dr inż. Monika Stankiewicz

Zachodniopomorska Szkoła Biznesu

Teoria zbiorów przybliżonych w analizie sprzedaży wydarzeń kulturalnych

Streszczenie:
Celem niniejszego opracowania jest przedstawienie możliwości wykorzystania teorii zbiorów przybliżonych do analizy sprzedaży biletów na wydarzenia kulturalne. W pracy podjęto próbę opracowania algorytmu pozwalającego rozwiązać problem decyzyjny związany z wyborem najlepiej sprzedających się pozycji repertuarowych. W pracy poruszono także problem oceny wiarygodności zgromadzonych danych w kontekście zgodności pojedynczych obiektów z pozostałymi na podstawie zachodzących między nimi zależności. Punktem wyjścia do przeprowadzenia badań było przypuszczenie, iż liczba sprzedanych biletów danego rodzaju jest związana z typem wydarzenia, jakie jest oferowane, a przeprowadzona analiza sprzedaży miała określić, wśród jakiej grupy klientów dana oferta cieszy się (lub nie) zainteresowaniem. Przeprowadzone w tym obszarze badania pozwoliły jednoznacznie wnioskować, jaki klient korzysta z oferty. W perspektywie daje to możliwość dostosowania do klienta planów repertuarowych i zwiększenia sprzedaży biletów na wydarzenia kulturalne.

Słowa kluczowe: teoria zbiorów przybliżonych, analiza sprzedaży wydarzeń kulturalnych

Wprowadzenie

Na dzień przeprowadzenia badań w Polsce działa 11 teatrów muzycznych (1 państwowy: Teatr Wielki – Opera Narodowa, 6 samorządowych: Opera Śląska w Bytomiu, Państwowa Opera Bałtycka w Gdańsku, Opera Krakowska, Teatr Wielki w Łodzi, Opera na Zamku w Szczecinie, Warszawska Opera Kameralna, 4 samorządowe współprowadzone: Opera Wrocławska, Opera i Filharmonia Podlaska, Opera Nova – Państwowa Opera w Bydgoszczy, Teatr Wielki im. S. Moniuszki w Poznaniu) oraz około 15 mniejszych, kameralnych. Wszystkie z nich wywierają znaczny wpływ na życie artystyczne i kulturalne miast.

Wybrana i poddana analizie metodą zbiorów przybliżonych prof. Pawlaka instytucja kultury (na prośbę której nie ujawniono nazwy) to jedyna scena muzyczna działająca w obrębie swojego województwa. Przedmiotem jej działalności jest realizacja własnych przedstawień teatralnych i muzycznych, widowisk baletowych i koncertów estradowych, organizacja występów gościnnych zespołów i solistów z kraju i zagranicy oraz innych przedsięwzięć artystycznych. Specyfika prowadzonej przez nią działalności usługowej wymaga utrzymywania stale w repertuarze około 10 tytułów – z zastrzeżeniem, że każdy tytuł nie może być powtarzany więcej niż 10 do 40 razy. Ważna jest zatem elastyczność planowania repertuaru i dbałość o to, aby w miejsce przedstawień zagranych powstawały nowe. Jak każda instytucja kultury, badany podmiot prowadzi bardzo aktywną działalność artystyczną, proponując widzom zróżnicowany repertuar muzyczny – od opery przez balet, spektakle dziecięce po operetkę i musical, grając w kraju ok. 80 przedstawień rocznie.

W rozważanym podmiocie nigdy nie przeprowadzono analizy danych sprzedażowych ani badań z wykorzystaniem metod czy narzędzi wspomagania decyzji. Z wieloletnich obserwacji samej instytucji wynika jednak, iż każde przedstawienie przez nią udostępniane ogląda przeciętnie ok. 20 tys. osób. Daje to szansę na maksymalnie 40 wykonań każdego tytułu w ciągu kilku sezonów.

Ustalanie repertuaru stanowi istotny element działalności każdej instytucji kultury. Pozyskiwanie szerszej widowni i nowych widzów wiąże się z wystawianiem różnorodnej oferty do sprzedaży, przeznaczonej zarówno dla widzów dorosłych, jak i dzieci i młodzieży. Różne grupy odbiorców kultury wysokiej wymagają również wprowadzenia zróżnicowanej oferty biletowej.

Środowisko badań

Instytucja jednorazowo udostępniła próbę danych sprzedażowych za okres od października 2008 do marca 2009 roku. W badanym okresie w repertuarze zaproponowano widzom: 3 tytuły operetek, 4 tytuły oper, 2 koncerty, 1 musical, 1 balet, 1 spektakl gościnny, 2 tytuły spektakli dla dzieci. Każdy tytuł grany był w wymienionym okresie od 2 do 19 razy. Łącznie na wszystkie spektakle sprzedano 30 757 biletów, w tym: 14 924 biletów normalnych, 1 250 pracowniczych, 9 597 biletów zbiorowych, 728 biletów familijnych. Instytucja przygotowała na sezon artystyczny 2008/2009 (trwający od października do czerwca) 3 premiery. W badanym okresie dysponowała widownią o 536 miejscach.
W sprzedaży dostępne były:

- bilety normalne (pełnopłatne),
- bilety familijne: 2 osoby dorosłe + dziecko do 12. roku życia,
- bilet zbiorowy: co najmniej 10 biletów z upustem cenowym.

Rozdysponowano również zaproszenia (bilet w cenie 50% wartości biletu normalnego) i noty (bilet bezpłatny, przeznaczony dla gości i sponsorów Teatru).

Raport ze sprzedaży biletów daje wskazania, wśród jakiej grupy klientów dany repertuar cieszy się większym zainteresowaniem, np. bilet zbiorowy kupowany jest głównie przez szkoły i tylko na określone typy spektakli. Zmniejszona ich sprzedaż mogłaby być podstawą przypuszczeń, że należałoby podjąć zintensyfikowane działania mające na celu informowanie placówek szkolnych o nowych propozycjach repertuarowych lub wystawienie dodatkowych spektakli w godzinach przedpołudniowych.

Przyjęta metoda badawcza

Teoria zbiorów przybliżonych prof. Pawlaka[1] jest matematycznym podejściem do pojęć nieostrych, metodą analizy danych[2] opisującą rzeczywistość ograniczoną liczbą atrybutów. Każdy opis dotyczy grupy elementów o tych samych wartościach, nierozróżnialnych z punktu widzenia opisu.

Jak podkreśla prof. Tadeusiewicz zbiory przybliżone oferują bardzo ogólne podejście, oparte na nierozróżnialności elementów przy dostępnych danych. „Jest to coś, co istnieje obiektywnie, trzeba tylko odpowiednią metodą informatyczną uwzględnić to w systemie sztucznej inteligencji". Często porównuje się je z metodą zbiorów rozmytych oferujących podejście numeryczne, opartych na mierze należenia elementu do zbioru, wyrażonej liczbą pomiędzy 0 a 1 (często wyznaczaną arbitralnie). Obie metody zdaniem profesora mają zwolenników i przeciwników, i tak samo zalety jak i wady[3].

Teoria zbiorów przybliżonych określa system informacyjny, tablice decyzyjne, jakość klasyfikacji, redukt, dolne i górne przybliżenia, relacje nierozróżnialności i reguły decyzyjne. Na potrzeby niniejszego artykułu przyjęto definicje za R.K. Nowickim[4]:

[1] Z. Pawlak, Information systems – theoretical foundations, Information Systems, vol. 6, no. 8. Pergamon Press Ltd, Great Britain 1981, s. 205-215.
[2] Z. Pawlak, Zbiory przybliżone. Nowa matematyczna metoda analizy danych, (red. mer.) S. Janeczko, Osiągnięcia Nauki i Techniki. Kierunki Rozwoju i Metody, Konwersatorium Politechniki Warszawskiej, Wkładka do Miesięcznika Politechniki Warszawskiej nr 5, 2004.
[3] R. Tadeusiewicz, http://ryszardtadeusiewicz.natemat.pl/150285,polska-wyspa-w-archipelagu-sztucznej-inteligencji-zbiory-przyblizone (06.04.2016r).
[4] R.K. Nowicki, Rozmyte systemy decyzyjne w zadaniach z ograniczoną wiedzą, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2009, s. 10-11.

1) Systemem informacyjnym nazywamy uporządkowaną czwórkę:

$$SI = < U, Q, V, f > \quad (1)$$

gdzie:
U – przestrzeń rozważań – zbiór obiektów lub stanów,
Q – zbiór cech (atrybutów) tych obiektów,
V – zbiór wartości cech,
f – funkcja informacyjna.

Przy czym $U, Q, V \neq \phi$ i skończone. Poszczególne cechy $q \in Q$ przyjmują wartości ze zbiorów V_q, więc zachodzi $V = \bigcup_{q \in Q} V_q$.

Funkcja informacyjna jest funkcją zupełną i przyporządkowuje wartości cechom obiektów, $f : U \times Q \to V$, czyli cecha *q* obiektu *x* ma wartość $\upsilon_q(x) = f(x, q)$ dla każdego $q \in Q$ i $x \in U$.

2) Tablicą decyzyjną nazywamy uporządkowaną piątkę $DT = < U, C, D, V, f >$,
gdzie:
U – przestrzeń rozważań – zbiór obiektów lub stanów,
C – zbiór cech (atrybutów) warunkowych tych obiektów,
D – zbiór atrybutów decyzyjnych,
V – zbiór wartości cech,
f – funkcja informacyjna, która przyporządkowuje wartości cechom obiektów, $f : U(C \cup D) \to V$, czyli cecha $q \in (C \cup D)$ obiektu $x \in U$ ma wartość $\upsilon_q(x) = f(x, q)$.

Zbiór cech Q w tablicy decyzyjnej dzieli się na dwa rozłączne zbiory: C – atrybutów warunkowych i D – atrybutów decyzyjnych. Każdy wpis w tablicy odpowiada regułom typu: JEŻELI *wartość atrybutu warunkowego*, TO *wartość atrybutu decyzyjnego*. Jej zadaniem jest przechowywanie wiedzy wykorzystywanej we wnioskowaniu. D występuje często jako zbiór jednoelementowy $D = d$ [5]. Zależności pomiędzy atrybutami omówiono w pracy[6].

Klasą abstrakcji $[\hat{x}]_R \subseteq U$ w pewnej niepustej przestrzeni U, dla określonego elementu $\hat{x} \in U$ oraz relacji równoważności $R \subseteq U \times U$ nazywa się następujący podzbiór przestrzeni U: $[\hat{x}]_R = \{x \in U : \hat{x}Rx\}$.

Warto wskazać, iż w literaturze *spójność tablicy decyzyjnej* definiowana jest na dwa sposoby. Pierwszy z nich, zaproponowany przez A. Skowrona, L. Polkowskiego i J. Komorowskiego[7], mówi, że tablica decyzyjna A jest spójna (deterministyczna), jeżeli $|\partial_A(x)| = 1$ dla każdego $x \in U$, w przeciwnym razie A jest niespójna (niedeterministyczna). Łatwo zauważyć, że tablica decyzyjna A jest spójna wtedy i tylko wtedy, gdy $POS_A(d) = U$. Ponadto, jeżeli $\partial_B = \partial_B{}'$, to $POS_B(d) = POS_B{}'(d)$ dla każdej pary z niepustych zbiorów $B, B' \subseteq A$. Natomiast Z. Pawlak w swoich pracach[8] spójność tablicy decyzyjnej warunkuje spójnością reguł w niej zawartych. Zgodnie z jego twierdzeniem tablica zawierająca niespójne reguły decyzyjne jest automatycznie niespójna, a w odwrotnym przypadku, tj. gdy zawiera tylko spójne reguły, jest spójna.

---

[5] R. Susmaga, Experiments in Incremental Computation of Reducts, [w:] L. Polkowski, Skowron A. (Eds.), Rough Sets in Knowledge Discovery 2: Applications, case studies and software systems, Physica-Verlag Heidelberg, New York 1998, s. 530-553.
[6] Z. Pawlak, Data analysis – the rough sets perspective, [w:] J. Chojcan, J. Łęski (red.), Zbiory rozmyte i ich zastosowanie, Wydawnictwo Politechniki Śląskiej, Gliwice 2001, s. 173-183.
[7] J. Komorowski, L. Polkowski, A. Skowron, Rough Sets, A Tutorial, [w:] S.K. Pal, A. Skowron (eds.) Rough Fuzzy Hybridization, A New Trend in Decision Making, Springer, Singapore 1999, s. 1-98.
[8] Z. Pawlak, Data Analysis with Rough Set Theory, Proceedings of CODATA'96, October 1996, Tsukuba, Japan.

3) Relacją równoważności nazywamy taką relację, która jest w swojej dziedzinie zwrotna, symetryczna i przechodnia i pozwala na podział przestrzeni ilorazu zbioru U przez relację R. Elementy tego podziału to rodzina klas abstrakcji (zbiory rozłączne).

Relacją $\widetilde{P}$ **(nierozróżnialności**, *indiscernibility relation*) nazywamy relację $\widetilde{P}$ określoną w przestrzeni $U \times U$ zdefiniowaną następująco:

$$x\widetilde{P}\hat{x} \Leftrightarrow f(x,q) = f(\hat{x},q) \forall q \in P,$$

gdzie:

$x, \hat{x} \in U$ – dowolne obiekty,

$P \subseteq Q$ _ pewien zbiór cech tych obiektów,

f – funkcja informacyjna określona w definicji systemu informacyjnego.

4) Zbiorem przybliżonym nazywamy parę zbiorów $\{\underline{R}X, \overline{R}X\}$, gdzie:

- zbiór $\underline{R}X$ jest R – dolną aproksymacją zbioru $X \subseteq U$,
- zbiór $\overline{R}X$ jest R – górną aproksymacją zbioru $X \subseteq U$.

Zbiory $\underline{R}X$ i $\overline{R}X$ są określone następująco:

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\}, \quad (2)$$

$$\overline{R}X = \{x \in U : [x]_R \cap X \neq \phi\}. \quad (3)$$

Zbiór obiektów, które:

- należą do aproksymowanego zbioru X, to R – dolna aproksymacja,
- należą do zbioru X oraz jednoznacznie ani nie należą, ani należą, to R – górna aproksymacja,
- nie należą do R – górnej aproksymacji i nie należą do zbioru X [9].

Za J. Stefanowskim przyjęto definicje: *Współczynnik dokładności przybliżenia* definiowany jest jako:

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}, \quad (4)$$

gdzie:

$|X|$ – liczność niepustego zbioru X.

*Klasa decyzyjna***.** Niech będzie dana tablica decyzyjna $DT = (U, A \cup \{d\})$, gdzie zbiór V_d wartości atrybutu decyzyjnego d jest równy $\{v_d^1, ..., v_d^r\}$. Atrybut decyzyjny d definiuje podział $CLASS_{DT}(d) = \{X_1, X_2, ..., X_r\}$ zbioru obiektów U, gdzie:

$X_k = \{x \in U : f(d,x) = v_d^k\}$ dla $1 \leq k \leq r$. Podział $CLASS_{DT}(d)$ jest nazywany klasyfikacją obiektów w DT, a X_k jest k-tą klasą decyzyjną.

5) Jakość przybliżenia klasyfikacji definiujemy jako:

$$\gamma(B,d) = \frac{|POS_B(d)|}{|U|} = \frac{\sum_{k=1}^{r} \underline{B}(X_k)}{|U|}, \quad (5)$$

gdzie: $POS_B(d)$ – B – pozytywny obszar klasyfikacji.

Za jakość przybliżenia klasyfikacji odpowiada jądro. Zawiera ono te atrybuty, których usunięcie z tablicy decyzyjnej powoduje spadek jakości przybliżenia.

[9] R.K. Nowicki, wyd. cyt., s. 10-11.

6) Redukt[10] jest minimalnym podzbiorem atrybutów, który umożliwia taką samą klasyfikację elementów uniwersum, jak i całego zbioru atrybutów. Inaczej mówiąc, atrybuty, które nie należą do reduktu, są zbędne w odniesieniu do klasyfikacji elementów uniwersum.

Reduktem[11] w tablicy decyzyjnej $IT = (U, A)$ nazywamy podzbiór $B \subseteq A$ wtedy i tylko wtedy, gdy:

dla każdego $x \in U$ zachodzi $I_B(x) = I_A(x)$, dla każdego podzbioru $C \subset B$ pierwszy warunek jest niespełniony.

W przypadku tablicy decyzyjnej $DT = (U, A \cup \{d\})$ można podać inną definicję reduktu. Podzbiór $B \subseteq A$ nazywany jest reduktem względnym w tablicy DT wtedy i tylko wtedy, gdy: $\gamma(B, x) = \gamma(A, x)$.

Pojęcie reduktu (wykorzystywanego do wykrywania nadmiarowości danych w tablicy informacyjnej) definiowane jest tylko dla obiektów obecnych w tablicy decyzyjnej (nie są uwzględniane żadne obiekty poza nią). W tablicy może się pojawić więcej niż jeden redukt, a ich część wspólna określana jest jądrem (*core*). Poszukiwanie wszystkich reduktów to problem o złożoności wykładniczej. *Jądro zbioru atrybutów* A jako części wspólnej reduktów zbioru A definiuje się $CORE(A) = \cap RED(A)$. Redukty mają za zadanie edukcję nadmiarowej informacji i ekstrakcję reguł decyzyjnych. Ich właściwości omówiono dokładniej m.in. w pracy Z. Pawlaka[12].

7) Reguły decyzyjne R dla tablicy decyzyjnej $DT = (U, A \cup \{d\})$ mają następującą postać:

jeżeli P , to $\quad (f(d, x) = v_d^j)$, (6)

gdzie:

P – koniunkcja warunków $(f(a_i, x) = v_{ai})$,

d – atrybut decyzyjny przyjmujący wartości z dziedziny V_d .

J. Stefanowski w swojej rozprawie wyróżnia dwa typy reguł: generowane z przykładów należących do dolnych przybliżeń klas decyzyjnych – tzw. *reguły możliwe* (*possibile*) i do górnych przybliżeń – tzw. *reguły pewne* (*certain*). Ich zapis przedstawił przy użyciu poniższych definicji.

Niech $[w_i]$ oznacza zbiór obiektów spełniających wyrażenie w_i, a więc $\{x \in U : f(a_i, x) = v_{ai}\}$ Zauważmy, że $[w_i \wedge w_j] = [w_i] \cap [w_j]$, stąd dla uproszczenia stosowany będzie zapis: jeżeli P , to Q , gdzie: $Q = (f(d, x) = v_d^j)$.

Mówimy, że obiekt *x* jest dopasowany do części warunkowej reguły *r* lub reguła pokrywa (*covers*) obiekt, jeśli $x \in [P]$, a wspiera (*supports*) regułę *r*, jeśli $x \in [P \wedge Q]$.

Regułą pewną w tablicy DT jest reguła decyzyjna *r* spełniająca warunek $[P] \subseteq [Q]$, a *regułą możliwą* reguła spełniająca warunek $[P] \subseteq \overline{A}(X_j)$ przy założeniu, że $X_j = [Q]$[13].

Minimalne reguły decyzyjne to takie, które mają minimalną liczbę argumentów (warunkowych), czyli zostały wygenerowane dla najkrótszych reduktów.

Zbiory przybliżone jako metoda radzenia sobie z niespójnymi (lub niekompletnymi) danymi, jest jedną z najszybciej rozwijających się metod sztucznej inteligencji. Jak wskazuje prof. Tadeusiewicz „w prostocie tej metody jest bardzo duży potencjał rozwoju […] i już od trzydziestu lat naukowcy na całym świecie teorię i praktykę zbiorów przybliżonych rozszerzają, wzbogacają i stosują do różnych celów - i wydaje się, że jeszcze na długo im starczy tej inspiracji"[14]. Jednak pomimo szeregu zastosowań teorii zbiorów przybliżonych (m.in. w matematyce, informatyce, finansach, marketingu, lingwistyce, farmakologii, naukach o ziemi, naukach społecznych, biznesie, przemyśle, medycynie[15]) i kilku tysięcy publikacji (badań i opracowań) na jej temat (tu jako przykłady można

[10] Z. Pawlak, Rough set Elements. [w:] L. Polkowski, A. Skowron (eds.), Rough Sets in Knowledge Discovery Heidelberg, Physica-Verlag 1998, s. 10-30.
[11] J. Stefanowski, Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy, Rozprawa habilitacyjna, Wydawnictwo Politechniki Poznańskiej, Seria Rozprawy 361, Poznań 2001.
[12] Z. Pawlak, Zbiory przybliżone…
[13] J. Stefanowski, Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy, Rozprawa habilitacyjna, Wydawnictwo Politechniki Poznańskiej, Seria Rozprawy 361, Poznań 2001.
[14] R. Tadeusiewicz, http://ryszardtadeusiewicz.natemat.pl/150285,polska-wyspa-w-archipelagu-sztucznej-inteligencji-zbiory-przyblizone, (06.04.2016r).
[15] W. Podraza, Modyfikacja zastosowania teorii zbiorów przybliżonych w medycynie w celu ograniczenia błędów przypadkowych, PAK 8/2000.

podać prace: M. Libery i W. Waligóry[16], M. Klimkiewicza i K. Moczulskiej[17], M. Łatuszyńskiej, A. Wawrzyniak, B. Wąsikowskiej, E. Galindo, J. Cruz Sandovalbrak[18], B. Kostek[19], E. Stępień[20], A. Wawrzyniak i B.Wąsikowskiej[21], A. Mrózek i L. Płonki[22], B. Ludwiszewskiego, K. Redlarskiego i J. Wachowicza[23], A. Piegata[24]) to w literaturze przedmiotu nie zauważa się (brak ich lub nie zostały nigdy opublikowane) przykładów zastosowań w obszarze kultury. Ponadto, analizując wyniki badań uzyskane w przytoczonych przykładach, nasuwa się wniosek, iż warto byłoby je dodatkowo poddać krytycznej ocenie eksperta (niektórzy autorzy to sugerują, niemniej jednak w żadnej z pozycji nie znalazły się takie interpretacje).

Na otrzymanej z Teatru próbie danych i w oparciu o przedstawioną metodologię przeprowadzono badania, których rezultaty przedstawiono w kolejnej części pracy.

Raport z przeprowadzonych badań

Przeprowadzona analiza miała na celu opracowanie algorytmu decyzyjnego pozwalającego przewidzieć liczbę widzów, którzy odwiedzą Teatr w zależności od tego, jaki typ wydarzenia będzie się odbywać, czy będzie to premiera, czy stała pozycja w repertuarze. Informacja o typie sprzedanego biletu miała również pozwolić na oszacowanie przewidywanych wpływów z wydarzenia.

Badaniu poddano półroczny repertuar instytucji na podstawie raportów generalnych ze sprzedaży (dla zbiorczej liczby spektakli).

W badaniach przyjęto następujące:

1) atrybuty warunkowe:

Atrybut *Typ wydarzenia (q1)*

q1	Etykieta lingwistyczna
1	Operetka
2	Opera
3	Balet
4	Spektakl gościnny
5	Spektakl dla dzieci
6	Koncert
7	Musical

Atrybut *Premiera (q2)*

q2	Etykieta lingwistyczna
1	Tak
2	Nie

Atrybut *Typ biletu (q3)*

q3	Etykieta lingwistyczna
1	Familijny
2	Normalny
3	Nota
4	Zaproszenie
5	Pracowniczy
6	Zbiorowy

2) atrybuty decyzyjne:

Liczba sprzedanych biletów

d1	Etykieta lingwistyczna
1	0-49 biletów (mało)
2	50-109 biletów (średnia ilość)
3	110 i więcej biletów (dużo)

[16] M. Libera, W. Waligóra, Zastosowanie teorii zbiorów przybliżonych do oceny wpływu właściwości materiałowych elementów ze stali 100Cr6 na ich powierzchniową trwałość zmęczeniową, Komisja Budowy Maszyn PAN, vol. 30 nr 2. Poznań 2010.

[17] M. Klimkiewicz, K. Moczulska, Zastosowanie zbiorów przybliżonych do analizy satysfakcji klientów serwisu pojazdu, Inżynieria Rolnicza 2008; 1(99): 165-172.

[18] M. Łatuszyńska, A. Wawrzyniak, B. Wąsikowska, E. Galindo, J. Cruz Sandoval, Teoria zbiorów przybliżonych w wykrywaniu reguł zachowań zakupowych kobiet i mężczyzn podczas kupowania telefonów komórkowych, Studia Informatica nr 35/2014. s. 65-76.

[19] B. Kostek, Zastosowania metod inteligentnych w akustyce, Modelowanie inżynierskie nr 32. s. 273-280. Gliwice 2006.

[20] E. Stępień, Analiza opinii o dostępności książki w bibliotece metodami zbiorów przybliżonych: studium metodologiczne, PTINT Praktyka i Teoria Informacji Naukowej i Technicznej. Tom 8, nr 2 (30), s. 7-15. Warszawa 2000.

[21] A. Wawrzyniak, B. Wąsikowska, Badanie preferencji wyborczych mieszkańców województwa zachodniopomorskiego przy użyciu zbiorów przybliżonych, Studies & Proceedings of Polish Association for Knowledge Management Nr 56, 2011. s.218-229.

[22] A. Mrózek, L. Płonka, Analiza danych metodą zbiorów przybliżonych. Zastosowania w ekonomii, medycynie i sterowaniu, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.

[23] B. Ludwiszewski, K. Redlarski, J.Wachowicz, Wykorzystanie zbiorów przybliżonych w analizie Kansei. Proceedings of the Conference: Interfejs użytkownika - Kansei w praktyce, Wydawnictwo PJWSTK. Warszawa 2010, s. 21–27.

[24] A. Piegat, Zastosowania zbiorów przybliżonych w ekonomii, wykłady - rękopis powielony, Uniwersytet Szczeciński, Szczecin 2005.

Definicje te pozwoliły opracować tabelę informacyjną (której fragment przedstawia tab. 1) dotyczącą liczby widzów odwiedzających instytucję kultury.

Lp.	Typ wydarzenia	Premiera	Bilet	Liczba sprzedanych biletów	Liczbasprzedanych biletów na spektakl
1.	musical	tak	normalny	50-109 (średnio)	97
2.	musical	tak	nota	110-więcej (dużo)	152
3.	musical	tak	pracowniczy	0-49 (mało)	22
4.	musical	tak	zaproszenie	50-109 (średnio)	101
5.	musical	tak	zbiorowy	0-49 (mało)	37
6.	spektakl dla dzieci	nie	familijny	50-109 (średnio)	57
7.	spektakl dla dzieci	nie	normalny	50-109 (średnio)	94
8.	spektakl dla dzieci	nie	nota	50-109 (średnio)	106
...	..				
88.	koncert	nie	Zbiorowy	0-49 (mało)	43

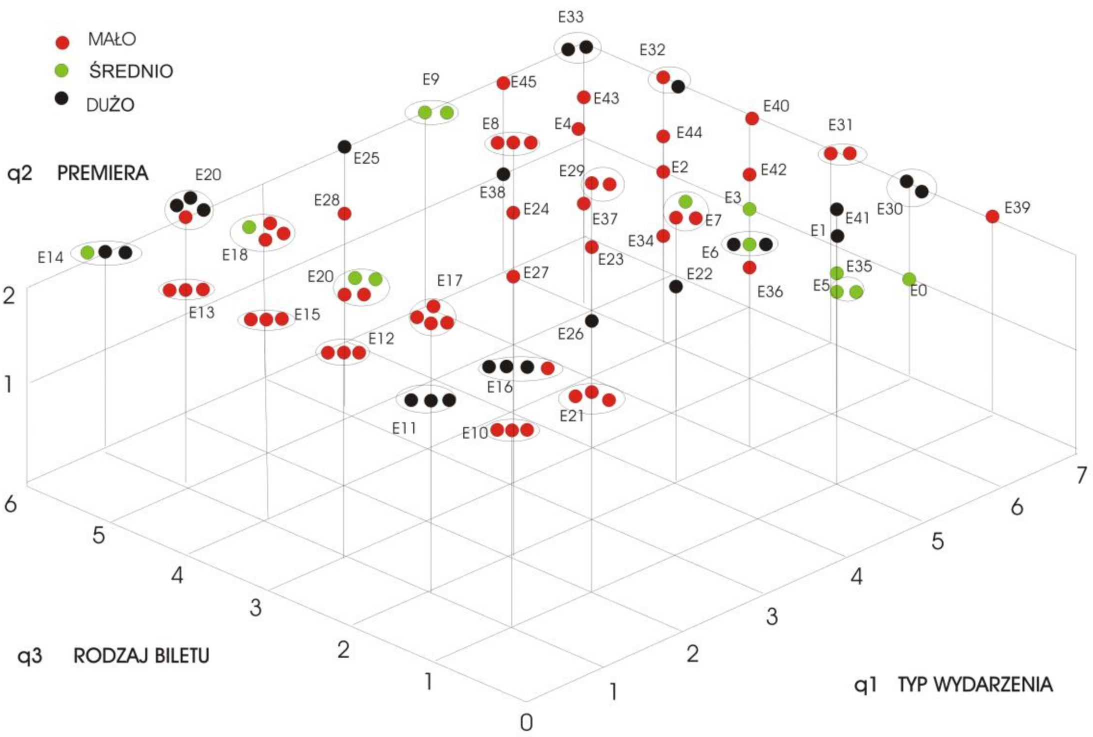
Tabela 1. Tabela informacyjna
Źródło: Opracowanie własne.

Tab. nr 1. została następnie przekształcona do postaci zakodowanej (której fragment przedstawia tab. nr 2.):

Lp.	q1	q2	q3	d_1
1.	7	1	2	2
2.	7	1	3	3
3.	7	1	5	1
4.	7	1	4	2
5.	7	1	6	1
6.	5	2	1	2
7.	5	2	2	2
8.	5	2	3	2
...				
88.	6	2	6	1

Tabela 2. Postać zakodowana danych
Źródło: Opracowanie własne.

Na podstawie przedstawionej tab. nr 2., przy pełnym zbiorze 3 atrybutów warunkowych, rodzina P* zawiera 46 elementarnych zbiorów warunkowych:
E0 = [1]; E1 = [2]; E2 = [3]; E3 = [4]; E4 = [5]; E5 = [6 42]; E6 = [7 45 60]; E7 = [8 43 61];
E8 = [9 44 62]; E9 = [10 46]; E10 = [11 37 56]; E11 = [12 36 55]; E12 = [13 38 57]; E13 = [14 39 54]; E14 = [15 41 59]; E15 = [16 40 58]; E16 = [17 27 52 68]; E17 = [18 29 48 67]; E18 = [19 30 49 66]; E19 = [20 32 51 64]; E20 = [21 31 50 65]; E21 = [22 28 53]; E22 = [23]; E23 = [24]; E24 = [25]; E25 = [26]; E26 = [33]; E27 = [34]; E28 = [35]; E29 = [47 63]; E30 = [69 78]; E31 = [70 80]; E32 = [71 81]; E33 = [72 83]; E34 = [73]; E35 = [74]; E36 = [75]; E37 = [76]; E38 = [77]; E39 = [79]; E40 = [82]; E41 = [84]; E42 = [85]; E43 = [86]; E44 = [87]; E45 = [88], których prezentację graficzną przedstawiono na rys. nr 1.



Rysunek 1. Zbiory elementarne
Źródło: Opracowanie własne.

Na podstawie zbiorów elementarnych wyróżniono 3 zbiory decyzyjne:

1) X_1 = [3 5 9 11 13 14 16 18 19 20 22 24 25 28 29 30 31 34 35 37 38 39 40 43 44 47 48 53 54 56 57 58 61 62 63 65 66 68 70 71 73 75 76 79 80 82 86 87 88]
card (X_1)= 49,

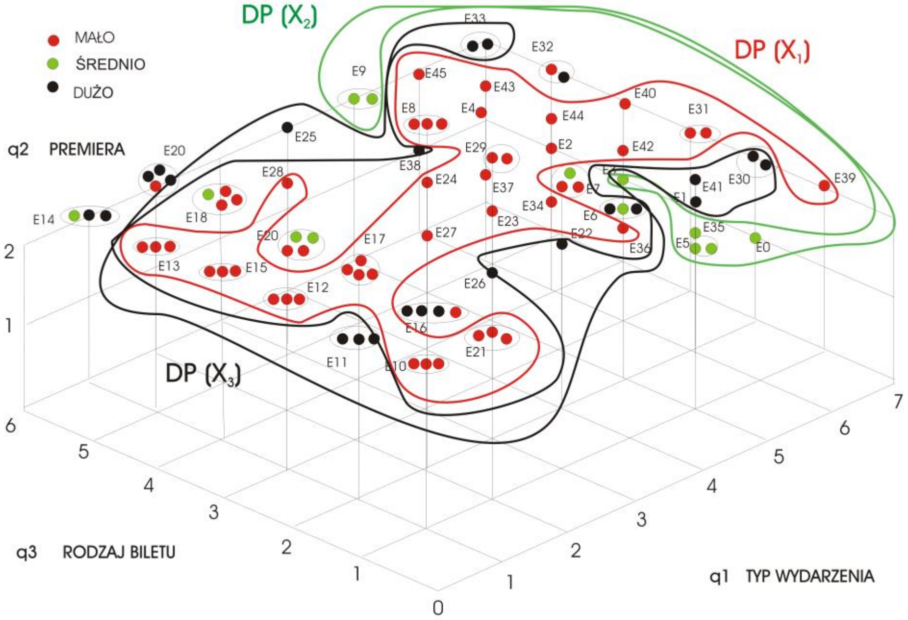
2) X_2 = [1 4 6 7 8 10 21 41 42 46 49 50 74]
card (X_2)= 13,

3) X_3 = [2 12 15 23 26 27 32 33 36 45 51 52 55 59 60 64 69 72 77 78 81 83 84]
card (X_3)= 23,

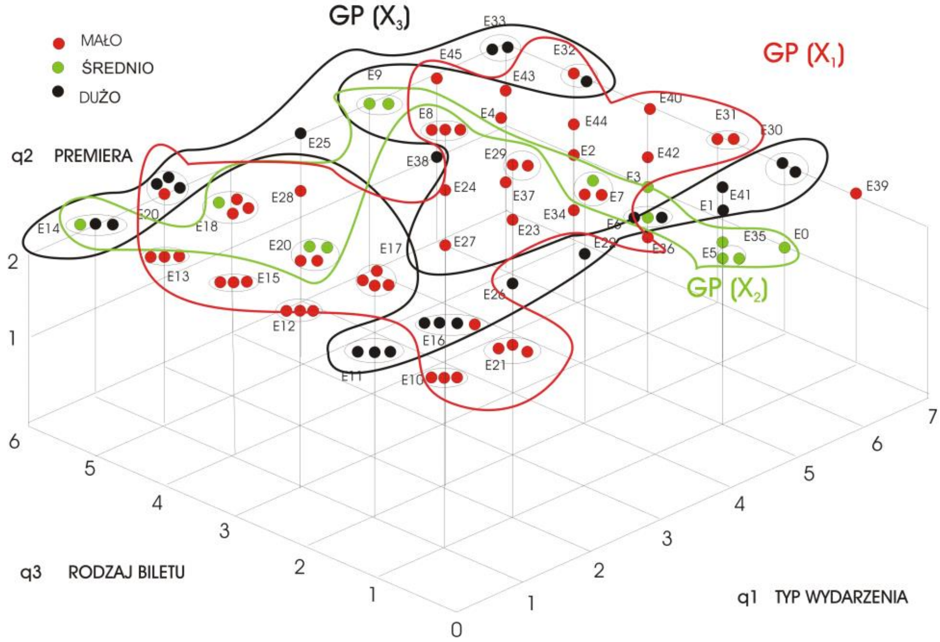
a następnie określono dolne i górne przybliżenia konceptów decyzyjnych X_1, X_2 i X_3 (rys. 2 i 3) oraz jakość i dokładność rodziny konceptów decyzyjnych X_1, X_2 i $X_{3:}$

Jakość przybliżenia rodziny konceptów decyzyjnych:

$$\gamma = \frac{pos(D^*)}{card(U)} = 61/88 = 0{,}69 = k$$



Rysunek 2. Dolne przybliżenia konceptów decyzyjnych X_1, X_2 i X_3
Źródło: Opracowanie własne.



Rysunek 3. Górne przybliżenia konceptów decyzyjnych X_1, X_2 i X_3
Źródło: Opracowanie własne.

Pozytywny obszar rodziny D_ przy atrybutach warunkowych {q1, q2, q3} określono następująco:
PosD_ = E2 ∪ E4 ∪ E8 ∪ E10∪ E12 ∪ E13 ∪ E15 ∪ E17 ∪ E21 ∪ E23 ∪ E24 ∪ E27 ∪ E28 ∪ E29 ∪ E31 ∪ E34 ∪ E36 ∪ E37 ∪ E39 ∪ E40 ∪ E42 ∪ E43 ∪ E44 ∪ E45 ∪ E0 ∪ E3 ∪ E5 ∪ E9 ∪ E35 ∪ E1 ∪ E11 ∪ E22 ∪ E25 ∪ E26 ∪ E30 ∪ E33 ∪ E38 ∪ E41
PosD_ = [3 5 9 11 13 14 16 18 22 24 25 28 29 34 35 37 38 39 40 44 47 48 53 54 56 57 58 62 63 67 70 73 75 76 79 80 82 85 86 87 88 1 4 6 10 42 46 74 2 12 23 26 33 36 55 69 72 77 78 83 84]

Obszar pozytywny rodziny konceptów decyzyjnych:

$$pos(D^*) = \sum_{i=1}^{4} card\,(\mathrm{DP}(\mathrm{X_i})) = 61$$

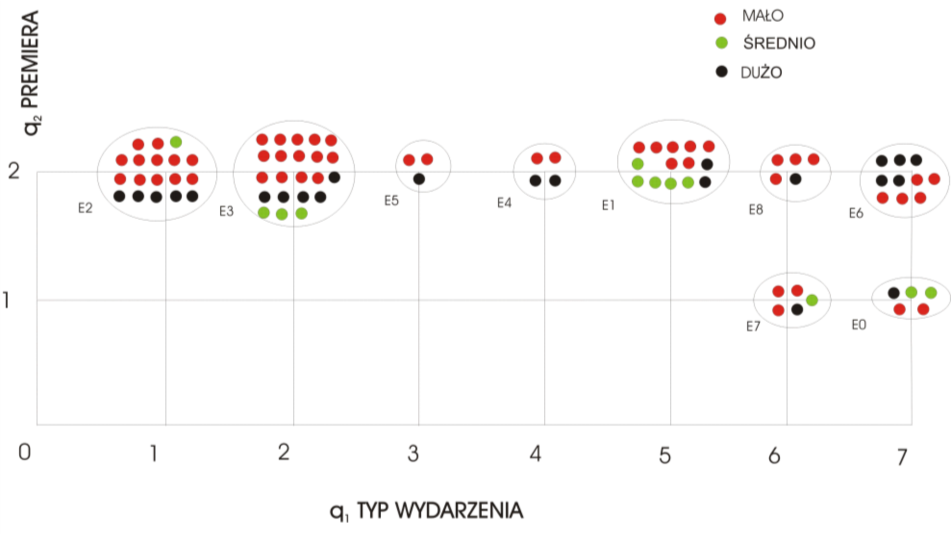
Pozytywny obszar zawiera 61 przykładów. Jakość przybliżenia rodziny D_ wynosi: (D_) = 61/88 = 0,693, co oznacza to, iż 69,3% przykładów pozwala generować reguły pewne. Dokładność przybliżenia rodziny konceptów decyzyjnych D_ wynosi:

$$\beta = \frac{pos(D^*)}{\sum_{i=1}^{4} card(GP(X_i))} =$$

= 61/111= 0,549 (D_) = 61/111 = 0, 549

Kolejnym etapem badań było wyznaczenie reduktów bezwzględnych i względnych zbioru atrybutów warunkowych oraz względnej istotności atrybutów warunkowych. Na tej postawie można było wyróżnić następujące zbiory elementarne (graficzna prezentacja w wybranej przestrzeni na rys. nr 4):

E_0 = [1 2 3 4 5]	card (E_0)= 5
E_1 = [6 7 8 9 10 42 43 44 45 46 47 60 61 62 63]	card (E_1)= 15
E_2 = [11 12 13 14 15 16 36 37 38 39 40 41 54 55 56 57 58 59]	card (E_2)= 18
E_3 = [17 18 19 20 21 22 27 28 29 30 31 32 48 49 50 51 52 53 64 65 66 67 68]	card (E_3)= 23
E_4 = [23 24 25 26]	card (E_4)= 4
E_5 = [33 34 35]	card (E_5)= 3
E_6 = [69 70 71 72 78 79 80 81 82 83]	card (E_6)= 10
E_7 = [73 74 75 76 77]	card (E_7)= 5
E_8 = [84 85 86 87 88]	card (E_8)= 5



Rys. 4. Elementarne zbiory warunkowe w przestrzeni {q1, q2}
Źródło: Opracowanie własne.

Kolejno podjęto próby redukcji atrybutów, w wyniku czego wywnioskowano, że:

1. Obszar PosD_1 nie zawiera przykładów w przestrzeni q1, q2, natomiast obszar PosD_ w przestrzeni q1, q2, q3 zawiera 29 przykładów. Oznacza to, iż atrybut q3 jest rdzeniem zbioru P, a zbiór P1 = q1, q2 nie jest reduktem względnym zbioru P = q1, q2, q3. Jakość przybliżenia rodziny D_1 wynosi (D_1) = 0, co oznacza, że nie są generowane reguły pewne. Względna istotność atrybutu q3 wynosi 1, z czego wynika, że atrybut q3 jest nieusuwalny.
2. Jakość przybliżenia rodziny D_2 wynosi (D_2) = 44.
   44= 0, 5 co oznacza, że 50% przykładów pozwala na generowanie reguł pewnych. Względna istotność atrybutu q2 wynosi 0,2785 i jest różna od zera, z czego wynika, że atrybut q2 jest nieusuwalny.
3. Obszar PosD_3 zawiera 4 przykłady w przestrzeni q2, q3, natomiast obszar PosD_ w przestrzeni q1, q2, q3 zawiera 46 przykładów. Oznacza to, iż atrybut q1 jest rdzeniem zbioru P, a zbiór P3 = q2, q3 nie jest reduktem względnym zbioru P = q1, q2, q3.

Jakość przybliżenia rodziny D_3 wynosi (D_3) = 4/80 = 0,0454 co oznacza, że 4,5% przykładów pozwala na generowanie reguł pewnych.

Względna istotność atrybutu q1 wynosi 0,9345 i jest różna od zera, z czego wynika, że atrybut q1 jest nieusuwalny.

Podsumowując krótko powyższe: ponieważ nie istnieje taki minimalny i niezależny podzbiór Pi zbioru P, który by posiadał identyczną relację nierozróżnialności (identyczną rodzinę zbiorów elementarnych Ei), zbiór P nie posiada reduktów bezwzględnych. Zbiór atrybutów warunkowych P{q1, q2, q3} nie posiada reduktów bezwględnych ani względnych, więc atrybuty warunkowe q1, q2 oraz q3 stanowią rdzeń bezwzględny i jednocześnie rdzeń względny zbioru atrybutów warunkowych.

W następnym kroku opracowano algorytm decyzyjny, wyznaczono wsparcie, pewność i siłę dla reguł nieuproszczonych i podzielono algorytm na część dobrze i źle zdefiniowaną. Część dobrze zdefiniowaną uzyskano po usunięciu reguł sprzecznych z pełnego algorytmu decyzyjnego, na podstawie zbiorów elementarnych E_i stanowiących dolne przybliżenie konceptów decyzyjnych X_i. Fragment zaprezentowano w tab. nr 3.

Reguła	q1	q2	q3	d1	Liczba przykładów	Wsparcie	Pewność	Siła
R1	7	1	2	2	1	1	1	0,01
R2	7	1	3	3	1	1	1	0,01
R3	7	1	5	1	1	1	1	0,01
R4	7	1	4	2	1	1	1	0,01
R5	7	1	6	1	1	1	1	0,01
R6	5	2	1	2	2	2	1	0,02
R9	5	2	5	1	3	3	1	0,03
R10	5	2	6	2	2	2	1	0,02
R11	1	2	1	1	3	3	1	0,03
$R_{23/12/29/38/50}$	NOT 2 OR 5	2	2	3	3	3	1	0,03
...								
$R_{51/13/24/39/30/18}$	NOT 5	2	3	1	3	3	1	0,03

Tabela 3. Zbiór wszystkich użytecznych reguł w formie zakodowanej
Źródło: Opracowanie własne.

Po uproszczeniu reguł podobnych ponownie wyliczono wsparcie, pewność i siłę reguł uproszczonego algorytmu decyzyjnego (części dobrze i źle zdefiniowanej) oraz zbadano je pod kątem ich użyteczności.

Interpretacja wyników przeprowadzonych badań i wnioski

Zastosowanie teorii zbiorów przybliżonych prof. Pawlaka pozwoliło na podstawie próby danych pomiarowych uzyskać zestaw 27 reguł charakteryzujących określony proces decyzyjny. Przeprowadzone badania potwierdziły ścisłą korelację typu wydarzenia, rodzaju sprzedanego biletu i statusu wydarzenia, jako tych, które ściśle determinują liczbę widzów.

Analiza reguł implikujących niską sprzedaż wydarzeń kulturalnych: (d1 = 1) → [(q1=7) AND (q2=1) AND (q3=5)] OR [(q1=7) AND (q2=1) AND (q3=6)] OR [(q1=7) AND (q2=2) AND (q3=1)] OR [(q1=2) AND (q2=2) AND (q3=1)] OR [(q1=5) AND (q2=2) AND (q3=4)] OR [(q1=6) AND (q2=1) AND (q3=4)] OR [(q1=6) AND (q2=1) AND (q3=3)] OR [(q1=6) AND (q2=1) AND (q3=5)] OR [(q1=4) AND (q2=2) AND (q3=4)] OR [(q1=1) AND (q2=2) AND (q3=1)] OR [(q1=1) AND (q2=2) AND (q3=4)] OR [(q1=6) AND (q2=2) AND (q3=4)] OR [(q1=6) AND (q2=2) AND (q3=6)] OR [(q1=3) AND (q2=2) AND (q3=5)] OR [(q1=5) AND (q2=2) AND (q3=5)] OR [(q1=6) AND (q2=2) AND (q3=5)] OR [(q1=1) AND (q2=2) AND (q3=5)] OR [(q2=2)] AND [(q3=3)] AND [(q1= NOT 5)]
pozwoliła na wyciągnięcie następujących wniosków:

1. Pracownicy teatru niechętnie kupują bilety na premiery musicalu, popremierowe spektakle dla dzieci, operetki i balet oraz koncerty (bez względu na to, czy jest to premiera, czy nie);
2. Rodziny rzadko wybierają popremierowe opery, operetki i musicale;

3. Bilety-zaproszenia na premierę musicalu lub koncertu, popremierowe spektakle gościnne, operetkę, spektakl dla dzieci nie cieszą się zainteresowaniem;
4. Zorganizowane grupy widzów rzadko wybierają popremierowe koncerty.

O niewielkim zainteresowaniu wymienionymi wydarzeniami świadczy również fakt, że ich udział w sprzedaży stanowi jedynie 9%.

Analiza reguł implikujących średnią liczbę sprzedanych biletów: (d1 = 2) → [(q1=6) AND (q2=1) AND (q3=2)] OR [(q1=7) AND (q2=1) AND (q3=2)] OR [(q1=7) AND (q2=1) AND (q3=4)] OR [(q1=5) AND (q2=2) AND (q3=1)] OR [(q1=5) AND (q2=2) AND (q3=6)]

Wnioski

1. Na premierowe koncerty (bilety w najwyższej cenie) chętnie przychodzą klienci indywidualni (bilet normalny);
2. Rodziny wybierają popremierowe spektakle dla dzieci;
3. Na premierę musicalu sprzedano najwięcej biletów normalnych. (Jest to ważny wynik, ponieważ bilet normalny ma najwyższą cenę i generuje największe wpływy do kasy Teatru).

Na wymienione wydarzenia sprzedaje się ok. 20% wszystkich biletów.

Analiza reguł implikujących wysoką sprzedaż: (d1 = 3) → [(q1=7) AND (q2=1) AND (q3=3)] OR [(q1=6) AND (q2=1) AND (q3=6)] OR [(q1=4) AND (q2=2) AND (q3=6)] OR [(q2=2)] AND [(q3=2)] AND [(q1=NOT 2 OR 5)] potwierdza, iż:

- pracownicy Teatru kupili najwięcej biletów na popremierowy musical;
- klienci indywidualni nie wybierali popremierowych oper i spektakli dla dzieci. Najwięcej widzów odnotowano zaś na premierze musicalu oraz na premierowych spektaklach gościnnych – wydano wówczas ponad 20% ze wszystkich wystawionych biletów na notę (wynika to z tego, iż zwyczajem Teatru jest zapraszanie na premiery gości, którym przekazywane są bezpłatne bilety).

Jako podsumowanie przeprowadzonych badań należy wskazać, iż liczba przykładów (danych przekazanych do badań przez instytucję) była daleko niewystarczająca dla tak dużej liczby możliwych reguł i nie dziwi, że wykryte reguły były słabo wsparte przykładami. W przyszłości należałoby rozszerzyć badania o dane za inne sezony artystyczne, a ostateczną interpretację wyników poddać krytycznej ocenie eksperta. Niemniej jednak cel pracy został osiągnięty, potwierdzono możliwość wykorzystania teorii zbiorów przybliżonych jako praktyczne narzędzie do analizy sprzedaży wydarzeń kulturalnych. Podjęta próba wnioskowania (nawet przy ograniczonej liczbie danych) dała już nawet na tym etapie matematyczne uzasadnienie popularności konkretnych propozycji repertuarowych (uzyskano informację które atrybuty decydują o wyższej sprzedaży) oraz wstępne rekomendacje dla dalszych planów instytucji.

The theory of rough sets in the analysis of sales of cultural events

Summary:
This article focuses on the possibilities of potential of the rough set theory to analyze sales of tickets for cultural events. An attempt was carried out for develop an algorithm enabling solve the problem of decision-making related to the choice of the best-selling repertoire items. The study also addresses the problem of assessing the reliability of the data collected in the context of compliance another isolated objects on the basis of dependence between them. The starting point for research was speculated that the number of tickets sold for a given type is associated with type of events that are offered and sales analysis have been performed to determine, among which groups of customers a given offer a enjoy (or no) interest. Carried out in this area research have unambiguously concluded, which client uses this offer. In the longer time gives the possibility adapt to customer of repertoire plans and enhance ticket sales for cultural events.

Keywords: rough sets theory, sales of tickets for cultural events