnione (S): $O_4, O_7, O_8, O_{12}, O_{13}, O_{14}$,
- słabo uzdolnione (N): O_1, O_9 .

Tabela 4. Różnice w ocenach obiektów za pomocą klasyfikacji symbolicznej oraz modelowania przybliżonego

Obiekty	Ocena	
	zbiory symboliczne	zbiory przybliżone
O_2	N	*
O_5	W	S
O_6	W	S
O_{11}	*	S
O_{15}	N	S
O_{16}	S	*
O_{18}	S	N

Źródło: opracowanie własne.

Niektóre osoby zostały ocenione za pomocą każdej z metod inaczej, co wynika z tego, że przyjęcie do opisu obiektów zmiennych symbolicznych pozwala uwzględnić odpowiedzi wszystkich studentów poddanych ankietyzacji, a nie tylko odpowiedzi dominujące. Powoduje to mniejszą stratę informacji. Zatem wyniki klasyfikacji (a także ocena) mogą się różnić od wyników uzyskanych na podstawie badania, w którym uwzględniono jedynie wypowiedzi większości studentów. Różnice te przedstawiono w tab. 4.

Aby ocenić, która klasyfikacja dała bardziej wiarygodny wynik, wykorzystano następujące mierniki jakości stosowane w przypadku klasyfikacji tradycyjnej:

$$L = \sum_s \sum_{O_k, O_l \in A_s} d(O_k, O_l), \text{ oraz } P = \sum_s \sum_{O_k \in A_s, O_l \notin A_s} d(O_k, O_l).$$

Miara L jest sumą wszystkich sum odległości grupowych, mierzy zróżnicowanie obiektów wewnątrz klas. Najlepszy jest podział, dla którego wartość tego miernika jest najmniejsza. Zgodnie z tym kryterium w wyniku takiego podziału otrzymano klasy obiektów o najbardziej zbliżonych wartościach poszczególnych wag w zmiennych symbolicznych. Miernik P mierzy zróżnicowanie między grupami, poprawniejszy podział zatem można uzyskać za pomocą metody, która zalicza obiekty do poszczególnych klas w taki sposób, aby suma sum odległości danego obiektu od obiektów nienależących do danej klasy była jak największa. Dla klasyfikacji nauczycieli uzyskanej w wyniku zastosowania zbiorów przybliżonych otrzymano $L_1 = 82,405$, $P_1 = 107,975$, natomiast podział uzyskany w wyniku zastosowania zmiennych symbolicznych charakteryzują wartości mierników $L_2 = 69,156$, $P_2 = 204,393$.

Otrzymane wyniki świadczą o tym, że w wyniku klasyfikacji obiektów symbolicznych otrzymano bardziej zróżnicowane grupy obiektów o większym podobieństwie wewnątrz grup niż w przypadku modelowania przybliżonego. Dzięki przyjęciu do opisu obiektów zmiennych symbolicznych uwzględnione zostały odpowiedzi wszystkich ankietowanych studentów. Natomiast w modelowaniu przybliżonym na wynik oceny nauczycieli wpływ miały jedynie odpowiedzi najczęściej wybierane przez studentów.

6. Podsumowanie

W artykule porównano wyniki klasyfikacji wybranej grupy nauczycieli akademickich pod względem wpływu ich osobowości na popularność wykładów ze statystyki, otrzymanej za pomocą modelowania przybliżonego oraz klasyfikacji obiektów o zmiennych symbolicznych z wagami. Wyznaczony miernik Szmigła (por. [Szmigiel 1981]), oparty na tablicy kontyngencji, oznacza dość dużą niezgodność podziałów (wartość 0,25). Badanie jakości otrzymanych podziałów wskazuje na to, że rezultaty oceny zdolności dydaktycznych nauczycieli akademickich, otrzymane w wyniku zastosowania metody Warda do zbioru obiektów opisa-

nych zmiennymi symbolicznymi z wagami, są bardziej wiarygodne niż otrzymane za pomocą reguł decyzyjnych przybliżonego modelu frekwencji. Zastosowanie zmiennych symbolicznych z wagami pozwoliło wykorzystać każdą dostępną informację do oceny wpływu osobowości nauczycieli na frekwencję studentów na wykładach ze statystyki matematycznej.

Na jakość otrzymanych wyników może mieć wpływ także niezbyt duża liczba badanych obiektów oraz występowanie jednoelementowych grup. Nie można wówczas dokonać oceny niektórych nauczycieli. Nauczyciele jednak niechętnie poddają się tego rodzaju badaniom. Nie uwzględniono w obu klasyfikacjach takich czynników, jak ocena uzyskana za kurs czy wrodzone zdolności, które determinują wypowiedzi studentów.

Pozostaje również kwestia, czy badanie powinno być przeprowadzone bezpośrednio po zakończeniu kursu, czy też dopiero wówczas, gdy studenci ukończą studia. Z jednej strony po zakończeniu kursu studenci mają w pamięci wyraźny obraz przeprowadzonych wykładów, metod dydaktycznych stosowanych przez nauczyciela, z drugiej zaś strony osobowość nauczyciela i wiedza, którą przekazał, dopiero po czasie zostają docenione.

Literatura

- Jajuga K. (1984), *O sposobach określania liczby klas w zagadnieniu klasyfikacji i klasyfikacji rozmytej*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 262, AE, Wrocław, s. 17-29.
- Mrózek A., Płonka L. (1999), *Analiza danych metodą zbiorów przybliżonych*, Akademicka Oficyna Wydawnicza PLJ, Warszawa.
- Machowska-Szewczyk M. (2006), *Ocena wpływu osobowości nauczyciela na popularność wykładów ze statystyki matematycznej*, Materiały XI Sesji Naukowej Informatyki, Wydział Informatyki, Politechnika Szczecińska, s. 279-288.
- Machowska-Szewczyk M., Sompolska-Rzechuła A. (2007), *Zastosowanie zbiorów przybliżonych do oceny popularności wykładów ze statystyki matematycznej*, [w:] Taksonomia 14, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1169, AE, Wrocław, s. 141-148.
- Nowak E. (1990), *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa.
- Szmigiel C. (1981), *Modyfikacja wskaźnika zgodności kryteriów podziału*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 190, AE, Wrocław.
- Wilk J. (2006), *Miary odległości obiektów opisanych zmiennymi symbolicznymi z wagami*, [w:] Taksonomia 13, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1126, AE, Wrocław, s. 224-236.

**COMPARISON OF CLASSIFICATION RESULTS OF UNIVERSITY
TEACHERS REGARDING PERSONALITY INFLUENCE
ON POPULARITY OF STATISTICS LECTURES**

Summary

In the article the classification results which use objects described by symbolical variables with weights are compared to results of grouping based on rough sets elements. The quality research of the received divisions shows that results of didactics rate of university teachers, obtained as a result of using Ward's method to an objects collection described by symbolical variables, are more reliable than those received by decision-making rules of an approximate model of attendance.