Techniki informacyjne w badaniach systemowych

Pod redakcją

Piotra Kulczyckiego
Olgierda Hryniewicza
Janusza Kacprzyka

Wydawnictwa Naukowo-Techniczne
Warszawa

Opiniodawcy:
prof. dr hab. inż. Jakub Gutenbaum
prof. dr hab. inż. Krzysztof Kuźmiński

Okładkę i strony tytułowe projektował *Paweł G. Rubaszewski*
Redaktor techniczny *Ewa Kosińska*
Korekta *Zespół*
Przygotowanie do druku *Imtex*

Wydanie książki dofinansowane przez Instytut Badań Systemowych PAN

Wydawnictwa Naukowo-Techniczne
00-048 Warszawa, ul. Mazowiecka 2/4
tel. 0-22 826 72 71, e-mail: wnt@wnt.pl
www.wnt.pl

ISBN 978-83-204-3271-8

8

Zbiory przybliżone we wspomaganiu decyzji

Zdzisław Pawlak[1], *Roman Słowiński*[1]

Streszczenie

Pojęcie zbioru stanowi podstawę współczesnej matematyki, gdyż wszystkie pojęcia matematyczne dadzą się sprowadzić do pojęcia zbioru. Niezależnie od matematyków i filozofów także inżynierowie zainteresowali się pojęciem zbioru. Okazało się, że wiele problemów praktycznych nie da się sformułować ani rozwiązać, używając klasycznego, cantorowskiego pojęcia zbioru. Prezentowana w tym rozdziale teoria zbiorów przybliżonych jest jedną z propozycji odejścia od klasycznej definicji zbioru w celu modelowania pojęć nieostrych. Z praktycznego punktu widzenia teoria ta jest nową metodą analizy danych. W szczególności podejście zbiorów przybliżonych oparte na relacji dominacji umożliwia analizę danych, w których występuje porządek preferencyjny, a atrybuty odgrywają rolę kryteriów oceny. Podejście to umożliwia przetwarzanie informacji porządkowej bez pośrednictwa jej reprezentacji ilościowej oraz prowadzi do wygenerowania reguł decyzyjnych lub drzew decyzyjnych, które tworzą model preferencji odkryty z danych. Dzięki powyższym możliwościom teoria zbiorów przybliżonych stanowi atrakcyjne podejście do wspomagania decyzji wielokryterialnych w zakresie klasyfikacji, wyboru lub porządkowania oraz w warunkach ryzyka i niepewności.

Słowa kluczowe: teoria zbiorów przybliżonych, wnioskowanie, odkrywanie wiedzy, wielokryterialne wspomaganie decyzji, teoria zbiorów przybliżonych oparta na dominacji.

[1] Instytut Informatyki Teoretycznej i Stosowanej PAN, ul. Bałtycka 5, 44-100 Gliwice; Wyższa Szkoła Informatyki Stosowanej i Zarządzania, ul. Newelska 6, 01-447 Warszawa.
[2] Instytut Informatyki Politechniki Poznańskiej, ul. Piotrowo 2, 60-965 Poznań; Instytut Badań Systemowych PAN, ul. Newelska 6, 01-447 Warszawa; e-mail: roman.slowinski@cs.put.poznan.pl

8.1. Wstęp

Pojęcie zbioru stanowi podstawę współczesnej matematyki, ponieważ wszystkie pojęcia matematyczne dadzą się do niego sprowadzić. Dlatego teoria zbiorów stworzona przez Cantora [3] jest fundamentem, na którym jest zbudowana cała matematyka.

Niezależnie od matematyków i filozofów również inżynierowie zainteresowali się pojęciem zbioru. Okazało się, że wiele problemów praktycznych nie da się sformułować ani rozwiązać przy użyciu klasycznego, cantorowskiego pojęcia zbioru.

W 1965 roku profesor Lotfi Zadeh [41], z University of California w Berkeley, zaproponował inne niż cantorowskie pojęcie zbioru. Przyjął mianowicie, że elementy mogą należeć do zbioru w pewnym stopniu, a nie w pełni, jak to ma miejsce w klasycznej teorii zbiorów. Propozycja ta znalazła bardzo wiele zastosowań i zapoczątkowała lawinę badań na temat „teorii zbiorów rozmytych" (*fuzzy set theory*), jak nazwano teorię Zadeha.

Jeszcze inną propozycję rozumienia pojęcia zbioru podał Pawlak [27], przedstawiając w 1982 roku **teorię zbiorów przybliżonych** (*rough set theory*). Z logicznego punktu widzenia teoria ta jest nowym matematycznym podejściem do pojęć nieostrych, a z praktycznego punktu widzenia jest ona nową metodą analizy danych.

W klasycznej teorii mnogości zbiór jest definiowany przez swoje elementy, przy czym nie jest tu potrzebna żadna dodatkowa wiedza o elementach uniwersum, z których tworzymy zbiory. W teorii zbiorów przybliżonych przeciwnie, zakładamy, że mamy pewne dane o elementach uniwersum i dane te są wykorzystywane w tworzeniu zbiorów. Elementy, o których mamy identyczną informację, są nierozróżnialne i tworzą tzw. **zbiory elementarne**. Stanowią one podstawę rozumowania w teorii zbiorów przybliżonych. Suma dowolnych zbiorów elementarnych jest nazywana **zbiorem definiowalnym**. Zbiory, które nie są zbiorami definiowalnymi, są nazywane **zbiorami przybliżonymi**.

Oczywiście zbiory definiowalne można jednoznacznie scharakteryzować przez własności ich elementów, natomiast zbiorów przybliżonych nie można scharakteryzować w ten sposób. Dlatego w teorii zbiorów przybliżonych zostały wprowadzone pojęcia **dolnego** i **górnego przybliżenia zbioru**, które pozwalają każdy zbiór niedefiniowalny (przybliżony) scharakteryzować za pomocą dwóch zbiorów definiowalnych – jego dolnego i górnego przybliżenia.

Dolnym przybliżeniem zbioru są wszystkie te elementy, które w świetle posiadanej wiedzy mogą być zaklasyfikowane jednoznacznie do rozważanego zbioru, zaś górnym przybliżeniem zbioru są wszystkie te elementy, których nie można wykluczyć, w świetle posiadanej wiedzy, z danego zbioru. Różnica między górnym a dolnym przybliżeniem jest nazywana **obszarem brzegowym** (**brzegiem**) zbioru. Na przykład zbiór „liczb parzystych" jest pojęciem definiowalnym (ostrym), gdyż każdą liczbę naturalną możemy jednoznacznie zaklasyfikować jako parzystą lub nieparzystą. Natomiast zbiór „zdolnych studentów" jest pojęciem przybliżonym (nieostrym), gdyż nie o każdym studencie możemy jednoznacznie twierdzić, że jest on zdolny bądź nie.

Zbiór jest przybliżony wtedy i tylko wtedy, gdy jego obszar brzegowy jest niepusty.

Z praktycznego punktu widzenia teoria zbiorów przybliżonych jest nową metodą analizy danych. Umożliwia ona

- szukanie zależności między danymi,
- redukcję danych,
- określenie istotności danych,
- generowanie reguł decyzyjnych z danych,

i wiele innych rzeczy.

Do zalet teorii zbiorów przybliżonych należą

- brak założeń odnośnie do danych (np. prawdopodobieństwa czy rozmytości),
- tolerowanie niespójności w zbiorze danych i częściowego braku danych,
- szybkie algorytmy analizy danych,
- łatwość interpretacji wyników,
- prostota matematyczna.

Metoda zbiorów przybliżonych znalazła liczne zastosowania, między innymi w

- medycynie,
- farmakologii,
- bankowości,
- lingwistyce,
- rozpoznawaniu mowy,
- ochronie środowiska,
- bazach danych.

Teoria zbiorów przybliżonych odgrywa również ważną rolę w analizie problemów decyzyjnych [28, 29, 36]. W szczególności pozwala ona na

- analizę danych, w których występuje porządek preferencyjny, a atrybuty pełnią rolę kryteriów oceny;
- ocenę ważności kryteriów;
- redukcję zbioru kryteriów;
- przetwarzanie informacji porządkowej bez pośrednictwa jej reprezentacji numerycznej;
- generowanie reguł decyzyjnych lub drzew decyzyjnych, będących modelem preferencji odkrytym z danych;
- wspomaganie decyzji wielokryterialnych dotyczących klasyfikacji, wyboru lub porządkowania bądź wspomaganie decyzji w warunkach ryzyka i niepewności.

Teoria zbiorów przybliżonych ma wiele związków z innymi dziedzinami, a zwłaszcza z

- teorią ewidencji Dempstera–Shafera,
- teorią zbiorów rozmytych,

- teorią decyzji i modelowaniem preferencji,
- metodami wnioskowania na podstawie podobieństwa do znanych przypadków,
- metodami wnioskowania bayesowskiego,
- metodami wnioskowania boolowskiego.

Mimo to może ona być rozpatrywana jako niezależna i samodzielna dyscyplina naukowa.

Teoria zbiorów przybliżonych nie jest alternatywą dla innych istniejących metod analizy danych, a raczej je uzupełnia i może być stosowana łącznie z nimi.

Na temat teorii zbiorów przybliżonych i jej zastosowań opublikowano na świecie do tej pory blisko trzy tysiące prac oraz kilkanaście książek. Wzbudziła ona spore zainteresowanie, głównie w USA, Kanadzie, Japonii, Chinach i Indiach. Prace na jej temat prowadzone są także w Europie. Również w Polsce kilka ośrodków badawczych zajmuje się tą teorią oraz jej zastosowaniami.

Do tej pory odbyło się wiele międzynarodowych konferencji na temat teorii zbiorów przybliżonych oraz jej zastosowań. Ponadto na wielu renomowanych, międzynarodowych konferencjach organizowano specjalne sesje poświęcone tej teorii. Teoria zbiorów przybliżonych wykazała swą użyteczność w wielu dziedzinach oraz wzbudziła spore zainteresowanie na świecie nie tylko wśród informatyków, ale również wśród logików i filozofów. Mimo to wymaga ona dalszych badań, w szczególności w zakresie jej podstaw matematycznych oraz możliwości zastosowań w różnych dziedzinach.

Materiał przedstawiony w tym rozdziale jest podzielony na trzy części. W części pierwszej (podrozdz. 8.2) są zawarte uwagi na temat wnioskowania z danych, z którym teoria zbiorów przybliżonych ma silne związki praktyczne. W części drugiej (podrozdz. 8.3) są podane podstawowe pojęcia teorii zbiorów przybliżonych, a w części trzeciej (podrozdz. 8.4 i 8.5) są omówione jej rozszerzenia i zastosowania w problemach decyzyjnych, które charakteryzują się istnieniem porządku preferencyjnego w danych.

8.2. Wnioskowanie z danych

Podstawowe pojęcia teorii zbiorów przybliżonych mogą być sformułowane całkowicie ogólnie, jednakże z punktu widzenia zastosowań tej teorii wygodnie jest wyrazić je za pomocą terminów analizy danych, którą można uważać za szczególny przypadek wnioskowania indukcyjnego.

Przypomnijmy, że mamy dwa główne rodzaje wnioskowania: dedukcyjne i indukcyjne. **Wnioskowanie dedukcyjne** daje narzędzia służące do wyprowadzania zdań prawdziwych z innych zdań prawdziwych. Prowadzi ono zawsze do konkluzji prawdziwych. Teoria dedukcji ma dobrze ugruntowane i powszechnie przyjęte podstawy teoretyczne. Wnioskowanie dedukcyjne jest głównym narzędziem stosowanym w rozumowaniach matematycznych i poza nią nie znalazło zastosowania.

W naukach przyrodniczych (np. w fizyce) podstawową rolę odgrywa **wnioskowanie indukcyjne**. Jego cechą charakterystyczną jest to, że nie wychodzi się tu, jak w logice dedukcyjnej, od aksjomatów wyrażających wiedzę ogólną o interesującym nas świecie, lecz od pewnych faktów częściowych o badanej rzeczywistości (przykładów), które są następnie uogólniane, tworząc wiedzę o szerszym świecie niż ten, który stanowił punkt wyjścia wnioskowania.

W przeciwieństwie do wnioskowania dedukcyjnego wnioskowanie indukcyjne nie prowadzi do wniosków prawdziwych, a jedynie do wniosków prawdopodobnych (możliwych). Również w przeciwieństwie do logiki dedukcji logika indukcji nie ma jednolitych, ogólnie przyjętych podstaw teoretycznych. Rozstrzyganie prawdziwości hipotez w logice indukcji odbywa się nie, jak w logice dedukcji, drogą formalnego rozumowania, a na podstawie eksperymentu. Fizyka jest tu najlepszą ilustracją.

Powstanie komputerów i ich nowatorskie zastosowania przyczyniły się istotnie do gwałtownego wzrostu zainteresowania wnioskowaniem indukcyjnym. Dziedzina ta rozwija się dzięki informatyce niezwykle dynamicznie. Uczenie maszynowe, odkrywanie wiedzy, wnioskowanie z danych, tworzenie systemów eksperckich to przykłady nowych kierunków we wnioskowaniu indukcyjnym. Również badania nad teorią indukcji zawdzięczają informatyce nowe impulsy. Jednakże do sytuacji, jaką mamy w logice dedukcji, jest jeszcze bardzo daleka droga. Nie widać bowiem na horyzoncie zarysu teorii indukcji mającej taki status jak teoria dedukcji.

Podsumowując, cechy charakterystyczne wyżej wymienionych rodzajów wnioskowania są następujące:

- wnioskowanie dedukcyjne
  - zastosowania: matematyka,
  - pełna teoria,
  - wnioskowanie zawsze prawdziwe,
  - weryfikacja hipotez – dowód,
- wnioskowanie indukcyjne
  - zastosowania: nauki przyrodnicze i techniczne,
  - częściowe teorie,
  - wnioski prawdopodobne (możliwe),
  - weryfikacja hipotez – eksperyment.

Tak więc analizę danych należy traktować jako szczególny przypadek wnioskowania indukcyjnego.

8.3. Zbiory przybliżone – pojęcia podstawowe

W celu lepszego i bardziej intuicyjnego zrozumienia podstaw tej teorii zacznijmy od prostego przykładu. Rozpatrzmy zbiór danych podanych w tabeli 8.1. Wiersze tabeli opisują pacjentów (obiekty), a kolumny tabeli oznaczone jako *Ból głowy*, *Ból mięśni* oraz *Temperatura* reprezentują symptomy choroby i są nazywane **atrybutami warunkowymi**. Kolumna oznaczona jako *Grypa* definiuje podział pacjentów na dwie klasy decyzyjne: chorujących na grypę (*Grypa-tak*) i niechorujących (*Grypa-nie*), i nazywana jest **atrybutem decyzyjnym**.

Tabela 8.1. Zbiór danych (przykładów) – tablica decyzyjna

Pacjent	*Ból głowy*	*Ból mięśni*	*Temperatura*	*Grypa*
1	*nie*	*tak*	*podwyższona*	*tak*
2	*tak*	*nie*	*podwyższona*	*tak*
3	*tak*	*tak*	*wysoka*	*tak*
4	*nie*	*tak*	*normalna*	*nie*
5	*tak*	*nie*	*podwyższona*	*nie*
6	*nie*	*nie*	*wysoka*	*tak*

Tabele takiego rodzaju nazywamy tablicami decyzyjnymi. Rzeczywiste tablice decyzyjne mogą zawierać dużo więcej atrybutów i obiektów. Ponadto oprócz atrybutów jakościowych mogą zawierać również atrybuty ilościowe, przyjmujące wartości liczbowe. Niektóre wartości atrybutów dla pewnych obiektów mogą być także nieznane.

Problem będący przedmiotem naszego zainteresowania może być określony następująco: znaleźć zależność między występowaniem grypy (lub jej niewystępowaniem) a symptomami opisującymi pacjentów. Inaczej mówiąc, celem jest opisanie zbioru obiektów $\{1, 2, 3, 6\}$ (lub zbioru $\{4, 5\}$) w kategoriach wartości atrybutów warunkowych.

Zauważmy, że analizowane dane są niespójne, gdyż przypadki 2 i 5 dostarczają sprzecznych informacji, tzn. obaj pacjenci są opisani tymi samymi wartościami atrybutów warunkowych, lecz są przydzieleni do różnych klas decyzyjnych. Oznacza to, że zbiór pacjentów cierpiących z powodu grypy nie może być jednoznacznie opisany za pomocą symptomów *Ból głowy*, *Ból mięśni*, *Temperatura*. Możemy jednakże opisać ten zbiór w sposób przybliżony. Opierając się na posiadanych danych, można stwierdzić, że:

- $\{1, 3, 6\}$ jest maksymalnym zbiorem pacjentów, których przydzielenie do klasy *Grypa-tak* jest **pewne**, na podstawie opisu atrybutami warunkowymi;
- $\{1, 2, 3, 5, 6\}$ jest zbiorem pacjentów, których przydzielenie do klasy pacjentów *Grypa-tak* jest **możliwe**, na podstawie opisu atrybutami warunkowymi;
- $\{2, 5\}$ jest zbiorem pacjentów, którzy nie mogą być jednoznacznie przydzieleni do klasy *Grypa-tak* ani do klasy *Grypa-nie*, ze względu na sprzeczny opis atrybutami warunkowymi.

Dwa pierwsze zbiory są przybliżeniami klasy decyzyjnej *Grypa-tak*, nazywanymi, odpowiednio, jej **dolnym** i **górnym** przybliżeniem.

Przedstawimy teraz powyższe rozważania bardziej formalnie.

Tablice danych, takie jak podana w powyższym przykładzie, są często nazywane **systemami informacyjnymi**.

System informacyjny jest czwórką (U, A, V, f), gdzie U jest niepustym i skończonym zbiorem obiektów zwanym **uniwersum**, A jest niepustym i skończonym zbiorem *atrybutów*. $V = \bigcup_{a \in A} V_a$, przy czym V_a jest **dziedziną** atrybutu $a \in A$, oraz $f : U \times A \to V$ jest **funkcją informacyjną**, taką że $\forall a \in A$, $x \in U$, $f(a, x) \in V_a$.

Z każdym podzbiorem atrybutów $P \subseteq A$ jest związana binarna relacja $I(P)$, nazywana **relacją nierozróżnialności**, zdefiniowana jako

$$I(P) = \{(x, y) \in U \times U : f(a, x) = f(a, y),\ \forall a \in P\}.$$

Jeśli $(x, y) \in I(P)$, to obiekty x i y są **nierozróżnialne** ze względu na podzbiór atrybutów P. Relacja nierozróżnialności jest relacją równoważności. Rodzinę wszystkich klas abstrakcji relacji $I(P)$, tj. podział U za pomocą P, oznaczamy $U/I(P)$. Z kolei $P(x)$ oznacza klasę abstrakcji relacji $I(P)$ zawierającą obiekt x. Klasy te nazywane są zbiorami ***P*-elementarnymi**.

Jeżeli w systemie informacyjnym wyróżniamy rozłączne zbiory atrybutów warunkowych C i atrybutów decyzyjnych D (gdzie $A = C \cup D$), to system taki jest nazywany **tablicą decyzyjną**.

Niech $S = (U, A, V, f)$ będzie systemem informacyjnym, X niepustym podzbiorem U oraz $P \subseteq A$. Celem jest opisanie zbioru X w kategoriach wartości atrybutów z P. Prowadzi to do zdefiniowania dwóch zbiorów $\underline{P}(X)$ i $\overline{P}(X)$, nazywanych, odpowiednio, ***P*-dolnym przybliżeniem i *P*-górnym przybliżeniem** zbioru X. Definicje tych zbiorów mają następujące postaci:

$$\underline{P}(X) = \{x \in U : P(x) \subseteq X\}, \qquad \overline{P}(X) = \{x \in U : P(x) \cap X \neq \emptyset\}.$$

Zbiór $BN_P(X) = \overline{P}(X) - \underline{P}(X)$ jest nazywany ***P*-brzegiem** zbioru X.

Dolne przybliżenie $\underline{P}(X)$ zbioru X jest zbiorem obiektów, które można z **pewnością** zaliczyć do X na podstawie zbioru atrybutów P, podczas gdy obiekty z $\overline{P}(X)$ mogą być tylko uznane za **być może** należące do X, w świetle atrybutów P. P-brzeg $BN_P(X)$ zawiera obiekty, których nie można jednoznacznie przydzielić do X z uwagi na sprzeczny opis w terminach atrybutów P. Z kolei obiekty z $U - \overline{P}(X)$ z pewnością nie należą do X. O zbiorze X mówimy, że jest ***P*-przybliżony**, jeśli $BN_P(X) \neq \emptyset$, w przeciwnym razie jest on ***P*-definiowalny** (dokładny).

Z własności tych widać, że przybliżenia stanowią wnętrze i domknięcie w topologii generowanej przez dane. Zauważmy też, że konstrukcja przybliżeń ma charakter **obliczeń granularnych**, gdyż operuje na blokach obiektów nierozróżnialnych przez atrybuty P, czyli na zbiorach P-elementarnych [30].

Zbiór przybliżony X może być scharakteryzowany ilościowo za pomocą **współczynnika dokładności przybliżenia**:

$$\alpha_P(X) = \frac{|\underline{P}(X)|}{|\overline{P}(X)|},$$

gdzie $|X|$ oznacza liczność zbioru X.

Przybliżenia zbiorów są podstawowymi pojęciami teorii zbiorów przybliżonych i służą do opisu nieprecyzyjnej wiedzy o interesujących nas zjawiskach. Ponadto są one używane do znajdowania zależności w danych.

Wróćmy do przykładu z tabeli 8.1. Jeżeli przyjmiemy, że atrybuty *Ból mięśni*, *Ból głowy* i *Temperatura* są atrybutami warunkowymi, a atrybut *Grypa* jest atrybutem decyzyjnym, to taka tabela jest tablicą decyzyjną. Wiersze tablicy decyzyjnej określają reguły decyzyjne, które można wyrazić w postaci wyrażeń

„Jeżeli ..., to ...". Każda reguła decyzyjna wyznacza decyzje, które muszą być podjęte, jeśli warunki podane w tablicy są spełnione.

Reguły decyzyjne są ściśle związane z przybliżeniami. Dolne przybliżenia klas decyzyjnych wyznaczają **deterministyczne reguły decyzyjne**, tj. takie reguły, które jednoznacznie wyznaczają decyzje na podstawie warunków; natomiast górne przybliżenia klas decyzyjnych wyznaczają **niedeterministyczne reguły decyzyjne**, tj. reguły, które nie wyznaczają jednoznacznie decyzji na podstawie warunków.

I tak na przykład reguła decyzyjna odpowiadająca wierszowi 3:

> Jeżeli (*Ból głowy*, *tak*) i (*Ból mięśni*, *tak*) i (*Temperatura*, *wysoka*), to (*Grypa*, *tak*)

jest deterministyczna, ponieważ warunkom występującym w tej regule odpowiada w tablicy decyzyjnej jednoznaczna decyzja (*Grypa*, *tak*); natomiast reguła decyzyjna odpowiadająca wierszowi 2:

> Jeżeli (*Ból głowy*, *tak*) i (*Ból mięśni*, *nie*) i (*Temperatura*, *podwyższona*), to (*Grypa*, *tak*)

jest niedeterministyczna, gdyż warunkom występującym w tej regule odpowiadają w tablicy decyzyjnej dwie decyzje (*Grypa*, *tak*) oraz (*Grypa*, *nie*).

Teoria zbiorów przybliżonych oferuje nieprzedstawione tutaj matematyczne narzędzia do generowania, analizy i upraszczania reguł decyzyjnych na podstawie tablic decyzyjnych (patrz np. [23, 24, 31, 37, 40, 42]).

8.4. Podejście zbiorów przybliżonych oparte na relacji dominacji

8.4.1. Wnioskowanie indukcyjne a wiedza dziedzinowa

Jednym z powodów fiaska sztucznej inteligencji w latach sześćdziesiątych XX wieku było usiłowanie opracowania ogólnych strategii rozwiązywania problemów decyzyjnych. Okazało się, że efektywne wspomaganie decyzji oparte na wiedzy wyindukowanej z danych wymaga wzięcia pod uwagę **wiedzy dziedzinowej** związanej z semantyką danych [32]. Tylko wtedy wzorce odkryte z danych mogą być poprawne i potencjalnie użyteczne. Przez **wzorce** rozumie się zależności istniejące w danych, zwane trendami, regułami, hipotezami itp. W pracy [33] wyróżniono trzy kategorie wiedzy dziedzinowej, mające istotny wpływ na wzorce odkrywane z danych:

(i) dziedziny atrybutów, tj. zbiory wartości atrybutów mające sens dla ludzkiej percepcji;
(ii) podział atrybutów na warunkowe i decyzyjne, ograniczający wzorce do związków między wartościami atrybutów warunkowych i decyzyjnych;
(iii) porządek preferencyjny w dziedzinie niektórych atrybutów i semantyczna korelacja między tymi atrybutami, narzucająca na wzorce konieczność przestrzegania zasady dominacji.

Trudno sobie wyobrazić pominięcie kategorii (i) w algorytmach odkrywania wiedzy – bez tej wiedzy dziedzinowej odkryte wzorce są wyrażone w języku wartości oryginalnych pomiarów, który jest zazwyczaj nieadekwatny do ludzkiej percepcji, na przykład pomiar temperatury ciała za pomocą dokładnego termometru i jego medyczna percepcja. Wzorce odkrywane przy znajomości wyłącznie tej wiedzy dziedzinowej nazywają się **regułami asocjacyjnymi** [1]; pokazują one silne związki między wartościami pewnych atrybutów, bez dzielenia ich z góry na warunkowe i decyzyjne.

W uczeniu maszynowym powszechnie zakłada się znajomość wiedzy dziedzinowej (i) oraz (ii) – można wtedy rozpatrywać podział zbioru obiektów w tablicy decyzyjnej na klasy decyzyjne za pomocą atrybutów decyzyjnych. Odkrywane wzorce mają wtedy postać **reguł** lub **drzew decyzyjnych** reprezentujących związki funkcyjne między atrybutami warunkowymi a decyzyjnymi [25].

Jeśli wzorce odkryte z danych mają być wykorzystane do wspomagania decyzji, to szczególne znaczenie ma uwzględnienie przy ich odkrywaniu wiedzy dziedzinowej (iii). Wyjaśnijmy to na przykładzie wystawiania ocen uczniom w szkole. Załóżmy, że wśród atrybutów opisujących wyniki uczniów są ocena z matematyki (*Mat*), ocena z fizyki (*Fiz*) oraz ocena ogólna (*OG*). Niech dziedziny tych atrybutów składają się z trzech wartości: *zła*, *średnia* i *dobra*. Ta wiedza odpowiada kategorii (i). Jeśli chodzi o wiedzę kategorii (ii), to jest ona również znana, ponieważ jasne jest, że *Mat* i *Fiz* są atrybutami warunkowymi, a *OG* jest atrybutem decyzyjnym. Porządek preferencyjny w dziedzinie wartości atrybutów jest oczywisty: ocena *dobra* jest lepsza od *średniej*, a ta jest lepsza od *złej*. Wiadomo ponadto, że atrybut *Mat* jest semantycznie skorelowany z atrybutem *OG*, tak jak *Fiz* z *OG*. Jest to wiedza dziedzinowa kategorii (iii).

Atrybuty z dziedziną uporządkowaną według preferencji są nazywane **kryteriami**, ponieważ dotyczą oceny w określonej skali preferencji. **Zwykłe atrybuty** nie mają skali preferencji, mimo że mogą przyjmować wartości liczbowe.

Semantyczna korelacja między dwoma kryteriami (odpowiadającymi warunkowi i decyzji) oznacza, że polepszenie oceny obiektu na jednym kryterium nie powinno spowodować pogorszenia wartości na drugim kryterium, jeśli wartości pozostałych atrybutów i kryteriów są niezmienione. W naszym przykładzie polepszenie oceny ucznia z *Mat* lub *Fiz*, przy niezmienionej wartości drugiej oceny, nie powinno pogarszać oceny ogólnej *OG*, lecz raczej ją polepszać. Mówiąc ogólnie, semantyczna korelacja między kryteriami warunkowymi a decyzyjnymi wymaga, żeby obiekt x **dominujący** obiekt y na wszystkich kryteriach warunkowych (tzn. x mający oceny co najmniej tak dobre jak y na wszystkich kryteriach warunkowych) powinien również **dominować** y na wszystkich kryteriach decyzyjnych (tzn. x powinien być oceniony co najmniej tak dobrze jak y na wszystkich kryteriach decyzyjnych). Zasada ta jest zwana **zasadą dominacji** (lub **zasadą Pareto**) i jest to jedyna obiektywna zasada wielokryterialnego porównywania obiektów, której racjonalności nikt nie podważa.

Jakie wzorce mogą być odkryte z danych dotyczących oceny uczniów? To pytanie wiąże się z pytaniem o wpływ wiedzy dziedzinowej (iii) na odkrywane wzorce. Odpowiedź na pierwsze pytanie jest następująca: są to reguły decyzyjne postaci „Jeżeli ..., to ...". Każda reguła decyzyjna jest określona przez **profil**

warunkowy i **profil decyzyjny**, które są wektorami wartości progowych na wybranych atrybutach i kryteriach, odpowiednio, po stronie warunkowej i decyzyjnej. Odpowiedź na drugie pytanie jest taka, że z powodu semantycznej korelacji między kryteriami warunkowymi a decyzyjnymi profile warunkowe i profile decyzyjne reguł decyzyjnych muszą spełniać zasadę dominacji. Mówimy, że jeden profil **dominuje** drugi, jeśli oba profile mają takie same wartości na zwykłych atrybutach, a wartości kryteriów pierwszego profilu są nie gorsze od wartości kryteriów drugiego profilu.

Wyjaśnijmy zasadę dominacji na przykładzie reguł dotyczących oceny uczniów. Załóżmy, że dwie reguły wyindukowane z danych szkolnych przedstawiają związki między ocenami z *Mat* i *Fiz* po stronie warunkowej i *OG* po stronie decyzyjnej:

reguła #1: jeżeli *Mat* = *średnia* i *Fiz* = *średnia*, to *OG* = *dobra*,

reguła #2: jeżeli *Mat* = *dobra* i *Fiz* = *średnia*, to *OG* = *średnia*.

Powyższe reguły nie spełniają zasady dominacji, ponieważ profil warunkowy reguły #2 dominuje profil warunkowy reguły #1, podczas gdy profil decyzyjny reguły #2 jest zdominowany przez profil decyzyjny reguły #1. Stąd te dwie reguły są niespójne z zasadą dominacji, czyli są błędne w świetle wiedzy dziedzinowej (iii).

Można by powiedzieć, z drugiej strony, że reguły te są prawdziwe, gdyż są one poparte przykładami z tablicy decyzyjnej, lecz to nie zmienia negatywnej oceny wartości wzorców popartych przez **niespójne** przykłady. Niespójność ma wiele możliwych źródeł, którymi są na przykład

- brakujące atrybuty (zwykłe lub kryteria) w opisie obiektów; być może tablica decyzyjna nie zawiera takiego kryterium jak *opinia wychowawcy* (*OW*), wyrażona w czasie zebrania rady pedagogicznej, gdy ustalano ocenę ogólną *OG*;
- niestały charakter preferencji decydenta; być może członkowie rady pedagogicznej zmienili zdanie na temat wpływu oceny z *Mat* na ocenę ogólną *OG* w trakcie rozpatrywania ocen kolejnych uczniów.

Uwzględnienie powyższych niespójności ma podstawowe znaczenie dla odkrywania wiedzy, gdyż chcemy, aby była ona poprawna i użyteczna do wspomagania decyzji. Nie można ich potraktować jako szumu bądź błędu, który należy pominąć lub zmieszać z danymi spójnymi za pomocą operacji uśredniania czy wygładzania. Niespójności te powinny zostać zidentyfikowane i ewentualnie wykorzystane do konstrukcji wzorców niepewnych.

Dodajmy, że jeśli wiedza dziedzinowa (iii) byłaby pominięta przy odkrywaniu wzorców, to nie można byłoby zidentyfikować powyższych niespójności. Rzeczywiście, nie byłoby nic błędnego w regułach #1 i #2: są one poparte przez różne przykłady rozróżnialne za pomocą rozpatrywanych atrybutów.

Jak zaznaczyliśmy we wstępie, teoria zbiorów przybliżonych pozwala scharakteryzować daną klasę decyzyjną, gdy niespójności w opisie obiektów uniemożliwiają jej jednoznaczną charakterystykę poprzez własności jej elementów. Pojęcia **dolnego** i **górnego przybliżenia zbioru** umożliwiają przedstawienie każdego zbióru niedefiniowalnego (z powodu niespójności) za pomocą dwóch zbiorów definiowalnych, które reprezentują wiedzę pewną i możliwą. Zauważmy, że w podrozdziale 8.2 przedstawiliśmy ten paradygmat w kontekście wiedzy dziedzinowej (i) + (ii). Zwróciliśmy wtedy uwagę na to, że konstrukcja przybliżeń ma charakter **granularny** ze względu na operacje na **zbiorach P-elementarnych** stanowiących bloki obiektów nierozróżnialnych za pomocą atrybutów P (**granule wiedzy**). Jest to paradygmat bardzo ogólny, który pozostaje ważny w kontekście wiedzy dziedzinowej (i) + (ii) + (iii), jednak postać granul będzie w tym kontekście inna niż w przypadku wiedzy dziedzinowej (i) + (ii).

Adaptację powyższego paradygmatu na przypadek wiedzy dziedzinowej (i) + (iii) przedstawiono w [21]. Połączenie koncepcji przybliżeń zbiorów z koncepcją granul opartych na relacji dominacji, zamiast dotychchczasowej relacji nierozróżnialności, pozwoliło na zastosowanie teorii zbiorów przybliżonych do analizy danych, w których występuje porządek preferencyjny i semantyczna korelacja między kryteriami warunkowymi i decyzyjnymi (wiedza dziedzinowa (i) + (ii) + (iii)). Powstałe podejście nosi nazwę DRSA (*Dominance-based Rough Set Approach*) [8, 10, 14, 15, 17, 35, 36, 37].

W następnym punkcie przedstawiamy koncepcję granul wiedzy zdolnych do wykorzystania wiedzy dziedzinowej (iii) w procesie odkrywania wzorców.

8.4.2. Granule wiedzy w postaci stożków dominacji

Zachowując oznaczenia z podrozdziału 8.3, przyjmiemy ponadto, że $X_C = \prod_{q=1}^{|C|} V_q$ i $X_D = \prod_{q=1}^{|D|} V_q$ oznaczają, odpowiednio, przestrzenie atrybutów warunkowych i decyzyjnych. Punkty przestrzeni X_C i X_D są wektorami możliwych ocen obiektów, odpowiednio, za pomocą atrybutów warunkowych $C = \{1, \ldots, |C|\}$ i decyzyjnych $D = \{1, \ldots, |D|\}$. Ocena obiektu x na atrybucie $q \in A$ jest oznaczona przez x_q. Relacja nierozróżnialności na U ze względu na zbiór atrybutów decyzyjnych D dokonuje podziału U na skończoną liczbę klas decyzyjnych $\boldsymbol{Cl} = \{Cl_t, t = 1, \ldots, n\}$. Każdy obiekt $x \in U$ należy do jednej i tylko jednej klasy $Cl_t \in \boldsymbol{Cl}$.

Zakładając znajomość wiedzy dziedzinowej (i) + (ii) dla każdego podzbioru $P \subseteq C$ i podzbioru $R \subseteq D$, granule wiedzy są klasami abstrakcji relacji nierozróżnialności $I(P)$ i $I(R)$, czyli ograniczonymi podzbiorami przestrzeni X_P i X_R, wynikającymi z podziału U indukowanego, odpowiednio, przez $I(P)$ i $I(R)$. W tej sytuacji celem odkrycia za pomocą wnioskowania indukcyjnego są wzorce klasyfikacji będące funkcjami wyrażającymi granule $R(x)$ za pomocą granul $P(x)$ w przestrzeni atrybutów warunkowych X_P, dla każdego $x \in U$ i $P \subseteq C$.

Gdy do wiedzy dziedzinowej (i) + (ii) dołączymy (iii), wtedy relacja nierozróżnialności nie jest w stanie stworzyć granul w X_C i X_D zdolnych do uwzględnienia porządku preferencyjnego i zasady dominacji. Aby temu zaradzić, relacja nierozróżnialności musi być zastąpiona relacją dominacji zarówno w X_P, jak i w X_R ($P \subseteq C$ i $R \subseteq D$). Załóżmy bez utraty ogólności, że wszystkie atrybuty warunkowe w C i decyzyjne w D są kryteriami oraz że C i D są skorelowane semantycznie.

Niech $\succeq_q$ będzie **relacją słabej preferencji** na U (zwaną też **relacją przewyższania**) reprezentującą preferencję na zbiorze obiektów ze względu na kryterium $q \in \{C \cup D\}$; $x_q \succeq_q y_q$ oznacza, że „x_q jest co najmniej tak dobry jak y_q ze względu na kryterium q".

Z drugiej strony mówimy, że x **dominuje** y ze względu na $P \subseteq C$ (krótko, x ***P*-dominuje** y) w przestrzeni atrybutów warunkowych X_P (oznaczenie: xD_Py), jeśli $x_q \succeq_q y_q$ dla wszystkich kryteriów $q \in P$. Zakładając, bez utraty ogólności, że dziedziny wszystkich kryteriów są liczbowe, tzn. $X_q \subseteq \boldsymbol{R}$ dla każdego $q \in C$, oraz że są uporządkowane w ten sposób, iż preferencja rośnie z wartością, można powiedzieć, że xD_Py jest równoważne: $x_q \geqslant y_q$ dla wszystkich $q \in P$, $P \subseteq C$. Zauważmy, że dla każdego $x \in X_P$ zachodzi xD_Px, tzn. P-dominacja jest zwrotna. Analogiczną definicję dominacji można sformułować dla przestrzeni atrybutów decyzyjnych X_R (oznaczenie: xD_Ry), $R \subseteq D$.

Relacje dominacji xD_Py i xD_Ry ($P \subseteq C$ i $R \subseteq D$) są stwierdzeniami ukierunkowanymi, w których x jest podmiotem, a y jest obiektem odniesienia.

Jeśli $x \in X_P$ jest obiektem odniesienia, $P \subseteq C$, to można zdefiniować zbiór obiektów $y \in X_P$ P-dominujących x, zwany **zbiorem *P*-dominującym**:

$$D_P^+(x) = \{y \in U : yD_Px\}.$$

Jeśli $x \in X_P$ jest podmiotem, $P \subseteq C$, to można zdefiniować zbiór obiektów $y \in X_P$ P-zdominowanych przez x, zwany **zbiorem *P*-zdominowanym**:

$$D_P^-(x) = \{y \in U : xD_Py\}.$$

Zbiory P-dominujący $D_P^+(x)$ i P-zdominowany $D_P^-(x)$ odpowiadają **pozytywnym i negatywnym stożkom dominacji** w X_P, z początkiem w punkcie x.

Jeśli chodzi o przestrzeń atrybutów decyzyjnych X_R, $R \subseteq D$, to relacja R-dominacji pozwala na zdefiniowanie zbiorów

$$Cl_R^{\geqslant}(x) = \{y \in U : yD_Rx\}, \qquad Cl_R^{\leqslant}(x) = \{y \in U : xD_Ry\}.$$

$Cl_R^{\geqslant}(x)$ i $Cl_R^{\leqslant}(x)$ odpowiadają **pozytywnym** i **negatywnym stożkom dominacji** w X_R, z początkiem w punkcie x.

W tym wypadku granule wiedzy są otwartymi zbiorami w przestrzeniach X_P i X_R, zdefiniowanymi przez stożki dominacji, odpowiednio, $D_P^+(x)$, $D_P^-(x)$ ($P \subseteq C$) i $Cl_R^{\geqslant}(x)$, $Cl_R^{\leqslant}(x)$ ($R \subseteq D$). Także w tej sytuacji celem odkrycia za pomocą wnioskowania indukcyjnego są wzorce klasyfikacji będące funkcjami wyrażającymi granule $Cl_R^{\geqslant}(x)$, $Cl_R^{\leqslant}(x)$ za pomocą granul $D_P^+(x)$, $D_P^-(x)$ w przestrzeni atrybutów warunkowych X_P, dla każdego $x \in U$ i $P \subseteq C$, $R \subseteq D$. Funkcje te będą reprezentowane przez zbiory reguł decyzyjnych lub drzewa decyzyjne o specjalnej składni.

8.4.3. Podejście zbiorów przybliżonych oparte na dominacji (DRSA)

W sytuacjach praktycznych, przy odkrywaniu wzorców z danych, zbiór D atrybutów decyzyjnych jest zazwyczaj jednoelementowy, $D = \{d\}$. Przyjmijmy to założenie, aczkolwiek nie jest ono konieczne dla DRSA. Atrybut decyzyjny d dokonuje, tak samo jak poprzednio zbiór D, podziału U na skończoną liczbę klas decyzyjnych $\boldsymbol{Cl} = \{Cl_t,\ t = 1, \ldots, n\}$. Innymi słowy, $Cl_t = \{x \in U : x_q = t\}$, $t = 1, \ldots, n$. Im wyższy numer klasy, tym lepsza klasa. Ponieważ obiekty są jednoznacznie przyporządkowane do klas, oznaczenia zbiorów $Cl_t^{\geqslant}(x)$ i $Cl_t^{\leqslant}(x)$ można uprościć do $Cl_t^{\geqslant}$ i $Cl_t^{\leqslant}$, $t = 1, \ldots, n$. Są to teraz złożenia klas decyzyjnych, odpowiednio, w górę i w dół, zdefiniowane jako

$$Cl_t^{\geqslant} = \bigcup_{s \geqslant t} Cl_s, \qquad Cl_t^{\leqslant} = \bigcup_{s \leqslant t} Cl_s, \qquad t = 1, \ldots, n,$$

co odpowiada określeniom: **„klasa co najmniej Cl_t"** i **„klasa co najwyżej Cl_t"**.

Zauważmy, że $Cl_1^{\geqslant} = U$, $Cl_n^{\leqslant} = U$ oraz dla $t = 2, \ldots, n$ mamy $Cl_t^{\geqslant} = U - Cl_{t-1}^{\leqslant}$, tzn. wszystkie obiekty, które nie należą do klasy Cl_t lub lepszej, należą do klasy Cl_{t-1} lub gorszej.

Wyjaśnijmy, jak pojęcie zbioru przybliżonego zostało uogólnione w podejściu DRSA w celu umożliwienia obliczeń granularnych na stożkach dominacji (szczegółowe wyprowadzenie można znaleźć w pracach [8, 10, 14, 15, 17, 36, 37].

Dla danego zbioru kryteriów $P \subseteq C$ zaliczenie obiektu $x \in U$ do złożenia klas w górę $Cl_t^{\geqslant}$, $t = 2, \ldots, n$, jest **niespójne z zasadą dominacji**, gdy zachodzi jedna z poniższych sytuacji:

- obiekt x należy do klasy Cl_t lub lepszej, lecz jest P-zdominowany przez obiekt y należący do klasy gorszej od Cl_t, tzn. $x \in Cl_t^{\geqslant}$, lecz $D_P^+(x) \cap Cl_{t-1}^{\leqslant} \neq \emptyset$;
- obiekt x należy do klasy gorszej od Cl_t, lecz P-dominuje obiekt y należący do klasy Cl_t lub lepszej, tzn. $x \notin Cl_t^{\geqslant}$, lecz $D_P^-(x) \cap Cl_t^{\geqslant} \neq \emptyset$.

Jeśli dla danego zbioru kryteriów $P \subseteq C$ zaliczenie $x \in U$ do $Cl_t^{\geqslant}$, $t = 2, \ldots, n$, jest niespójne z zasadą dominacji, to mówimy, że x należy do $Cl_t^{\geqslant}$ **z wątpliwością** (wątpliwość spowodowana niespójnością przykładów decyzji jest pojęciem podstawowym wykorzystywanym w różnych rozszerzeniach teorii zbiorów przybliżonych, polegających na zastąpieniu relacji nierozróżnialności ogólniejszą relacją – patrz [38]). Zatem x należy do $Cl_t^{\geqslant}$ **bez wątpliwości**, biorąc pod uwagę $P \subseteq C$, jeśli $x \in Cl_t^{\geqslant}$ i nie ma niespójności z zasadą dominacji. To oznacza, że wszystkie obiekty P-dominujące x należą do $Cl_t^{\geqslant}$, tzn. $D_P^+(x) \subseteq Cl_t^{\geqslant}$. Geometrycznie odpowiada to zawieraniu się wszystkich obiektów ze stożka pozytywnej dominacji $D_P^+(x)$, z początkiem w x, w stożku pozytywnej dominacji $Cl_t^{\geqslant}$, z początkiem w Cl_t.

Ponadto x **być może należy do** $Cl_t^{\geqslant}$, biorąc pod uwagę $P \subseteq C$, gdy zachodzi jedna z poniższych sytuacji:

- zgodnie z decyzją d obiekt x należy do $Cl_t^{\geqslant}$,
- zgodnie z decyzją d obiekt x nie należy do $Cl_t^{\geqslant}$, lecz obiekt ten jest niespójny w sensie zasady dominacji z obiektem y należącym do $Cl_t^{\geqslant}$.

Używając pojęcia wątpliwości, x być może należy do $Cl_t^{\geqslant}$, biorąc pod uwagę $P \subseteq C$, jeśli x należy do $Cl_t^{\geqslant}$ z wątpliwością lub bez. Ze względu na zwrotność relacji dominacji D_P powyższe sytuacje mogą być podsumowane następująco: biorąc pod uwagę $P \subseteq C$, x **być może należy do** klasy Cl_t lub lepszej, jeśli wśród obiektów P-zdominowanych przez x istnieje obiekt y należący do klasy Cl_t lub lepszej, tzn. $D_P^-(x) \cap Cl_t^{\geqslant} \neq \emptyset$. Geometrycznie odpowiada to niepustemu przecięciu zbioru obiektów ze stożka negatywnej dominacji $D_P^-(x)$, z początkiem w x, ze stożkiem pozytywnej dominacji $Cl_t^{\geqslant}$, z początkiem w Cl_t.

Dla $P \subseteq C$ zbiór wszystkich obiektów należących bez wątpliwości do $Cl_t^{\geqslant}$ tworzy ***P*-dolne przybliżenie** złożenia klas $Cl_t^{\geqslant}$, oznaczone przez $\underline{P}(Cl_t^{\geqslant})$; natomiast zbiór wszystkich obiektów być może należących do $Cl_t^{\geqslant}$ tworzy ***P*-górne przybliżenie** złożenia klas $Cl_t^{\geqslant}$, oznaczone przez $\overline{P}(Cl_t^{\geqslant})$:

$$\underline{P}(Cl_t^{\geqslant}) = \{x \in U : D_P^+(x) \subseteq Cl_t^{\geqslant}\},$$
$$\overline{P}(Cl_t^{\geqslant}) = \{x \in U : D_P^-(x) \cap Cl_t^{\geqslant} \neq \emptyset\}, \qquad t = 2, \ldots, n.$$

Analogicznie można zdefiniować ***P*-dolne przybliżenie** i ***P*-górne przybliżenie** złożenia klas $Cl_t^{\leqslant}$:

$$\underline{P}(Cl_t^{\leqslant}) = \{x \in U : D_P^-(x) \subseteq Cl_t^{\leqslant}\},$$
$$\overline{P}(Cl_t^{\leqslant}) = \{x \in U : D_P^+(x) \cap Cl_t^{\leqslant} \neq \emptyset\}, \qquad t = 1, \ldots, n-1.$$

Przybliżenia P-dolne i P-górne zdefiniowane powyżej mają następującą własność **zawierania** dla każdego $P \subseteq C$, charakterystyczną dla teorii zbiorów przybliżonych:

$$\underline{P}(Cl_t^{\geqslant}) \subseteq Cl_t^{\geqslant} \subseteq \overline{P}(Cl_t^{\geqslant}), \qquad t = 2, \ldots, n;$$
$$\overline{P}(Cl_t^{\leqslant}) \subseteq Cl_t^{\leqslant} \subseteq \overline{P}(Cl_t^{\leqslant}), \qquad t = 1, \ldots, n-1.$$

Wszystkie obiekty należące do $Cl_t^{\geqslant}$ lub $Cl_t^{\leqslant}$ z wątpliwością wynikającą z niespójności z zasadą dominacji tworzą ***P*-brzegi** złożenia klas $Cl_t^{\geqslant}$ lub $Cl_t^{\leqslant}$ oznaczone, odpowiednio, przez $BN_P(Cl_t^{\geqslant})$ i $BN_P(Cl_t^{\leqslant})$. Mogą być one wyrażone za pomocą przybliżeń dolnych i górnych w następujący sposób:

$$BN_P(Cl_t^{\geqslant}) = \overline{P}(Cl_t^{\geqslant}) - \underline{P}(Cl_t^{\geqslant}), \qquad t = 2, \ldots, n,$$
$$BN_P(Cl_t^{\leqslant}) = \overline{P}(Cl_t^{\leqslant}) - \underline{P}(Cl_t^{\leqslant}), \qquad t = 1, \ldots, n-1.$$

P-dolne i P-górne przybliżenia złożeń klas $Cl_t^{\geqslant}$ i $Cl_t^{\leqslant}$ mają istotną własność **komplementarności**, również charakterystyczną dla podstawowej teorii zbiorów przybliżonych. Mówi ona, że jeśli obiekt x należy bez wątpliwości do klasy Cl_t lub lepszej, to niemożliwe jest, by mógł należeć do klasy Cl_{t-1} lub gorszej, tzn. $\underline{P}(Cl_t^{\geqslant}) = U - \overline{P}(Cl_{t-1}^{\leqslant})$, $t = 2, \ldots, n$.

Dzięki własności komplementarności $BN_P(Cl_t^{\geqslant}) = BN_P(Cl_{t-1}^{\leqslant})$, dla $t = 2, \ldots, n$, co oznacza, że jeśli obiekt x należy z wątpliwością do klasy Cl_t lub lepszej, to również należy z wątpliwością do klasy Cl_{t-1} lub gorszej.

Z punktu widzenia odkrywania wiedzy P-dolne przybliżenia złożeń klas reprezentują **wiedzę pewną**, podczas gdy P-górne przybliżenia reprezentują **wiedzę możliwą**, a P-brzegi reprezentują **wiedzę wątpliwą** dostarczaną przez kryteria $P \subseteq C$.

Powyższe definicje przybliżeń są oparte na ścisłym zastosowaniu zasady dominacji. Tymczasem w praktycznej analizie danych warto czasem osłabić tę zasadę i zaakceptować ograniczoną liczbę niespójnych obiektów w dolnych przybliżeniach, szczególnie gdy zbiór danych jest duży. Tak rozszerzona wersja podejścia DRSA jest nazywana VC-DRSA (*Variable-Consistency DRSA*) [20]. Poniżej wyjaśniamy, na czym to rozszerzenie polega.

Dla każdego $P \subseteq C$ mówimy, że $x \in U$ należy **bez wątpliwości na poziomie spójności** $l \in (0, 1]$ do złożenia klas $Cl_t^{\geqslant}$, jeśli $x \in Cl_t^{\geqslant}$ i co najmniej $l * 100\%$ wszystkich obiektów $y \in U$ P-dominujących x również należy do $Cl_t^{\geqslant}$, tzn.

$$\frac{|D_P^+(x) \cap Cl_t^{\geqslant}|}{|D_P^+(x)|} \geqslant l, \qquad t = 2, \ldots, n.$$

Parametr l jest nazywany **poziomem spójności**, ponieważ można nim kontrolować stopień spójności między obiektami należącymi do $Cl_t^{\geqslant}$ bez wątpliwości. Innymi słowy, jeśli $l < 1$, to co najwyżej $(1 - l) * 100\%$ wszystkich obiektów $y \in U$ P-dominujących x nie należy do $Cl_t^{\geqslant}$ i tym samym przeciwstawia się zaliczeniu x do $Cl_t^{\geqslant}$.

Analogicznie, dla każdego $P \subseteq C$ mówimy, że $x \in U$ należy **bez wątpliwości na poziomie spójności** $l \in (0, 1]$ do złożenia klas $Cl_t^{\leqslant}$, jeśli $x \in Cl_t^{\leqslant}$ i co najmniej $l * 100\%$ wszystkich obiektów $y \in U$ P-zdominowanych przez x również należy do $Cl_t^{\leqslant}$, tzn.

$$\frac{|D_P^-(x) \cap Cl_t^{\leqslant}|}{|D_P^-(x)|} \geqslant l, \qquad t = 1, \ldots, n-1.$$

Zatem dla każdego $P \subseteq C$ każdy obiekt $x \in U$ jest **wątpliwy** lub **niewątpliwy na poziomie spójności** l ze względu na przynależność do złożenia klas w górę $Cl_t^{\geqslant}$ $(t = 2, \ldots, n)$ lub w dół $Cl_t^{\leqslant}$ $(t = 1, \ldots, n-1)$.

Pojęcie niewątpliwego obiektu na poziomie spójności l prowadzi w naturalny sposób do nowej definicji ***P*-dolnych przybliżeń** złożeń klas decyzyjnych $Cl_t^{\geqslant}$ i $Cl_t^{\leqslant}$:

$$\underline{P}^l(Cl_t^{\geqslant}) = \{x \in Cl_t^{\geqslant} : \frac{|D_P^+(x) \cap Cl_t^{\geqslant}|}{|D_P^+(x)|} \geqslant l\}, \qquad t = 2, \ldots, n,$$

$$\underline{P}^l(Cl_t^{\leqslant}) = \{x \in Cl_t^{\leqslant} : \frac{|D_P^-(x) \cap Cl_t^{\leqslant}|}{|D_P^-(x)|} \geqslant l\}, \qquad t = 1, \ldots, n-1.$$

Dla danego $P \subseteq C$ i poziomu spójności l możemy również zdefiniować **P-górne przybliżenia** $Cl_t^{\geqslant}$ i $Cl_t^{\leqslant}$, oznaczone, odpowiednio, przez $\overline{P}^l(Cl_t^{\geqslant})$ i $\overline{P}^l(Cl_t^{\leqslant})$, wykorzystując komplementarność $\underline{P}^l(Cl_{t-1}^{\leqslant})$ i $\underline{P}^l(Cl_{t+1}^{\geqslant})$ względem U:

$$\overline{P}^l(Cl_t^{\geqslant}) = U - \underline{P}^l(Cl_{t-1}^{\leqslant}), \qquad t = 2, \ldots, n,$$

$$\overline{P}^l(Cl_t^{\leqslant}) = U - \underline{P}^l(Cl_{t+1}^{\geqslant}), \qquad t = 1, \ldots, n-1.$$

$\overline{P}^l(Cl_t^{\geqslant})$ jest zbiorem obiektów, które należą do złożenia klas $Cl_t^{\geqslant}$ z **możliwą wątpliwością na poziomie spójności** $l \in (0, 1]$. Analogicznie, $\overline{P}^l(Cl_t^{\leqslant})$ jest zbiorem obiektów, które należą do złożenia klas $Cl_t^{\leqslant}$ z **możliwą wątpliwością na poziomie spójności** $l \in (0, 1]$. **P-brzegi** (obszary P-wątpliwe) złożeń klas $Cl_t^{\geqslant}$ i $Cl_t^{\leqslant}$ są zdefiniowane następująco:

$$BN_P^l(Cl_t^{\geqslant}) = \overline{P}^l(Cl_t^{\geqslant}) - \underline{P}^l(Cl_t^{\geqslant}), \qquad t = 2, \ldots, n,$$

$$BN_P^l(Cl_t^{\leqslant}) = \overline{P}^l(Cl_t^{\leqslant}) - \underline{P}^l(Cl_t^{\leqslant}), \qquad t = 1, \ldots, n-1.$$

Wariant **zmiennej spójności** w podejściu DRSA stwarza pewną elastyczność w zaliczaniu obiektów do dolnych i górnych przybliżeń złożeń klas. Można wykazać, że dla $0 < l' < l \leqslant 1$ oraz $t = 2, \ldots, n$

$$\underline{P}^l(Cl_t^{\geqslant}) \subseteq \underline{P}^{l'}(Cl_t^{\geqslant}) \quad \text{i} \quad \overline{P}^{l'}(Cl_t^{\geqslant}) \subseteq \overline{P}^l(Cl_t^{\geqslant}).$$

Dla każdego zbioru $P \subseteq C$ obiekty spójne w sensie zasady dominacji ze wszystkimi złożeniami klas w górę i w dół nazywamy **zaklasyfikowanymi P-poprawnie**. Dla $P \subseteq C$ **jakość przybliżenia klasyfikacji $\boldsymbol{Cl}$** za pomocą kryteriów P jest zdefiniowana jako stosunek liczby obiektów zaklasyfikowanych P-poprawnie do liczby wszystkich obiektów w uniwersum U. Ponieważ obiekty zaklasyfikowane P-poprawnie to takie, które nie należą do żadnego P-brzegu złożeń klas $Cl_t^{\geqslant}$ i $Cl_t^{\leqslant}$, więc jakość przybliżenia klasyfikacji $\boldsymbol{Cl}$ przez zbiór kryteriów P, można zapisać jako

$$\gamma_P(\boldsymbol{Cl}) = \frac{\left|\left(U - \left(\bigcup_{t\in\{2,\ldots,n\}} BN_P(Cl_t^{\leqslant})\right) \cup \left(\bigcup_{t\in\{1,\ldots,n-1\}} BN_P(Cl_t^{\geqslant})\right)\right)\right|}{|U|} =$$

$$= \frac{\left|\left(U - \left(\bigcup_{t\in\{2,\ldots,n\}} BN_P(Cl_t^{\geqslant})\right)\right)\right|}{|U|}.$$

$\gamma_P(\boldsymbol{Cl})$ można traktować jak miarę jakości wiedzy o klasyfikacji $\boldsymbol{Cl}$, którą można wyindukować z tablicy decyzyjnej zawierającej kryteria P.

Każdy nieredukowalny podzbiór $P \subseteq C$, taki że $\gamma_P(\boldsymbol{Cl}) = \gamma_C(\boldsymbol{Cl})$, jest nazywany **reduktem** klasyfikacji $\boldsymbol{Cl}$ i oznaczany przez $RED_{\boldsymbol{Cl}}$. Zauważmy, że dana tablica decyzyjna może mieć wiele reduktów. Przekrój wszystkich reduktów jest nazywany **rdzeniem** i oznaczany przez $CORE_{Cl}$. Kryteria należące do rdzenia $CORE_{Cl}$ nie mogą być usunięte z tablicy decyzyjnej bez konsekwencji utraty jakości wyindukowanej wiedzy o klasyfikacji $\boldsymbol{Cl}$. Oznacza to, że w zbiorze C znajdują się trzy rodzaje kryteriów:

- **niezbędne** kryteria zawarte w rdzeniu,
- **wymienne** kryteria zawarte w pewnych reduktach, lecz nie w rdzeniu,
- **nadmiarowe** kryteria, które nie są ani niezbędne, ani wymienne, czyli takie, które nie znalazły się w żadnym redukcie.

Dodajmy, że w ogólnym przypadku, gdy w zbiorze C są zwykłe atrybuty i kryteria, redukty są najmniejszymi ze względu na zawieranie zbiorami atrybutów i kryteriów, zdolnymi wyrazić całą istotną wiedzę zawartą w tablicy decyzyjnej – ta wiedza jest istotna dla wyjaśnienia wzorców ukrytych w danych, niekoniecznie zaś dla predykcji.

W pracy [12] zwrócono uwagę na fakt, że jakość przybliżenia klasyfikacjami ma własności funkcji zbioru (atrybutów), zwanej **miarą rozmytą** (*fuzzy measure*). Z tego powodu można ją wykorzystać do obliczenia wskaźników mówiących o istotności poszczególnych atrybutów/kryteriów, a także o sile interakcji między nimi. Użytecznymi wskaźnikami są na przykład wskaźniki wartości i interakcji Shapleya i Banzhafa, wskaźniki interakcji Murofushi-Sonedy i Roubensa oraz wskaźnik Möbiusa. Wszystkie one mogą pomóc w ocenie wartości informacyjnej atrybutów i kryteriów oraz w wyborze „najlepszego" reduktu.

8.4.4. Indukcja wzorców klasyfikacji w postaci reguł decyzyjnych

Zdefiniowane w poprzednim punkcie przybliżenia złożeń klas decyzyjnych „w górę" i „w dół" są punktem wyjścia dla indukcyjnego generowania reguł decyzyjnych opartych na dominacji. Dla danego złożenia klas $Cl_t^{\geqslant}$ reguły decyzyjne indukowane przy założeniu, że obiekty należące do dolnego przybliżenia $\underline{P}(Cl_s^{\geqslant})$ są pozytywne, a wszystkie inne negatywne, zalecają przydział do „klasy Cl_t lub lepszej". Analogicznie, dla danego złożenia klas $Cl_s^{\leqslant}$, reguły decyzyjne indukowane przy założeniu, że obiekty należące do dolnego przybliżenia $\underline{P}(Cl_s^{\leqslant})$ są pozytywne, a wszystkie inne negatywne, zalecają przydział do „klasy Cl_s lub gorszej". Z kolei reguły decyzyjne indukowane przy założeniu, że obiekty należące do przekroju $\overline{P}(Cl_s^{\leqslant}) \cap \overline{P}(Cl_t^{\geqslant})$ są pozytywne, a wszystkie inne negatywne, zalecają przydział do ciągu dwóch lub więcej klas od Cl_s do Cl_t $(s < t)$.

W celu uwzględnienia porządku preferencyjnego zaproponowano reguły decyzyjne o składni opartej na dominacji [8, 10, 14]. W zależności od źródła obiektów pozytywnych w procedurze indukcji reguł rozpatruje się następujące typy reguł decyzyjnych:

(1) $D_{\geqslant}$**-reguły pewne**, dla których obiektami pozytywnymi są obiekty należące do $\underline{P}(Cl_t^{\geqslant})$:

Jeżeli $x_{q1} \succeq_{q1} r_{q1}$ i $x_{q2} \succeq_{q2} r_{q2}$ i $\ldots x_{qp} \succeq_{qp} r_{qp}$, to x *należy do* $Cl_t^{\geqslant}$,
gdzie dla każdego w_q, $z_q \in X_q$ zapis „$w_q \succeq_q z_q$" oznacza, że „ocena w_q jest **co najmniej** tak dobra jak z_q".

(2) **$D_{\geqslant}$-reguły możliwe**, dla których obiektami pozytywnymi są obiekty należące do $\overline{P}(Cl_t^{\geqslant})$:
Jeżeli $x_{q1} \succeq_{q1} r_{q1}$ i $x_{q2} \succeq_{q2} r_{q2}$ i $\ldots$ $x_{qp} \succeq_{qp} r_{qp}$, to x *być może należy do* $Cl_t^{\geqslant}$.

(3) **$D_{\leqslant}$-reguły pewne**, dla których obiektami pozytywnymi są obiekty należące do $\underline{P}(Cl_t^{\leqslant})$:
Jeżeli $x_{q1} \preceq_{q1} r_{q1}$ i $x_{q2} \preceq_{q2} r_{q2}$ i $\ldots$ $x_{qp} \preceq_{qp} r_{qp}$, to x *należy do* $Cl_t^{\leqslant}$,
gdzie dla każdego w_q, $z_q \in X_q$ zapis „$w_q \preceq_q z_q$" oznacza, że „ocena w_q jest **co najwyżej** tak dobra jak z_q".

(4) **$D_{\leqslant}$-reguły możliwe**, dla których obiektami pozytywnymi są obiekty należące do $\overline{P}(Cl_t^{\leqslant})$:
Jeżeli $x_{q1} \preceq_{q1} r_{q1}$ i $x_{q2} \preceq_{q2} r_{q2}$ i $\ldots$ $x_{qp} \preceq_{qp} r_{qp}$, to x *być może należy do* $Cl_t^{\leqslant}$.

(5) **$D_{\geqslant\leqslant}$-reguły przybliżone**, dla których obiektami pozytywnymi są obiekty należące do $\overline{P}(Cl_s^{\leqslant}) \cap \overline{P}(Cl_t^{\geqslant})$ $(s < t)$:
Jeżeli $x_{q1} \succeq_{q1} r_{q1}$ i $\ldots$ $x_{qk} \succeq_{qk} r_{qk}$ i $x_{qk+1} \preceq_{qk+1} r_{qk+1}$ i $\ldots$ $x_{qp} \preceq_{qp} r_{qp}$, to x *należy do* $Cl_s \cup Cl_{s+1} \cup \ldots \cup Cl_t$.

W części warunkowej $D_{\geqslant\leqslant}$-reguł przybliżonych możemy mieć warunki „$x_q \succeq_q r_q$" i „$x_q \preceq_q r'_q$", gdzie $r_q \leqslant r'_q$, dla tego samego kryterium $q \in C$. Ponadto, jeśli $r_q = r'_q$, to te dwa warunki sprowadzają się do „$x_q \sim_q r_q$", gdzie dla każdego w_q, $z_q \in X_q$ zapis „$w_q \sim_q z_q$" oznacza, że „ocena w_q jest nierozróżnialna z z_q".

Ponieważ reguła decyzyjna jest relacją konsekwencji (specyficzną implikacją), zatem przez **regułę minimalną** rozumiemy taką relację konsekwencji, że nie istnieje inna tego typu relacja, której część warunkowa (CZW) byłaby co najmniej tak samo słaba (innymi słowy, reguła wykorzystująca podzbiór warunków i/lub słabsze warunki), a część decyzyjna (CZD) co najmniej tak samo silna (innymi słowy, $D_{\geqslant}$- lub $D_{\leqslant}$-reguła przydzielająca obiekty do tego samego złożenia klas lub do jego podzbioru albo $D_{\geqslant\leqslant}$-reguła przydzielająca obiekty do tego samego lub szerszego ciągu klas).

Reguły decyzyjne typu (1) i (3) reprezentują wiedzę pewną wyindukowaną z tablicy decyzyjnej, podczas gdy reguły typu (2) i (4) reprezentują wiedzę możliwą (być może wątpliwą), a reguły typu (5) reprezentują wiedzę wątpliwą.

Ponadto reguły typu (1) i (3) są **deterministyczne**, jeśli nie pokrywają obiektów spoza klas wskazanych w części decyzyjnej; w przeciwnym razie są one **niedeterministyczne**. Reguła niedeterministyczna jest scharakteryzowana przez **współczynnik wiarygodności**, określający prawdopodobieństwo, że obiekt spełniający część warunkową reguły spełnia również jej część decyzyjną. Reguły niedeterministyczne można otrzymać, zakładając, że źródłem reguł pewnych lub możliwych są przybliżenia zdefiniowane według podejścia DRSA ze zmienną spójnością.

Przejdźmy teraz do omówienia zastosowania reguł decyzyjnych do klasyfikacji obiektów mających określone oceny za pomocą kryteriów C. Stosując $D_{\geqslant}$-reguły do obiektu x, możliwe jest, że x albo pasuje do CZW co najmniej jednej z tych reguł, albo nie pasuje do CZW żadnej z tych reguł. W razie co najmniej jednokrotnego dopasowania rozsądnie jest przyjąć, że x należy do klasy Cl_t, będącej najniższą klasą złożenia „w górę" $Cl_t^{\geqslant}$ wynikającego z przecięcia wszystkich CZD reguł dopasowanych do x. Dokładniej mówiąc, jeśli x pasuje do CZW reguł $\rho_1, \rho_2, \ldots, \rho_m$, których CZD zalecają $x \in Cl_{t1}^{\geqslant}$, $x \in Cl_{t2}^{\geqslant}$, $\ldots$, $x \in Cl_{tm}^{\geqslant}$, to x należy przydzielić do klasy Cl_t, gdzie $t = \min\{t1, t2, \ldots, tm\}$. W razie braku dopasowania x należy przydzielić do Cl_1, czyli do klasy najgorszej, nie znaleziono bowiem reguły, która zalecałaby przydział x do klasy lepszej.

Analogicznie, stosując $D_{\leqslant}$-reguły do obiektu x, możliwe jest, że x albo pasuje do CZW co najmniej jednej z tych reguł, albo nie pasuje do CZW żadnej z tych reguł. W razie co najmniej jednokrotnego dopasowania rozsądnie jest przyjąć, że x należy do klasy Cl_t, będącej najwyższą klasą złożenia „w dół" $Cl_t^{\leqslant}$ wynikającego z przecięcia wszystkich CZD reguł dopasowanych do x. Dokładniej mówiąc, jeśli x pasuje do CZW reguł $\rho_1, \rho_2, \ldots, \rho_m$, których CZD zalecają $x \in Cl_{t1}^{\leqslant}$, $x \in Cl_{t2}^{\leqslant}$, $\ldots$, $x \in Cl_{tm}^{\leqslant}$, to x należy przydzielić do klasy Cl_t, gdzie $t = \max\{t1, t2, \ldots, tm\}$. W razie braku dopasowania x należy przydzielić do Cl_n, czyli do klasy najlepszej, nie znaleziono bowiem reguły, która zalecałaby przydział x do klasy gorszej.

Wreszcie, stosując $D_{\geqslant\leqslant}$-reguły do obiektu x, wnioskujemy, że x należy do sumy wszystkich złożeń klas występujących w CZD reguł dopasowanych do x. Brak dopasowania nie implikuje w tym wypadku żadnego przydziału. Inne procedury klasyfikacji, które biorą pod uwagę różne charakterystyki reguł, zostały omówione w pracy [2].

Zbiór reguł decyzyjnych jest **kompletny**, jeśli pokrywa wszystkie obiekty z tablicy decyzyjnej w ten sposób, że obiekt spójny jest zaklasyfikowany według powyższej zasady do swojej oryginalnej klasy, a obiekt niespójny jest zaklasyfikowany do ciągu klas wskazanych przez obiekty pozostające z nim w stanie niespójności. Zbiór reguł, który jest kompletny i nienadmiarowy (w tym sensie, że usunięcie z niego dowolnej reguły czyni go niekompletnym), nazywamy **minimalnym**.

Zauważmy, że składnia reguł decyzyjnych wyindukowanych z przybliżeń złożeń klas decyzyjnych według DRSA wykorzystuje pojęcie stożków dominacji: każdy profil warunkowy w CZW jest stożkiem dominacji w X_C i każdy profil decyzyjny w CZD jest stożkiem dominacji w X_D – w obu przypadkach stożek jest dodatni dla $D_{\geqslant}$-reguł i ujemny dla $D_{\leqslant}$-reguł.

Warto także zauważyć, że stożki dominacji mogą mieć początek w dowolnym punkcie przestrzeni X_C bez obawy, że otrzymane reguły będą zbyt specyficzne. Zatem, przeciwnie niż w podstawowym podejściu zbiorów przybliżonych opartym na relacji nierozróżnialności, w DRSA przestrzeń atrybutów warunkowych X_C nie musi być dyskretyzowana.

Procedury indukcji reguł dla DRSA przedstawiono dokładniej w pracy [19].

Dodajmy na zakończenie, że przybliżenia złożeń klas decyzyjnych wynikające z podejścia DRSA mogą być punktem wyjścia do indukowania monotonicznych drzew decyzyjnych, co przedstawiono w pracy [5].

8.4.5. Przykład zastosowania podejścia DRSA

Aby pokazać praktyczne walory DRSA, zastosujemy je do analizy próbki danych pochodzących z greckiego banku przemysłowego ETEVA, który finansuje firmy przemysłowe i handlowe w Grecji [39]. Próbka zawiera dane dotyczące 39 firm. Ekspert finansowy banku ETEVA dokonał przydziału każdej z tych firm do jednej z trzech klas ryzyka bankructwa. Decyzja o przydziale do klasy jest reprezentowana przez atrybut decyzyjny d przyjmujący następujące wartości:

$d = \mathrm{A}$, czyli „akceptowalna";

$d = \mathrm{U}$, czyli „niepewna";

$d = \mathrm{NA}$, czyli „nieakceptowalna".

Klasyfikacja dokonywana przez d jest oznaczona przez $\boldsymbol{Cl} = \{Cl_{\mathrm{A}}, Cl_{\mathrm{U}}, Cl_{\mathrm{NA}}\}$. Klasa Cl_{A} jest oczywiście lepsza od Cl_{U}, a ta z kolei jest lepsza od Cl_{NA}.

Dane zawierają także ocenę firm za pomocą następujących dwunastu kryteriów warunkowych tworzących zbiór C ($\uparrow$ oznacza preferencję rosnącą z wartością kryterium, a $\downarrow$ oznacza preferencję malejącą z wartością kryterium):

- A_1 = zyski przed oprocentowaniem i opodatkowaniem w stosunku do sumy aktywów, $\uparrow$
- A_2 = zysk netto w stosunku do majątku netto, $\uparrow$
- A_3 = suma pasywów w stosunku do sumy aktywów, $\downarrow$
- A_4 = suma pasywów w stosunku do obrotów, $\downarrow$
- A_5 = koszty obsługi kapitału w stosunku do sprzedaży, $\downarrow$
- A_6 = koszty ogólne i administracyjne w stosunku do sprzedaży, $\downarrow$
- A_7 = doświadczenie kierownictwa firmy, $\uparrow$ (bardzo niskie = 1, niskie = 2, średnie = 3, wysokie = 4, bardzo wysokie = 5),
- A_8 = pozycja niszowa firmy, $\uparrow$ (zła = 1, raczej zła = 2, średnia = 3, dobra = 4, bardzo dobra = 5),
- A_9 = infrastruktura firmy, $\uparrow$ (zła = 1, raczej zła = 2, średnia = 3, dobra = 4, bardzo dobra = 5),
- A_{10} = struktura organizacyjna firmy, $\uparrow$ (zła = 1, raczej zła = 2, średnia = 3, dobra = 4, bardzo dobra = 5),
- A_{11} = konkurencyjność firmy, $\uparrow$ (niska = 1, średnia = 2, wysoka = 3, bardzo wysoka = 4),
- A_{12} = elastyczność względem rynku, $\uparrow$ (bardzo niska = 1, niska = 2, średnia = 3, wysoka = 4, bardzo wysoka = 5).

Pierwszych sześć kryteriów ma charakter ilościowy (wskaźniki finansowe), a pozostałych sześć kryteriów ma charakter porządkowy. Dane w postaci finansowej tablicy decyzyjnej są przedstawione w tabeli 8.2.

Zauważmy, że między kryteriami warunkowymi ze zbioru C a decyzją d istnieje korelacja semantyczna, na przykład powiększenie zysku A_1 przy niezmienionych ocenach na pozostałych kryteriach nie powinno powiększać ryzyka bankructwa firmy, czyli nie powinno pogarszać decyzji d.

Tabela 8.2. Finansowa tablica decyzyjna

Firma	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	d
F1	16,4	14,5	59.82	2,5	7,5	5,2	5	3	5	4	2	4	A
F2	35,8	67,0	64.92	1,7	2,1	4,5	5	4	5	5	4	5	A
F3	20,6	61,75	75,71	3,6	3,6	8,0	5	3	5	5	3	5	A
F4	11,5	17,1	57,1	3,8	4,2	3,7	5	2	5	4	3	4	A
F5	22,4	25,1	49,8	2,1	5,0	7,9	5	3	5	5	3	5	A
F6	23,9	34,5	48,9	1,7	2,5	8,0	5	3	4	4	3	4	A
F7	29,9	44,0	57,8	1,8	1,7	2,5	5	4	4	5	3	5	A
F8	8,7	5,4	27,4	3,3	4,5	4,5	5	2	4	4	1	4	A
F9	25,7	29,7	46,8	1,7	4,6	3,7	4	2	4	3	1	3	A
F10	21,2	24,6	64,8	3,7	3,6	8,0	4	2	4	4	1	4	A
F11	18,32	31,6	69,3	4,4	2,8	3,0	4	3	4	4	3	4	A
F12	20,7	19,3	19,7	0,7	2,2	4,0	4	2	4	4	1	3	A
F13	9,9	3,5	53,1	4,5	8,5	5,3	4	2	4	4	1	4	A
F14	10,4	9,3	80,9	9,4	1,4	4,1	4	2	4	4	3	3	A
F15	17,7	19,8	52,8	3,2	7,9	6,1	4	4	4	4	2	5	A
F16	14,8	15,9	27,94	1,3	5,4	1,8	4	2	4	3	2	3	A
F17	16,0	14,7	53,5	3,9	6,8	3,8	4	4	4	4	2	4	A
F18	11,7	10,01	42,1	3,9	12,2	4,3	5	2	4	2	1	3	A
F19	11,0	4,2	60,8	5,8	6,2	4,8	4	2	4	4	2	4	A
F20	15,5	8,5	56,2	6,5	5,5	1,8	4	2	4	4	2	4	A
F21	13,2	9,1	74,1	11,21	6,4	5,0	2	2	4	4	2	3	U
F22	9,1	4,1	44,8	4,2	3,3	10,4	3	4	4	4	3	4	U
F23	12,9	1,9	65,02	6,9	14,01	7,5	4	3	3	2	1	2	U
F24	5,9	−27,7	77,4	−32,2	16,6	12,7	3	2	4	4	2	3	U
F25	16,9	12,4	60,1	5,2	5,6	5,6	3	2	4	4	2	3	U
F26	16,7	13,1	73,5	7,1	11,9	4,1	2	2	4	4	2	3	U
F27	14,6	9,7	59,5	5,8	6,7	5,6	2	2	4	4	2	4	U
F28	5,1	4,9	28,9	4,3	2,5	46,0	2	2	3	3	1	2	U
F29	24,4	22,3	32,8	1,4	3,3	5,0	2	3	4	4	2	3	U
F30	29,7	8,6	41,8	1,6	5,2	6,4	2	3	4	4	2	3	U
F31	7,3	−64,5	67,5	−2,2	30,1	8,7	3	3	4	4	2	3	NA
F32	23,7	31,9	63,6	3,5	12,1	10,2	3	2	3	4	1	3	NA
F33	18,9	13,5	74,5	10,0	12,0	8,4	3	3	3	4	3	4	NA
F34	13,9	3,3	78,7	25,5	14,7	10,1	2	2	3	4	3	4	NA
F35	−13,3	−31,1	63,0	−10,0	21,2	23,1	2	1	4	3	1	2	NA
F36	6,2	−3,2	46,1	5,1	4,8	10,5	2	1	3	3	2	3	NA
F37	4,8	−3,3	71,9	34,6	8,6	11,6	2	2	4	4	2	3	NA
F38	0,1	−9,6	42,5	−20,0	12,9	12,4	1	1	4	3	1	3	NA
F39	13,6	−9,1	76,0	11,4	17,1	10,3	1	1	2	1	1	2	NA

Najważniejsze pytania, na które szukamy odpowiedzi w procesie odkrywania wiedzy, są następujące:

- Czy dane zawarte w finansowej tablicy decyzyjnej są spójne?
- Jakie są redukty zbioru kryteriów, które zapewniają taką samą jakość przybliżenia klasyfikacji jak cały zbiór dwunastu kryteriów?
- Jakie wzorce w postaci reguł decyzyjnych można odkryć z finansowej tablicy decyzyjnej?
- Jakie są minimalne zbiory reguł pokrywających wszystkie przykłady z próbki?

Odpowiemy na te pytania, stosując DRSA. Aby uwzględnić porządek preferencyjny w klasyfikacji $\boldsymbol{Cl}$, rozpatrywane będą złożenia klas „w górę" $Cl_{\mathrm{U}}^{\geqslant}$, $Cl_{\mathrm{A}}^{\geqslant}$ i „w dół" $Cl_{\mathrm{NA}}^{\leqslant}$, $Cl_{\mathrm{U}}^{\leqslant}$.

Pierwszym wynikiem zastosowania DRSA jest stwierdzenie, że finansowa tablica decyzyjna jest ***spójna*** dla danego zbioru kryteriów C. Stąd C-dolne i C-górne przybliżenia złożeń klas $Cl_{\mathrm{U}}^{\geqslant}$, $Cl_{\mathrm{A}}^{\geqslant}$ i $Cl_{\mathrm{NA}}^{\leqslant}$, $Cl_{\mathrm{U}}^{\leqslant}$ są identyczne. Innymi słowy, jakość przybliżenia klasyfikacji dla C jest równa 1.

Drugim wynikiem jest zbiór 18 ***reduktów*** zapewniajacy tę samą jakość przybliżenia klasyfikacji co cały zbiór 12 kryteriów:

$$RED_{Cl}^{1} = \{A_1, A_4, A_5, A_7\}, \qquad RED_{Cl}^{2} = \{A_2, A_4, A_5, A_7\},$$
$$RED_{Cl}^{3} = \{A_3, A_4, A_6, A_7\}, \qquad RED_{Cl}^{4} = \{A_4, A_5, A_6, A_7\},$$
$$RED_{Cl}^{5} = \{A_4, A_5, A_7, A_8\}, \qquad RED_{Cl}^{6} = \{A_2, A_3, A_7, A_9\},$$
$$RED_{Cl}^{7} = \{A_1, A_3, A_4, A_7, A_9\}, \qquad RED_{Cl}^{8} = \{A_1, A_5, A_7, A_9\},$$
$$RED_{Cl}^{9} = \{A_2, A_5, A_7, A_9\}, \qquad RED_{Cl}^{10} = \{A_4, A_5, A_7, A_9\},$$
$$RED_{Cl}^{11} = \{A_5, A_6, A_7, A_9\}, \qquad RED_{Cl}^{12} = \{A_4, A_5, A_7, A_{10}\},$$
$$RED_{Cl}^{13} = \{A_1, A_3, A_4, A_7, A_{11}\}, \qquad RED_{Cl}^{14} = \{A_2, A_3, A_4, A_7, A_{11}\},$$
$$RED_{Cl}^{15} = \{A_4, A_5, A_6, A_{12}\}, \qquad RED_{Cl}^{16} = \{A_1, A_3, A_5, A_6, A_9, A_{12}\},$$
$$RED_{Cl}^{17} = \{A_3, A_4, A_6, A_{11}, A_{12}\}, \qquad RED_{Cl}^{18} = \{A_1, A_2, A_3, A_6, A_9, A_{11}, A_{12}\}.$$

Wszystkich 18 reduktów jest jednakowo dobrych i wystarczających do scharakteryzowania klasyfikacji dokonanej przez eksperta finansowego banku ETEVA w odniesieniu do 39 firm. Rdzeń zbioru kryteriów jest pusty ($CORE_{Cl} = \emptyset$), co oznacza, że żadne kryterium nie jest niezbędne, czyli niezastępowalne przez inne, w procesie przybliżania klasyfikacji $\boldsymbol{Cl}$. Ponadto wszystkie kryteria są wymienne i żadne kryterium nie jest nadmiarowe.

Trzecim wynikiem zastosowania DRSA jest zbiór ***wszystkich*** reguł decyzyjnych, które można wyindukowć z tablicy decyzyjnej. Otrzymaliśmy 74 reguły opisujące $Cl_{\mathrm{NA}}^{\leqslant}$, 51 reguł opisujących $Cl_{\mathrm{U}}^{\leqslant}$, 75 reguł opisujących $Cl_{\mathrm{U}}^{\geqslant}$ i 79 opisujących $Cl_{\mathrm{A}}^{\geqslant}$.

Czwartym wynikiem jest znalezienie ***minimalnych zbiorów*** reguł decyzyjnych. Istnieje kilka takich zbiorów; jeden z nich jest pokazany poniżej (w nawiasach jest podana liczba obiektów pokrywanych przez odpowiednią regułę, czyli tzw. siła reguły):

1. **Jeżeli** $f(x, A_3) \geqslant 67{,}5$ **i** $f(x, A_4) \geqslant -2{,}2$ **i** $f(x, A_6) \geqslant 8{,}7$, **to** $x \in Cl_{\mathrm{NA}}^{\leqslant}$; (4).

2. **Jeżeli** $f(x, A_2) \leqslant 3{,}3$ **i** $f(x, A_7) \leqslant 2$, **to** $x \in Cl_{\mathrm{NA}}^{\leqslant}$; (5).
3. **Jeżeli** $f(x, A_3) \geqslant 63{,}6$ **i** $f(x, A_7) \leqslant 3$ **i** $f(x, A_9) \leqslant 3$, **to** $x \in Cl_{\mathrm{NA}}^{\leqslant}$; (4).
4. **Jeżeli** $f(x, A_2) \leqslant 12{,}4$ **i** $f(x, A_6) \geqslant 5{,}6$, **to** $x \in Cl_{\mathrm{U}}^{\leqslant}$; (14).
5. **Jeżeli** $f(x, A_7) \leqslant 3$, **to** $x \in Cl_{\mathrm{U}}^{\leqslant}$; (18).
6. **Jeżeli** $f(x, A_2) \geqslant 3{,}5$ **i** $f(x, A_5) \leqslant 8{,}5$, **to** $x \in Cl_{\mathrm{U}}^{\geqslant}$; (26).
7. **Jeżeli** $f(x, A_7) \geqslant 4$, **to** $x \in Cl_{\mathrm{U}}^{\geqslant}$; (21).
8. **Jeżeli** $f(x, A_1) \geqslant 8{,}7$ **i** $f(x, A_9) \geqslant 4$, **to** $x \in Cl_{\mathrm{U}}^{\geqslant}$; (27).
9. **Jeżeli** $f(x, A_2) \geqslant 3{,}5$ **i** $f(x, A_7) \geqslant 4$, **to** $x \in Cl_{\mathrm{A}}^{\geqslant}$; (20).

Ponieważ minimalny zbiór reguł jest kompletny i złożony wyłącznie z $D_{\geqslant}$-reguł i $D_{\leqslant}$-reguł, więc zastosowanie tych reguł do 39 firm da w wyniku dokładną reklasyfikację wszystkich firm do ich oryginalnych klas.

Minimalne zbiory reguł reprezentują najbardziej zwięzłe i nienadmiarowe wzorce klasyfikacji, czyli podsumowanie wiedzy zawartej w finansowej tablicy decyzyjnej. Powyższy zbiór 9 reguł decyzyjnych korzysta z 18 warunków elementarnych na 8 kryteriach, tzn. jedynie z 3,85% ocen charakterystycznych zawartych w tablicy danych.

Znane metody uczenia maszynowego nie radzą sobie z danymi uporządkowanymi według preferencji, czyli nie uwzględniają wiedzy dziedzinowej typu (iii) (patrz punkt 8.4.1), która występuje w klasyfikacji wielokryterialnej. Z kolei wśród metod wielokryterialnego wspomagania decyzji istnieją metody dostosowane do klasyfikacji wielokryterialnej, jednak nie potrafią one odkrywać z danych wzorców klasyfikacji w postaci wyrażeń logicznych typu reguł decyzyjnych; w kontekście tych metod wzorzec klasyfikacji jest utożsamiany z modelem preferencji i zwykle jest to funkcja (użyteczności) lub relacja (przewyższania) [26].

Porównując DRSA z podstawową wersją zbiorów przybliżonych opartą na relacji nierozróżnialności (CRSA), zauważmy, że poza wszystkimi różnicami dotyczącymi zastosowania relacji dominacji w miejsce relacji nierozróżnialności DRSA nie wymaga dyskretyzacji atrybutów ilościowych. Ponadto składnia reguł oparta na dominacji jest bogatsza niż składnia reguł oparych na nierozróżnialności, gdyż w warunkach elementarnych tych pierwszych używa się relacji typu „$\leqslant$", „$=$", „$\geqslant$", podczas gdy w warunkach elementarnych tych drugich stosuje się relację „$=$". W konsekwencji reprezentacja wiedzy za pomocą reguł opartych na dominacji jest bardziej zwięzła.

8.5. Zastosowania podejścia zbiorów przybliżonych opartego na dominacji

W poprzednim podrozdziale przedstawiliśmy podejście zbiorów przybliżonych oparte na dominacji (DRSA), które jest dostosowane do analizy danych klasyfikacyjnych z uwzględnieniem wiedzy o uporządkowaniu dziedzin atrybutów według preferencji i o korelacji semantycznej atrybutów warunkowych i decyzyjnych (patrz wiedza dziedzinowa (iii) w punkcie 8.4.1). Atrybuty tego typu noszą nazwę kryteriów, a dane zawarte w tablicy decyzyjnej odpowiadają przykładom klasyfikacji wielokryterialnej.

Z punktu widzenia wspomagania decyzji przykłady te można potraktować jako informacje preferencyjne dostarczone przez decydenta w celu skonstruowania modelu jego preferencji. Faktycznie, w wyniku zastosowania DRSA do tej informacji preferencyjnej uzyskujemy **model preferencji decydenta** w postaci zbioru reguł decyzyjnych opartych na dominacji. Reguły decyzyjne mają specjalną składnię, do której wchodzi częściowy profil ocen na podzbiorze kryteriów i relacja dominacji między tym profilem a oceną obiektu. Jest to bardzo prosta i naturalna w interpretacji reprezentacja preferencji. Jej konkurentami w klasycznej teorii decyzji są funkcja użyteczności i relacja przewyższania. Modele te nie są tak naturalne w interpretacji jak reguły decyzyjne, a ponadto ich konstrukcja nastręcza wiele kłopotów, głównie z powodu trudności związanych z akwizycją specyficznych informacji preferencyjnych typu wag kryteriów, współczynników substytucji, progów preferencji, nierozróżnialności i weta. Tymczasem pozyskiwanie informacji preferencyjnych w przypadku DRSA jest stosunkowo proste – decydent podaje przykłady klasyfikacji lub jego decyzje są automatycznie rejestrowane *on line*.

Zaproponowanie „trzeciej drogi" w modelowaniu preferencji wymagało porównania nowego modelu regułowego z modelami klasycznymi. Autorzy podejścia DRSA dokonali obszernej charakteryzacji podstaw aksjomatycznych wszystkich trzech typów modeli preferencji [16, 18, 33], z jednej strony w terminach teorii wspólnego pomiaru (*conjoint measurement*), a z drugiej w terminach reguł decyzyjnych. W porównaniu z innymi tego typu charakteryzacjami operatorów agregacji aksjomaty zaproponowane we wspomnianych pracach nie wymagają żadnych wstępnych założeń odnośnie do skal kryteriów. Ważnym wnioskiem z tych badań jest udowodniony fakt, że model preferencji w postaci reguł decyzyjnych jest najogólniejszym modelem preferencji wśród wszystkich znanych operatorów agregacji. Model ten spełnia ponadto postulat przejrzystości i zrozumiałości modeli preferencji we wspomaganiu decyzji.

Podejście DRSA było także dostosowywane do innych problemów decyzyjnych niż klasyfikacja wielokryterialna. Ważną kategorią wielokryterialnych problemów decyzyjnych są problemy wyboru i porządkowania. W tych problemach decyzje wobec obiektów nie są podejmowane na podstawie ich bezwzględnych ocen, tak jak w problemie klasyfikacji wielokryterialnej, lecz na podstawie względnych ocen wynikających z porównania obiektów parami. Z tego powodu punktem wyjścia do analizy za pomocą DRSA jest nie tablica decyzyjna z przykładami klasyfikacji, lecz tablica porównań obiektów parami. Wiersze tej tablicy odpowiadają parom obiektów, a kolumny kryteriom warunkowym i decyzji w postaci globalnej relacji preferencji; każda para obiektów w takiej tablicy jest opisana przez cząstkowe relacje preferencji na poszczególnych kryteriach i przez globalną relację preferencji. Celem analizy DRSA jest konstrukcja przybliżeń globalnej relacji preferencji i indukcja reguł decyzyjnych z tych przybliżeń. Reguły decyzyjne zawierają warunki elementarne dotyczące relacji cząstkowych i decyzję określającą relację globalną dla pary obiektów. Wynik zastosowania tych reguł na zbiorze obiektów może być przedstawiony w postaci grafu relacji preferencji – do wypracowania zalecenia (wyboru lub uporządkowania) można zastosować procedurę eksploatacji grafu typu „bilansowania przepływu" (*net flow score*) [14, 22, 34].

Inne adaptacje DRSA do rozwiązywania problemów decyzyjnych dotyczą

- podejmowania decyzji w warunkach ryzyka i niepewności [13],
- analizy niekompletnych danych [9],
- powiązań DRSA z podejściem zbiorów rozmytych [6, 7, 11],
- problemów decyzyjnych z hierarchiczną strukturą atrybutów i kryteriów [4],
- indukowania reguł asocjacyjnych z danych uporządkowanych według preferencji [21].

8.6. Zakończenie

Podane tu sformułowanie podstawowych pojęć teorii zbiorów przybliżonych jest bardzo proste, jednakże do celów praktycznych jest ono niewystarczające. Aby podejście to mogło być stosowane do rozwiązywania rzeczywistych problemów, wymagało ono wielu rozszerzeń i uzupełnień. Pozwoliło to na stworzenie narzędzia matematycznego, które mogło być z powodzeniem zastosowane do rozwiązywania złożonych problemów.

W rozdziale tym skoncentrowaliśmy się na rozszerzeniu podstawowego podejścia zbiorów przybliżonych na problemy decyzyjne, dla których charakterystyczny jest porządek preferencji w danych. Informacja o tym porządku należy do specyficznej wiedzy dziedzinowej, której nie uwzględniają inne metody analizy danych i odkrywania wiedzy. Opisane rozszerzenie, zwane DRSA, charakteryzuje się zastosowaniem relacji dominacji w miejsce klasycznej relacji nierozróżnialności; stanowi ono nowe narzędzie służące z jednej strony do odkrywania wiedzy z danych, a z drugiej strony do wielokryterialnego wspomagania decyzji.

Zainteresowany czytelnik znajdzie informacje o innych rozszerzeniach w Internecie: http://www.roughsets.org.

W Internecie dostępne jest także przykładowe oprogramowanie: ROSETTA toolkit (http://www.idi.ntnu.no/~aleks/rosetta), RSES – Rough Set Exploration System (http://www.roughsets.org), ROSE – ROugh Set data Explorer (http://idss.cs.put.poznan.pl/site/software.html/) oraz 4eMka i JAMM – Dominance-based Rough Set Approach to Multicriteria Classification (http://idss.cs.put.poznan.pl/site/software.html/).

Teoria zbiorów przybliżonych nie jest alternatywą dla klasycznej teorii zbiorów. Jest ona pewną formalizacją pojęć nieprecyzyjnych (nieostrych) przy wykorzystaniu pojęć precyzyjnych. Pojęcie nieprecyzyjne jest przedstawione w tej teorii za pomocą dwóch pojęć precyzyjnych (dolnego i górnego przybliżenia), co pozwala na operowanie klasycznym aparatem teorii mnogości do wyrażania i analizy pojęć nieprecyzyjnych.

Literatura

[1] Agrawal R., Mannila H., Srikant R., Toivinen H., Verkamo I. (1996): Fast discovery of association rules. W: *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad *et al.* (red.), AAAI Press, s. 307–328.

[2] Błaszczyński J., Greco S., Słowiński R. (2007): Multi-criteria classification – A new scheme for applicaton of dominance-based decision rules. *European Journal of Operational Research*, Vol. 180, No. 3.

[3] Cantor G. (1884): Über unendliche, lineare Punktmannigfaltigkeiten. Arbeiten zur Mengenlehre aus dem Jahren 1872–1884. Teubner, Leipzig.

[4] Dembczyński K., Greco S., Słowiński R. (2002): Methodology of rough-set-based classification and sorting with hierarchical structure of attributes and criteria. *Control and Cybernetics*, Vol. 31, s. 891–920.

[5] Giove S., Greco S., Matarazzo B., Słowiński R. (2002): Variable consistency monotonic decision trees. W: *Rough Sets and Current Trends in Computing*, J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong (red.), LNAI 2475, Springer, Berlin, s. 247–254.

[6] Greco S., Inuiguchi M., Słowiński R. (2002): Dominance-based rough set approach using possibility and necessity measures. W: *Rough Sets and Current Trends in Computing*, J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong (red.), LNAI 2475, Springer, Berlin, s. 85–92.

[7] Greco S., Inuiguchi M., Słowiński R. (2004): A new proposal for fuzzy rough approximations and gradual decision rule representation. W: *Rough Fuzzy and Fuzzy Rough Sets*, D. Dubois, J. Grzymala-Busse, M. Inuiguchi, L. Polkowski (red.), *Transactions on Rought Sets II*, LNCS 3135, Springer, Berlin, s. 319–342.

[8] Greco S., Matarazzo B., Słowiński R. (1999): The use of rough sets and fuzzy sets in MCDM. Chapter 14. W: *Advances in Multiple Criteria Decision Making*, T. Gal, T. Stewart, T. Hanne (red.), Kluwer, Boston, s. 14.1–14.59.

[9] Greco S., Matarazzo B., Słowiński R. (2000): Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. W: *Decision Making: Recent Developments and Worldwide Applications*, S.H. Zanakis, G. Doukidis, C. Zopounidis (red.), Kluwer, Boston, s. 295–316.

[10] Greco S., Matarazzo B., Słowiński R. (2000): Extension of the rough set approach to multicriteria decision support. *INFOR*, Vol. 38, s. 161–196.

[11] Greco S., Matarazzo B., Słowiński R. (2000): Fuzzy extension of the rough set approach to multicriteria and multiattribute sorting. W: *Preferences and Decisions under Incomplete Knowledge*, J. Fodor, B. De Baets, P. Perny (red.), Physica-Verlag, Heidelberg, s. 131–151.

[12] Greco S., Matarazzo B., Słowiński R. (2001): Assessment of a value of information using rough sets and fuzzy measures. W: *Fuzzy Sets and their Applications*, J. Chojcan, J. Łęski (red.), Silesian University of Technology Press, Gliwice, s. 185–193.

[13] Greco S., Matarazzo B., Słowiński R. (2001): Rough set approach to decisions under risk. W: *Rough Sets and Current Trends in Computing*, W. Ziarko, Y. Yao (red.), LNAI 2005, Springer, Berlin, s. 160–169.

[14] Greco S., Matarazzo B., Słowiński R. (2001): Rough set theory for multicriteria decision analysis. *European Journal of Operational Research*, Vol. 129, s. 1–47.

[15] Greco S., Matarazzo B., Słowiński R. (2002): Multicriteria classification. Chapter 16.1.9. W: *Handbook of Data Mining and Knowledge Discovery*, W. Kloesgen, J. Zytkow (red.), Oxford University Press, New York, s. 318–328.

[16] Greco S., Matarazzo B., Słowiński R. (2002): Preference representation by means of conjoint measurement and decision rule model. W: *Aiding Decisions with Multiple Criteria – Essays in Honor of Bernard Roy*, D. Bouyssou, E. Jacquet-Lagreze, P. Perny, R. Slowinski, D. Vanderpooten, Ph. Vincke (red.), Kluwer, Boston, s. 263–313.

[17] Greco S., Matarazzo B., Słowiński R. (2002): Rough approximation by dominance relations. *International Journal of Intelligent Systems*, Vol. 17, No. 2, s. 153–171.

[18] Greco S., Matarazzo B., Słowiński R. (2004): Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough--set decision rules. *European Journal of Operational Research*, Vol. 158, s. 271–292.

[19] Greco S., Matarazzo B., Słowiński R., Stefanowski J. (2001): An algorithm for induction of decision rules consistent with dominance principle. W: *Rough Sets and Current Trends in Computing*, W. Ziarko, Y. Yao (red.), LNAI 2005, Springer, Berlin, s. 304–313.

[20] Greco S., Matarazzo B., Słowiński R., Stefanowski J. (2001): Variable consistency model of dominance-based rough set approach. W: *Rough Sets and Current Trends in Computing*, W. Ziarko, Y. Yao (red.), LNAI 2005, Springer, Berlin, s. 170–181.

[21] Greco S., Matarazzo B., Słowiński R., Stefanowski J. (2002): Mining association rules in preference-ordered data. W: *Foundations of Intelligent Systems*, M.-S. Hacid, Z.W. Raś, D.A. Zighed, Y. Kodratoff (red.), LNAI 2366, Springer, Berlin, s. 442–450.

[22] Greco S., Matarazzo B., Słowiński R., Stefanowski J., Tsoukias A. (1998): Exploitation of a rough approximation of the outranking relation in multicriteria choice and ranking. W: *Trends in Multicriteria Decision Making*, T.J. Stewart, R.C. van den Honert (red.), LNEMS 465, Springer, Berlin, s. 45–60.

[23] Grzymała-Busse J.W. (1992): LERS – a system for learning from examples based on rough sets. W: *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, R. Słowiński (red.), Kluwer, Dordrecht, s. 3–18.

[24] Grzymała-Busse J.W. (1997): A new version of the rule induction system LERS. *Fundamenta Informaticae*, Vol. 31, s. 27–39.

[25] Michalski R.S., Bratko I., Kubat M. (red.) (1998): *Machine Learning and Data Mining – Methods and Applications*. Wiley, New York.

[26] Mousseau V., Słowiński R. (1998): Inferring an ELECTRE TRI model from assignment examples. *Journal of Global Optimization*, Vol. 12, s. 157–174.

[27] Pawlak (1982): Rough sets. *International Journal of Information and Computer Sciences*, Vol. 11, No. 5, s. 341–356.

[28] Pawlak Z., Słowiński R. (1994): Decision analysis using rough sets. *International Transactions on Operational Research*, Vol. 1, s. 107–114.

[29] Pawlak Z., Słowiński R. (1994): Rough set approach to multi-attribute decision analysis. *European Journal of Operational Research*, Vol. 72, s. 443–459.

[30] Polkowski L., Skowron A. (1999): Calculi of granules based on rough set theory: approximate distributed synthesis and granular semantics for computing with words. W: *New Directions in Rough sets, Data Mining and Soft-Granular Computing*, N. Zhong, A. Skowron, S. Ohsuga (red.), LNAI 1711, Springer, Berlin, s. 20–28.

[31] Skowron A., Polkowski L. (1997): Decision algorithms: a survey of rough set-theoretic methods. *Fundamenta Informaticae*, Vol. 27, s. 345–358.

[32] Słowiński R. (2003): Od sztucznej inteligencji do sztucznego życia, czyli o aktywnej funkcji informatyki. *Pro Dialog*, Nr. 16, s. 51–76.

[33] Słowiński R., Greco S., Matarazzo B. (2002): Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. *Control and Cybernetics*, Vol. 31, s. 1005–1035.

[34] Słowiński R., Greco S., Matarazzo B. (2002): Mining decision-rule preference model from rough approximation of preference relation. W: *Proceedings of the 26th IEEE Annual International Conference on Computer Software and Applications* (COMPSAC 2002), Oxford, England, s. 1129–1134.

[35] Słowiński R., Greco S., Matarazzo B. (2002): Rough set analysis of preference-ordered data. W: *Rough Sets and Current Trends in Computing*, J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong (red.), LNAI 2475, Springer, Berlin, s. 44–59.

[36] Słowiński R., Greco S., Matarazzo B. (2005): Rough set based decision support. Chapter 16. W: *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, E. Burke, G. Kendall (red.), Kluwer, Boston, s. 475–527.
[37] Słowiński R., Stefanowski J., Greco S., Matarazzo B. (2000): Rough sets based processing of inconsistent information in decision analysis. *Control and Cybernetics*, Vol. 29, s. 379–404.
[38] Słowiński R., Vanderpooten D. (2000): A generalized definition of rough approximations based on similarity. *IEEE Transactions on Data and Knowledge Engineering*, Vol. 12, s. 331–336.
[39] Słowiński R., Zopounidis C. (1995): Application of the rough set approach to evaluation of bankruptcy risk. *Intelligent Systems in Accounting, Finance and Management*, Vol. 4, s. 27–41.
[40] Stefanowski J. (2001): Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy. Rozprawa habilitacyjna. Wydawnictwo Politechniki Poznańskiej, Seria Rozprawy, nr 361, Poznań.
[41] Zadeh L. (1965): Fuzzy sets. *Information and Control*, Vol. 8, s. 338–353.
[42] Ziarko W., Shan N. (1994): An incremental learning algorithm for constructing decision rules. W: *Rough Sets, Fuzzy Sets and Knowledge Discovery*, W.P. Ziarko (red.), Springer, London, s. 326–334.