

Metoda artykulacji preferencji decydenta na podstawie przykładów podjętych decyzji

Włodzimierz KWIATKOWSKI

Instytut Teleinformatyki i Cyberbezpieczeństwa, Wydział Cybernetyki, WAT,
ul. Gen. S. Kaliskiego 2, 00-908 Warszawa
wlodzimierz.kwiatkowski@wat.edu.pl

STRESZCZENIE: W artykule rozpatrywany jest problem wyznaczenia modelu preferencji decydenta na podstawie podjętych przez niego decyzji. Jest to podstawowe zagadnienie w zadaniach behawioralnej identyfikacji czy filtrowaniu informacji bazującym na współpracy. Wnioskowanie o preferencjach decydenta umożliwiła baza danych zawierająca wartości cech towarzyszących podjętym decyzjom. Jako podstawowy model preferencji służy wektor preferencji, określający wagi poszczególnych współrzędnych w procesie wyboru decyzji. Zaproponowana metoda artykulacji preferencji pozwala na uporządkowanie decyzji wyłącznie na podstawie proponowanego modelu, bez bezpośredniego odwoływania się do wskazanego wzorca.

SŁOWA KLUCZOWE: eksploracja danych, preferencje decydenta, wektor preferencji, atrybucja sytuacyjna.

1. Wprowadzenie

Użycie określenia "preferencje decydenta" opiera się na przyjęciu założenia, że postępowanie decydenta nie jest losowe, lecz racjonalne lub emocjonalne. Określenie to dotyczy sposobu wyboru decyzji, w tym wyboru decyzji na podstawie dokonywanych obserwacji. Indywidualizacja sposobu podejmowania decyzji może przejawiać się w odrębnym definiowaniu oceny użyteczności decyzji (zysków, strat), a także po prostu na własnym (np. emocjonalnym) sposobie wyboru decyzji spośród przedstawionych do oceny.

Przykłady działania decydenta (zaobserwowane albo wskazane przez niego jako akceptowalne) mogą być wykorzystywane do ujawnienia jego osobistych preferencji [9]. Rozpatrywany w niniejszym artykule problem dotyczy sytuacji, gdy każdemu możliwemu działaniu decydenta można przypisać wektor cech – niezależnie od tego, czy decydent z tego opisu korzysta. Podstawę wnioskowania

o preferencjach decydenta daje baza danych zawierająca wartości cech towarzyszących jego działaniom. Celem eksploracji danych jest wydobycie (ujawnienie) preferencji decydenta. Taką możliwość uzyskuje się, rzutując przykłady działań na przestrzeń cech. Określenie preferencji decydenta polega wtedy na określeniu znaczenia każdej cechy w procesie wyboru decyzji. Istotnym założeniem jest brak konieczności apriorycznego kategoryzowania cech¹.

W niniejszym artykule jako zasadę przyjmujemy założenie o braku wiedzy o tym, czy i na jakich obserwacjach decydent bazował. Można zauważyć, że jeśli decydent opierał wybór decyzji na znanych mu opisach decyzji w przestrzeni cech, to zadanie ujawnienia preferencji można sprowadzić do problemu odtwarzania odpowiednich reguł.

Celem artykulacji preferencji decydenta najczęściej jest atrybucja nie w przestrzeni obserwacji decydenta, lecz w przestrzeni danych dostępnych dla badacza lub będących przedmiotem jego zainteresowania. Możliwość uzyskiwania użytecznych wniosków występuje wówczas w przypadku, gdy eksplorowana przez badacza baza danych zawiera zapisy skorelowane z wykorzystywanymi przez decydenta informacjami. Od trafności doboru i jakości eksplorowanej bazy danych będzie zależeć użyteczność uzyskiwanych rozwiązań.

Celem niniejszego pracy jest zaproponowanie sposobu artykulacji preferencji, krócej mówiąc – modelu preferencji. Proponowany model bazuje na koncepcji wektora preferencji. Artykulacja preferencji polega na wyznaczeniu wag poszczególnych cech (*współrzędnych* tej przestrzeni, *atrybutów* relacyjnej bazy danych). Oczekuje się przy tym, że proponowana metoda powinna umożliwiać indywidualizację preferencji poszczególnych decydentów.

Najbardziej popularne metody eksploracji danych bazują na statystycznych modelach wykorzystywanych danych. Znajdują one skuteczne zastosowanie w przypadkach, gdy zbiór analizowanych danych jest wystarczająco obszerny. Przykładem są rozwiązania bazujące na kohortowych badaniach statystycznych, dotyczących określonych grup decydentów – o jednorodnych dla grupy preferencjach. Wymaganie, aby możliwe było artykułowanie preferencji dla każdego decydenta osobno, osłabia możliwość wykorzystywania takich badań.

¹ W przypadku podejmowania decyzji przez człowieka oczekuje się, że w wyniku eksploracji danych uzyska się model preferencji umożliwiający czytelną interpretację przyczyn postępowania decydenta, np. jako atrybucję przyczyn działania (ang. *causal attribution*). Omawiane w tym artykule zagadnienia są bliskie atrybucji sytuacyjnej (ang. *external attribution*), kiedy jako przyczyny działań decydenta upatrywane są czynniki zewnętrzne. Nie wykluczamy jednak możliwości, że na wybory decydenta mają wpływ cechy (emocje) decydenta. Całkowity brak korelacji między emocjami decydenta a czynnikami zewnętrznymi można uznać za raczej rzadki przypadek.

Możliwość indywidualizacji preferencji można uzyskać na podstawie wskazanych (lub podjętych) przez decydenta decyzji. Decyzje te są traktowane jako przykłady decyzji wzorcowych. Oznacza to konieczność zaproponowania takiej metody artykulacji preferencji, która jest skuteczna w przypadkach małej liczebności wzorcowych przykładów. Określenie "mała liczebność" dotyczy przy tym nie tylko możliwości wykorzystywania badań statystycznych. Należy uwzględnić także aspekt geometryczny, objawiający się w przypadku, gdy liczebność wzorców jest mniejsza od wymiaru przestrzeni cech. Można zauważyć, że przyjęte uwarunkowania problemu istotnie ograniczają możliwości atrybucji przyczyn działania.

Rozpatrywany problem artykulacji preferencji decydenta obejmuje modelowanie preferencji oraz wyznaczanie wartości parametrów przyjętego modelu. Proponowany wektorowy wskaźnik preferencji jest określony na przestrzeni cech charakteryzujących możliwe działania (tzn. towarzyszących lub kojarzonych z działaniami decydenta). Taki model preferencji pozwala odcinać zadanie artykulacji preferencji od zadania wyboru decyzji, tzn. opisuje zasady wyboru najlepszej decyzji w oderwaniu od wzorcowych przykładów działań decydenta. Prezentacja preferencji w proponowanej, wektorowej postaci jest powszechnie stosowana (często nawet wymagana formalnie), głównie ze względu na jej czytelność i łatwość wykorzystywania.

2. Prace związane

Rozpatrywany w artykułach [7] i [8] problem dotyczy porządkowania (ustalenia kolejności, szeregowania) decyzji zgodnie z preferencjami decydenta deklarowanymi poprzez wskazanie decyzji wzorcowych. W przedstawionych tam rozwiązaniach przyjęto, że podstawą takiego uporządkowania jest wykorzystanie funkcjonału bazującego na odległości wektora cech od wzorca. Zaproponowana w niniejszej pracy artykulacja preferencji również umożliwia porządkowanie decyzji, ale bez jawnego odwoływania się do wskazanego wzorca.

Problem wyboru decyzji na podstawie wskazanych przez decydenta przykładów występuje w zadaniach wyszukiwania informacji według zgłoszonych przez decydenta zapotrzebowań. Wykorzystywanie sygnalizowanych przez użytkownika wzorców jest istotnym problemem w zadaniach filtrowania informacji, bazujących na współpracy (ang. *collaborative filtering*) [1, 2, 10].

Problematyka optymalizacji decyzji na podstawie zadanych wzorców lub przykładów jest omawiana w pracach [5], [7]. Przedstawione w [7] rozwiązanie bazuje na opisie preferencji w postaci klastra wzorcowych decyzji. Jest ukierunkowane na bezpośrednie wyznaczanie decyzji zgodnie ze wskazanym przez decydenta wzorcem. Konsekwencją uznania wskazanych działań za

wzorcowe jest przyjęcie wskaźnika jakości decyzji jako odległości wektora cech badanego działania od wzorcowego klastra cech. W przedstawionych tam rozważaniach nie wyodrębnia się problematyki określenia preferencji decydena. Temat ten został podjęty w pracy [9]. Zaprezentowane tam rozważania także bazują na artykułowaniu preferencji jako klastra wzorców. Rozwiązany w tej pracy problem polega na poszerzaniu wzorca drogą uzupełniania składu wskazanego klastra.

Problem ujawniania preferencji można zaobserwować w pracach [1] i [2]. Prace te drążą temat *zrozumienia* znaczenia poszczególnych cech dla badanego wskaźnika jakości. Omawiany tam problem jest także przedstawiany przez autorów jako "estymacja *subiektywnych atrybutów*". Problem *zrozumienia* został w tym przypadku sformułowany jako "*interpretowalność* wyników klasyfikacji". Użyte określenie "subiektywnych atrybutów" jest bliskie modelowi preferencji w postaci "subiektywnego wektora preferencji" (ang. *Subjective Preference Vector*) [11].

Model wektorów preferencji PVM (ang. *Preference Vector Method*) [11] znajduje obecnie wiele zastosowań. Na uwagę zasługuje metoda PVM-VSI (ang. *Preference Vector Method - Vector Space of Increments*) [4]. Także aktualna jest technika TOPSIS (ang. *Technique for Order Preference by Similarity to an Ideal Solution*) [12]. Technika ta koresponduje z metodami optymalizacji na podstawie podobieństwa do rozwiązania idealnego [3, 4, 5].

3. Przestrzeń cech

Przyjmujemy, że elementy zadanego zbioru decyzji są ponumerowane od 1 do N . Dla każdej decyzji znana jest jej charakterystyka przedstawiona w postaci wektora liczb rzeczywistych. Wektor ten będziemy nazywać wektorem cech. Określenie to ma charakter umowny (poszczególne współrzędne tego wektora często uzyskiwane są po prostu jako towarzyszący decyzji wynik obserwacji czy pomiaru). Dla decyzji o indeksie k stosować będziemy następujące oznaczenie wektora cech:

$$\mathbf{a}_k = [a_{k,1}, a_{k,2}, \dots, a_{k,L}]^T, \quad \mathbf{a}_k \in R^L \quad (1)$$

Parametr L określa liczbę współrzędnych wektora cech. Wektory $\mathbf{a}_k \in R^L$ tworzą zbiór:

$$A = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N\}, \quad \mathbf{a}_k \in R^L \quad (2)$$

Wektory cech zestawiamy w postaci następującej macierzy:

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N], \quad \mathbf{a}_k \in R^L \quad (3)$$

Właściwości statystyczne cech są określone poprzez:

- wartość średnią:

$$\mathbf{m}_s = \frac{1}{N} \sum_{k=1}^N \mathbf{a}_k \quad (4)$$

- macierz kowariancji:

$$\mathbf{R}_s = \frac{1}{N-1} \sum_{k=1}^N (\mathbf{a}_k - \mathbf{m}_s)(\mathbf{a}_k - \mathbf{m}_s)^T \quad (5)$$

Przyjmujemy dalej, że

$$\det(\mathbf{R}_s) \neq 0 \quad (6)$$

Założenie to umożliwia stosowanie w przestrzeni cech metryki Mahalanobisa:

$$d_s(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{R}_s^{-1} (\mathbf{x} - \mathbf{y})} \quad (7)$$

gdzie: $\mathbf{x}, \mathbf{y} \in R^L$

4. Wzorce decyzji

Niech K oznacza liczebność klastra wzorcowych decyzji, czyli zaobserwowanych działań decydenta lub wskazanych przez niego jako akceptowalne. Skojarzone z tym klastrem wektory cech zestawiamy w postaci następującej macierzy:

$$\mathbf{A}_w = [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_K], \quad \hat{\mathbf{a}}_k \in A \subset R^L \quad (8)$$

Właściwości statystyczne cech wzorca są określone poprzez:

- wartość średnią:

$$\mathbf{m}_w = \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{a}}_k \quad (9)$$

- macierz kowariancji:

$$\mathbf{R}_w = \frac{1}{K-1} \sum_{k=1}^K (\hat{\mathbf{a}}_k - \mathbf{m}_w)(\hat{\mathbf{a}}_k - \mathbf{m}_w)^T \quad (10)$$

Podstawowym problemem jest zdefiniowanie metryki w przestrzeni cech – zgodnie z preferencjami decydenta (a nie odzwierciedlającej tylko statystyczne właściwości bazy danych). W praktyce macierz kowariancji \mathbf{R}_w może nie być dodatnio określona lub mogą wystąpić zbyt duże różnice jej wartości własnych

(macierz ta może być źle uwarunkowana). Także w tych przypadkach będziemy stosować metrykę Mahalanobisa, wykorzystując regularyzację zaproponowaną w pracy [8]:

$$d_w(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{y})} \quad (11)$$

gdzie:

$$\mathbf{R} = \rho \mathbf{R}_s + (1 - \rho) \mathbf{R}_w \quad (12)$$

Współczynnik regularyzacji ρ przyjmuje wartości z przedziału $[0, 1]$.

5. Wektorowy model preferencji

Niech $\mathbf{p}_w \in R^L$ oznacza wektor preferencji zgodny z zadaniem wzorcem. Poszczególne współrzędne tego wektora oznaczają wagi odpowiadających im cech. Niech wektor cech $\mathbf{x} \in R^L$ opisuje ocenianą decyzję, a funkcjonal $Q_w(\mathbf{x})$ określa sposób oceny decyzji. Proponowany model preferencji bazuje na następującej zależności oceny decyzji od opisującego ją wektora cech:

$$Q_w(\mathbf{x}) = \mathbf{p}_w^T \mathbf{x} \quad (13)$$

Wektor preferencji \mathbf{p}_w jest wyznaczany na podstawie oczekiwania, że wartości oceny $Q_w(\mathbf{x})$ będą bliskie jedności dla wszystkich wektorów $\hat{\mathbf{a}}_k$ cech wzorca. Wektor różnicy oznaczmy następująco:

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_K \end{bmatrix} = \begin{bmatrix} Q_w(\hat{\mathbf{a}}_1) - 1 \\ Q_w(\hat{\mathbf{a}}_2) - 1 \\ \dots \\ Q_w(\hat{\mathbf{a}}_K) - 1 \end{bmatrix} = \begin{bmatrix} \mathbf{p}_w^T \hat{\mathbf{a}}_1 - 1 \\ \mathbf{p}_w^T \hat{\mathbf{a}}_2 - 1 \\ \dots \\ \mathbf{p}_w^T \hat{\mathbf{a}}_K - 1 \end{bmatrix}$$

czyli

$$\mathbf{r} = \begin{bmatrix} \hat{\mathbf{a}}_1^T \\ \hat{\mathbf{a}}_2^T \\ \dots \\ \hat{\mathbf{a}}_K^T \end{bmatrix} \mathbf{p}_w - \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} = \mathbf{A}_w^T \mathbf{p}_w - \mathbf{z} \quad (14)$$

gdzie: $\mathbf{z} = [1, 1, \dots, 1]^T \in R^K$. Przedstawione oczekiwania dotyczące wyznaczonego modelu preferencji sformułujemy jako wymaganie minimalizacji następującej formy kwadratowej:

$$J(\mathbf{p}_w) = \mathbf{r}^T \mathbf{r} = (\mathbf{A}_w^T \mathbf{p}_w - \mathbf{z})^T (\mathbf{A}_w^T \mathbf{p}_w - \mathbf{z}) \quad (15)$$

Wektor $\bar{\mathbf{p}}_w$ minimalizujący ten funkcjonal ma postać [6]:

$$\bar{\mathbf{p}}_w = (\mathbf{A}_w \mathbf{A}_w^T)^{-1} \mathbf{A}_w \mathbf{z} \quad (16)$$

Warunkiem wystarczającym istnienia rozwiązania jest liniowa niezależność wektorów cech wzorca. Wykorzystanie wyznaczonego w ten sposób wektora preferencji $\bar{\mathbf{p}}_w \in R^L$ do oceny decyzji o cesze $\mathbf{x} \in R^L$ sprowadza się do wyznaczenia wartości funkcjonału:

$$\bar{Q}_w(\mathbf{x}) = (\bar{\mathbf{p}}_w^T \mathbf{x} - 1)^2 \quad (17)$$

Uzyskiwane w ten sposób wartości mogą stanowić podstawę porządkowania decyzji zgodnie z wyznaczonym modelem preferencji.

6. Eksperyment obliczeniowy

Celem badań było eksperymentalne potwierdzenie użyteczności zaproponowanej metody artykułowania preferencji decydenta. Obawy budzi fakt, że bezpośrednio wykorzystywanie wektora preferencji do wyznaczania oceny zadanej cechy \mathbf{x} zgodnie z liniową zależnością (13) oznacza nieograniczone preferowanie cech zgodnie z kierunkiem wektora preferencji $\bar{\mathbf{p}}_w$. Prowadzić to może do efektu nie zamierzonego przez decydenta. Zastosowanie zależności (17) ogranicza tę możliwość.

Oceny bezpośrednio zależne od wzorca będą bazować na odległości cechy \mathbf{x} analizowanej decyzji od wartości średniej wzorca:

$$D_w^2(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_w)^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{m}_w) \quad (18)$$

Oczywistym oczekiwaniem jest, że uporządkowanie decyzji otrzymane na podstawie ocen $\bar{Q}_w(\mathbf{x})$ bazujących na wektorze preferencji nie będzie istotnie różnić się od uporządkowania uzyskiwanego na podstawie odległości cech od wzorca. Możliwość wystąpienia takiej różnicy jest kosztem poniesionym za rezygnację z bezpośredniego wykorzystywania wzorca.

Oceny użyteczności zaproponowanego modelu preferencji dokonano na podstawie porównania wyników porządkowania decyzji w dwu przypadkach:

bezpośrednio na podstawie wskazanych przykładów i na podstawie wyznaczonego wcześniej modelu preferencji.

Użyte w badaniach dane pomiarowe zostały pobrane z archiwalnych baz danych stacji IMGW². W wybranym do analizy zbiorze pomiarów „decyzja” oznacza pojedynczą stację i jest scharakteryzowana wektorem $L = 6$ pomiarów (cech decyzji). Analizowana baza danych zawiera wyniki pomiarów $N = 56$ stacji. Wykorzystywane do obliczeń dane pomiarowe (nazywane dalej bazą danych), zostały przedstawione w tablicy.1. Analizowane decyzje będą wskazywane za pomocą indeksów bazy (kolumna 1. w tabeli).

Tablica 1. Baza danych

Indeks	Stacja	t	tx	tn	tng	RH	VA
1	BIELSKO-BIAŁA	3.8	6.6	-1.8	-8.2	57.6	4.5
2	ZAKOPANE	-0.9	5.0	-7.3	-13.0	68.9	3.9
3	NOWY SĄCZ	3.7	8.1	-1.6	-2.8	62.4	4.8
4	KROSNO	1.9	5.1	-2.0	-6.8	76.4	5.3
5	LESKO	1.3	3.7	-2.2	-6.7	72.8	4.8
6	JELEŃ GÓRA	4.5	10.1	-3.8	-5.8	65.3	5.5
7	KŁODZKO	2.0	3.9	-0.2	-2.6	73.4	5.2
8	OPOLE	3.6	6.2	-0.6	-2.5	69.4	5.4
9	RACIBÓRZ	3.5	5.7	-0.3	-2.8	69.8	5.4
10	CZĘSTOCHOWA	2.7	6.2	-1.2	-3.5	71.8	5.2
11	KATOWICE	3.3	7.2	-1.1	-3.1	67.3	5.1
12	KRAKÓW-BALICE	3.0	8.2	-2.5	-5.5	74.6	5.4
13	KIELCE-SUKÓW	2.4	7.3	-2.6	-3.8	74.6	5.2
14	TARNÓW	4.5	9.5	-0.8	-2.6	59.6	4.8
15	SANDOMIERZ	3.2	7.7	-1.7	-2.4	71.8	5.3
16	RZESZÓW-JASIONKA	3.4	6.9	-1.6	-3.0	70.4	5.4
17	ZAMOŚĆ	2.7	6.9	-0.8	-2.4	73.3	5.3
18	ZIELONA GÓRA	4.0	7.4	-0.3	-0.8	80.0	6.5
19	LEGNICA	4.3	7.5	-1.1	-3.0	71.3	5.9
20	LESZNO	4.5	8.3	-0.1	-2.3	71.6	6.0
21	WROCLAW	4.6	7.8	-0.4	-2.3	65.9	5.6
22	KALISZ	3.4	7.0	0.0	-0.6	76.6	5.9
23	WIELUŃ	3.3	6.5	-0.3	-0.8	69.6	5.4
24	ŁÓDŹ	3.5	7.1	-0.9	-2.5	69.4	5.4
25	SULEJÓW	3.2	6.8	-0.7	-1.4	73.0	5.5

² Źródło danych: IMGW Dane pomiarowo-obszaryjne. Dane meteorologiczne, dobowe, klimatyczne. Plik: k_d_07_2020.csv (2020_07_k.zip). URL: https://dane.imgw.pl/data/dane_pomiarowo_obszaryjne/dane_meteorologiczne/dobowe/klimat/2020/. Dostęp: 23.08.2021.

26	KOZIENICE	4.1	8.8	-0.3	-1.1	66.9	5.3
27	LUBLIN-RADAWIEC	2.5	7.6	-1.5	-4.6	76.4	5.4
28	WŁODAWA	2.4	7.3	-1.7	-2.7	77.8	5.5
29	SŁUBICE	5.4	9.6	1.0	0.4	77.1	6.9
30	GORZÓW WLKP.	4.3	8.1	0.2	-1.5	82.3	6.9
31	POZNAŃ	4.0	7.4	0.8	-0.6	76.4	6.2
32	KOŁO	3.3	6.5	-0.4	-1.3	74.8	5.7
33	PŁOCK	3.3	6.5	-0.4	-1.1	74.1	5.6
34	WARSZAWA	4.0	8.0	-0.9	-2.8	69.8	5.5
35	SIEDLCE	3.7	7.1	-0.1	-1.4	68.5	5.3
36	TERESPOL	3.2	6.4	-0.7	-1.9	75.1	5.7
37	ŚWINOUJŚCIE	4.9	7.4	2.5	-0.2	84.9	7.5
38	SZCZECIN	5.4	9.0	2.1	0.3	78.5	7.2
39	RESKO	3.5	5.9	1.1	-0.8	86.8	6.9
40	PIŁA	2.6	5.2	-0.7	-1.0	86.4	6.4
41	CHOJNICE	1.5	3.5	-0.6	-1.0	90.3	6.2
42	TORUŃ	2.9	5.7	-0.4	-3.1	80.3	6.0
43	MŁAWA	2.9	6.0	-0.8	-1.6	79.8	5.9
44	OLSZTYN	3.1	5.6	-0.1	-0.7	76.0	5.7
45	MIKOŁAJKI	3.3	5.9	0.8	-1.0	73.3	5.6
46	BIAŁYSTOK	3.2	7.0	0.0	-3.1	76.0	5.7
47	KOŁOBRZEG	3.8	5.9	1.6	1.4	85.9	6.9
48	KOSZALIN	2.9	4.7	1.0	0.3	87.3	6.6
49	USTKA	3.2	5.2	1.4	0.9	83.1	6.4
50	ŁEBA	3.1	5.7	1.0	0.2	88.1	6.7
51	ŁĘBORK	2.9	5.5	0.9	-1.3	83.6	6.3
52	HEL	3.5	5.5	2.2	-0.5	83.1	6.5
53	GDAŃSK ŚWIBNO	2.7	5.8	0.0	-1.4	85.1	6.3
54	ELBLĄG	1.8	4.4	-0.7	-0.8	84.6	5.8
55	KĘTRZYN	3.3	6.0	0.1	-0.7	76.9	5.9
56	SUWAŁKI	2.6	5.5	-0.6	-2.4	76.0	5.5

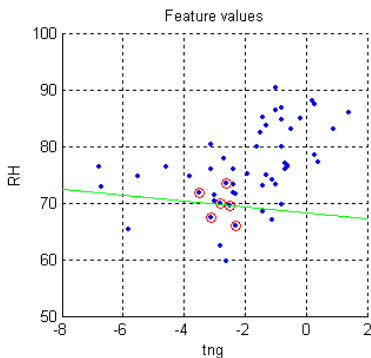
Przedstawimy wyniki obliczeń dla wzorca zdefiniowanego zbiorem indeksów $W = \{7, 8, 9, 10, 11, 21\}$. Wzorec ten scharakteryzowany jest zbiorem wektorów cech:

$$A_w = \{ \mathbf{a}_k \in A \subset R^L : k \in W \} = \{ \mathbf{a}_7, \mathbf{a}_8, \mathbf{a}_9, \mathbf{a}_{10}, \mathbf{a}_{11}, \mathbf{a}_{21} \} \quad (19)$$

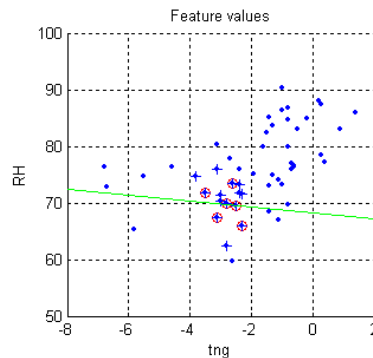
Wektory te są zestawiane w macierz:

$$\mathbf{A}_w = [\widehat{\mathbf{a}}_1, \widehat{\mathbf{a}}_2, \widehat{\mathbf{a}}_3, \widehat{\mathbf{a}}_4, \widehat{\mathbf{a}}_5, \widehat{\mathbf{a}}_6] = [\mathbf{a}_7, \mathbf{a}_8, \mathbf{a}_9, \mathbf{a}_{10}, \mathbf{a}_{11}, \mathbf{a}_{21}] \quad (20)$$

W celu uzyskania jednoznacznej wizualizacji wyników ograniczymy liczbę współrzędnych wektora cecho do $L = 2$ pomiarów (w bazie danych wybieramy kolumny *tng* oraz *RH*). Przyjęte do obliczeń dane zostały przedstawione na rys.1 (są przedstawione jako punkty niebieskie).



Rys. 1. Wizualizacja decyzji w przestrzeni cech. Cechy wzorcowe oznaczono kółkami koloru czerwonego. Kolorem zielonym oznaczono prostą aproksymującą zbiór cech wzorcowych.



Rys. 2. Wizualizacja decyzji w przestrzeni cech. Cechy zaliczone do rekomendacji oznaczono znakiem plus. Cechy wzorcowe oznaczono kółkami koloru czerwonego. Kolorem zielonym oznaczono prostą aproksymującą zbiór cech wzorcowych.

Podstawą dokonywanych porównań jest uporządkowanie decyzji bezpośrednio według zadanego wzorca. Porządkując w ten sposób całą bazę danych, uzyskuje się następującą sekwencję decyzji (rozpoczynającą się od decyzji najbliższych wzorcowi): 9,34,16,19,8,24,11,15,21,10,20,7,17,46,13,3,56,28,35,36,14,26,42,25,27,32,23,33,45,43,12,44,55,30,31,22,18,51,53,6,54,52,40,29,5,4,38,39,37,41,48,50,49,47,1,2.

Do oceny uzyskiwanego uporządkowania wykorzystamy współczynnik korelacji wektora po uporządkowaniu i wektora przed uporządkowaniem. Współrzędne wektorów oznaczają numer miejsca poszczególnych decyzji w opisywanym uporządkowaniu. Według tej zasady decyzja 9 ma numer 1, decyzja 34 - numer 2 itd. Wartość współczynnika korelacji uporządkowania decyzji według numeru w tabeli i uporządkowania według przyjętego do obliczeń wzorca wynosi 0,3277. Oczekuje się, że zarówno wartość współczynnika korelacji uporządkowania według przyjętego wzorca, jak i uporządkowania według modelu preferencji będzie zbliżona do 1.

Obliczony według wzoru (16) wektor preferencji ma wartość $\bar{\mathbf{p}}_w = [0.0076, 0.0147]^T$. Na rys.1 przedstawiono prostą aproksymującą zadany wzorec. Prostą tę tworzą punkty $\mathbf{x} \in R^2$ spełniające warunek: $\bar{\mathbf{p}}_w^T \mathbf{x} = 1$. Uporządkowanie według wektora preferencji uzyskuje się wykorzystując

sortowanie wartości funkcjonału (17) od wartości najmniejszej do wartości największej. Odpowiada to zadaniu porządkowania wektorów cech na podstawie ich odległości od prostej aproksymującej. W wyniku uzyskano w przykładzie następującą sekwencję: 9,34,8,24,35,16,23,5,19,10,26,20,15,11,12,21,7,17,25,13,45,4,33,27,36,32,6,2,46,56,3,44,31,22,28,55,29,14,38,42,43,18,30,52,51,1,49,54,53,37,40,39,47,48,50,41. Wartość współczynnika korelacji tej sekwencji (tzn. uporządkowania według wektora preferencji) i uporządkowania decyzji bezpośrednio według wzorca (tzn. uporządkowanej sekwencji przedstawionej wcześniej) wynosi 0,7761³.

Mimo widocznej dużej poprawy uporządkowania, widoczne jest zakłócanie uporządkowania przez wektory cech położone daleko od wzorca ale blisko prostej aproksymującej. Jest to immanentna wada wynikająca z bazowania na odległości nie od klastra wzorców, a od prostej aproksymującej ten klastr. Skutki można zmniejszyć, ograniczając się do porządkowania tylko wektorów cech położonych dostatecznie blisko wzorca. Z praktycznego punktu widzenia oznacza to ograniczenie liczby rekomendowanych decyzji. Na rys.2 przedstawiono wizualizację ograniczenia liczby rekomendowanych wzorców do 16. Uzyskana wówczas sekwencji uporządkowania według wektora preferencji jest następująca: 9,34,8,24,16,19,10,20,15,11,21,7,17,13,46,3. Wartość współczynnika korelacji tej sekwencji i uporządkowania decyzji bezpośrednio według wzorca (tj. sekwencji 9,34,16,19,8,24,11,15,21,10,20,7,17,46,13,3) wyniosła 0,9265⁴. Interpretację wyników ułatwia przedstawienie nazw miejscowości wskazanych jako wzorcowe: KŁODZKO, OPOLE, RACIBÓRZ, CZĘSTOCHOWA, KATOWICE, WROCŁAW. Wynik rekomendacji według odległości od wzorca przedstawia następująca sekwencja: RACIBÓRZ, WARSZAWA, RZESZÓW-JASIONKA, LEGNICA, OPOLE, ŁÓDŹ, KATOWICE, SANDOMIERZ, WROCŁAW, CZĘSTOCHOWA, LESZNO, KŁODZKO, ZAMOŚĆ, BIAŁYSTOK, KIELCE-SUKÓW, NOWY SĄCZ. Sekwencja rekomendacji według wektora preferencji jest następująca: RACIBÓRZ, WARSZAWA, OPOLE, ŁÓDŹ, RZESZÓW-JASIONKA, LEGNICA, CZĘSTOCHOWA, LESZNO, SANDOMIERZ, KATOWICE,

³ Wartość ta została obliczona jako współczynnik korelacji wektora numerów uporządkowania według odległości (czyli sekwencji: 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56) oraz wektora uporządkowania według modelu preferencji (czyli sekwencji: 1,2,5,6,19,3,27,45,4,10,22,11,8,7,31,9,12,13,24,15,29,46,28,25,20,26,40,56,14,17,16,32,35,36,18,33,44,21,47,23,30,37,34,42,38,55,53,41,39,49,43,48,54,51,52,50).

⁴ Wartość ta została obliczona jako współczynnik korelacji wektora numerów uporządkowania według odległości (czyli sekwencji: 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16) oraz wektora uporządkowania według modelu preferencji (czyli sekwencji: 1,2,5,6,3,4,10,11,8,7,9,12,13,15,14,16).

WROCLAW, KLÓDZKO, ZAMOŚĆ, KIELCE-SUKÓW, BIAŁYSTOK,
NOWY SĄCZ.

7. Wnioski końcowe

Agregacja zaobserwowanych działań w postaci modelu preferencji wydaje się być racjonalnym sposobem określania wzorca zachowań decydenta w dostępnej przestrzeni obserwacji. Idea ta ściśle koresponduje z koncepcją uczenia sztucznych sieci neuronowych. Przyjęty model wektora preferencji ściśle odpowiada pojedynczemu, liniowemu neuronowi. Wskazana odpowiedniość umożliwia świadome kształtowanie struktury sieci neuronów, wykorzystywanych np. w problemach identyfikacji behawioralnej.

Proponowana metoda artykulacji nie nakłada ograniczeń na badany sposób podejmowania decyzji. Uzyskanie ostrego (ale niekoniecznie adekwatnego) obrazu preferencji może ułatwić świadome kształtowanie go przez decydenta.

Jakość uzyskiwanego modelu jest zależna od dostępności badacza do danych silnie skorelowanych z okolicznościami towarzyszącymi działaniom decydenta. Uniwersalność i prostota proponowanej metody sprawia, że metoda ta może być wykorzystywana jako narzędzie do poszukiwania cech istotnych z punktu widzenia działań decydenta. Badania takie mają charakter porównawczy i polegają na wykonaniu serii eksperymentów obliczeniowych dla różnych baz danych lub różnie dobieranych atrybutów jednej bazy danych

Przyjęty model artykulacji preferencji ma charakter lokalny. Wynika to z faktu zastąpienia zadania wyszukiwania najmniejszej odległości od klastra wzorców zadaniem wyszukiwania najmniejszej odległości od różnorodności liniowej aproksymującej ten klastr. Wektor preferencji jest uzyskiwany także w przypadku małej liczby wskazań decyzji (cech) wzorcowych. Nie występuje problem regularyzacji, tak jak w przypadku bazowania na minimalizacji odległości względem wzorcowego klastra.

Wykorzystywanie funkcjonału (18) prowadzi do uzależniania ewaluacji decyzji tylko od wartości średniej \mathbf{m}_w cech wzorcowych. Większego wpływu poszczególnych wzorców na wynik można oczekiwać, biorąc za podstawę np. wartość średnią odległości badanego wektora cech od wszystkich wzorców, tzn. wykorzystując funkcjonał:

$$D_w^2(\mathbf{x}) = \frac{1}{N_w} \sum_{j \in I_w} (\mathbf{x} - \mathbf{a}_j)^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{a}_j) \quad (21)$$

gdzie: I_w – zbiór indeksów wzorców.

Literatura

- [1] BIELSKI A., TRZCIŃSKI T., *Pay Attention to Virality: Understanding Popularity of Social Media Videos with the Attention Mechanism*. IEEE/CVS Conference on Computer Vision and Pattern Recognition (CVPR) WorkShops, IEEE Access, 2018.
- [2] BIELSKI A., TRZCIŃSKI T., *Understanding Multimodal Popularity Prediction of Social Media Videos With Self-Attention*. IEEE Access, 2018.
- [3] HWANG C-L., YOON K., *Multiple Attribute Decision Making. Methods and Applications A State-of-the-Art Survey*. Springer, Berlin Heidelberg, 1981.
- [4] KANNCHEN M., *Using the PVM-VSI (Preference Vector Method - Vector Space of Increments) method in supporting the decision related to the purchase of an electric family car*. 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, 2021.
- [5] KWIATKOWSKI W., *Optymalizacja względem wzorców*. Biuletyn WAT, vol. 32, 3, 1983, s. 93-104.
- [6] KWIATKOWSKI W., *Metody automatycznego rozpoznawania wzorców*. BEL Studio, Warszawa, 2010.
- [7] KWIATKOWSKI W., *Recommendations as a result of decision evaluations based on reference examples*. Teleinformatics Review, No. 1-2, 2019, pp. 3-23.
- [8] KWIATKOWSKI W., *The regularization method in the classification task according to given examples*. Teleinformatics Review, No. 3-4, 2019, pp. 3-13.
- [9] KWIATKOWSKI W., *Rekomendacje jako wynik oceny preferencji na podstawie wskazanych przykładów*. Przegląd Teleinformatyczny, nr 1-4, 2021, s. 3-16.
- [10] MOBASHER B., DAI H., LUO T., NAKAGAWA M., *Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data*. in: Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01), 2001.
- [11] NERMEND K., *Wielokryterialna metoda wektora preferencji jako narzędzie wspomagające proces decyzyjny*. Przegląd Statystyczny, R. LXII - Zeszyt 1, 2015, str. 93-115.
- [12] ZAVADSKAS E.K., MARDANI A., TURSKIS Z., JUSOH A., KHALIL MD NOR, *Development of TOPSIS Method to Solve Complicated Decision-Making Problems — An Overview on Developments from 2000 to 2015*. International Journal of Information Technology & Decision Making, Vol. 15, No. 3.

**A method of articulating the decision maker's preferences based
on examples of decisions made**

ABSTRACT: The paper examines the problem of determining a decision-maker's preference model based on the decisions he makes. This is a fundamental issue in behavioral identification tasks or collaborative information filtering. Inferences about the decision-maker's preferences are made possible by a database containing the values of features accompanying the decisions made. The preference vector serves as the basic preference model, determining the weights of individual coordinates in the decision selection process. The proposed method of preference articulation allows for ordering decisions solely on the basis of the proposed model, without directly referring to the indicated pattern.

KEYWORDS: data mining, decision-maker's preferences, preference vector, external attribution.

Praca wpłynęła do redakcji: 29.09.2023 r.