

# Wnioskowanie z danych z brakującymi wartościami atrybutów

## Autoreferat

Rafał Latkowski

### Spis treści

<b>1</b>	<b>Ogólna charakterystyka dziedziny rozprawy</b>	<b>1</b>
<b>2</b>	<b>Aktualny stan badań związanych z przedmiotem rozprawy</b>	<b>2</b>
<b>3</b>	<b>Główne wyniki prac autora</b>	<b>3</b>
<b>4</b>	<b>Wykaz prac przedstawionych w ramach rozprawy</b>	<b>6</b>
<b>5</b>	<b>Omówienie głównych wyników prac wchodzących w skład rozprawy</b>	<b>7</b>
	<b>Literatura</b>	<b>9</b>

## 1 Ogólna charakterystyka dziedziny rozprawy

Problemy związane z wnioskowaniem z danych zawierających opis obiektów (potocznie zwane wnioskowaniem z danych) były i są intensywnie badane różnorodnymi metodami, począwszy od prac Legendre’a i Gaussa sprzed ponad 200 lat, aż po najnowsze prace z zakresu statystycznych metod uczenia maszynowego, metod zbiorów przybliżonych i obliczeń granularnych, czy też tzw. głębokich sieci neuronowych [BFOS84, HTF01, GBC16]. Częstą trudnością na jaką napotykają badacze jest niekompletny opis obiektów, zarówno w ogólnym sensie braku dostępnej informacji, jak i w szczególnym sensie niekompletności opisu w zdefiniowanej już wcześniej przestrzeni opisów. Jeśli dane definiujemy jako system informacyjny (w sensie Pawlaka [Paw83]), składający się ze skończonego zbioru obiektów oraz skończonego zbioru atrybutów, które są funkcjami nadającymi wartość atrybutu dla obiektu, to taka niekompletność opisu obiektu oznacza brakującą wartość atrybutu na tym obiekcie.

Problem wnioskowania na podstawie obiektów z brakującymi wartościami atrybutów jest znany w uczeniu maszynowym i poświęcono mu wiele prac w celu głębszego poznania tego problemu i opracowania metod wnioskowania w przypadku brakujących wartości atrybutów [LR87, MFK<sup>+</sup>04, Mar08, Gra12, HB14, vB18, Laa18]. W odróżnieniu jednak od analogicznego zagadnienia występującego w relacyjnych bazach danych, gdzie w latach ’70-’90 ubiegłego wieku wypracowano rozwiązania stosowane powszechnie jako standard [Lip79, IL84], w zakresie uczenia maszynowego nie wypracowano podejścia, które całościowo obejmowałoby aspekty związane z brakującymi wartościami. Stwierdzenie to jest zarówno prawdziwe w odniesieniu metod wnioskowania z wykorzystaniem teorii zbiorów przybliżonych [PSŻ95, GBW97, GBH01, GMS00, ST01a], jak i w odniesieniu do pozostałych metod takich jak np. sieci neuronowe [Mar08, GS10, MGR17, GBC16, SST<sup>+</sup>18].

Pod koniec lat '70 i na początku '80 ubiegłego wieku ukazały się m.in. prace Friedmana [Fri77], Kononenki [KBR84] i Breimana [BFOS84], które przedstawiają pewne metody umożliwiające poradzenie sobie z brakującymi wartościami podczas indukcji drzew decyzyjnych. Podstawową ideą wspólną dla zaproponowanych metod jest klasyfikacja obiektów przez więcej niż jedną ścieżkę w drzewie, co uzyskiwane jest przez wirtualny podział i fizyczną replikację obiektów danych. Taka metoda została zaimplementowana przez Quinlan'a w popularnym algorytmie C4.5 [Qui89], na którym wzorowanych jest wiele innych implementacji [WF00]. To podejście jest stosowane na szeroką skalę ze względu na dość dobre wyniki i prostą interpretację. Niestety, wszystkie tego typu metody powodują załamanie dobrej, liniowo-logarytmicznej złożoności obliczeniowej indukcji drzewa co opisane jest w mojej pracy [Lat03a]. Prowadzone również są prace nad metodami indukcji reguł decyzyjnych i asocjacyjnych, które pozwalają na bezpośrednie wnioskowanie na danych z brakującymi wartościami [PR07, WI00, CG14, GBW97].

Oprócz metod, które potrafią bezpośrednio przetwarzać dane z brakującymi wartościami atrybutów, opracowano również metody uzupełniania brakujących wartości [MFK<sup>+</sup>04, LR87, Gra12, HB14, vB18, Laa18], aby uzupełnione dane można było przetwarzać dowolną metodą wnioskowania, która nie zakłada możliwości wnioskowania na danych z brakującymi wartościami. Najprostsza metoda uzupełniania brakujących wartości, sprowadzająca się do podstawienia nieużywanej wartości z dziedziny, jest znana od samego początku uczenia maszynowego. Jednakże taka metoda prowadzi do znacznego spadku poprawności wnioskowania i jakości klasyfikacji. Stosowane metody uzupełniania brakujących wartości mogą być podzielone na proste, gdy nie jest konstruowany żaden model danych lub model ten jest relatywnie prosty [FH02, GBGBG99, GBH01], oraz złożone, gdzie uzupełnianie odbywa się na podstawie uprzednio wygenerowanego modelu danych [MFK<sup>+</sup>04, Gra12, Laa18], a nawet takie gdy uzupełnianie przeprowadzane jest wielokrotnie w celu lepszej estymacji statystycznych parametrów modelu [Rub87, HB14, vB18]. Jednakże proste uzupełnianie danych (za pomocą statystyk punktowych, np. średniej) często narusza strukturę danych, a uzupełnianie w oparciu o złożony model, a szczególnie uzupełnianie wielokrotne, można stosować tylko przy algorytmach, których wyniki (stanowiące parametry modelu) można w pewien sposób uśredniać, jak np. współczynniki regresji (ale już nie do drzew czy reguł decyzyjnych).

## 2 Aktualny stan badań związanych z przedmiotem rozprawy

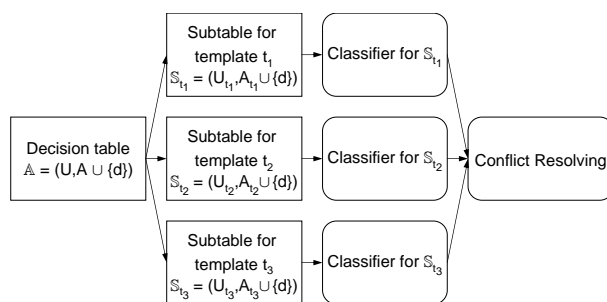
Problemy pojawiające się w związku z brakującymi wartościami atrybutów są badane również w ramach teorii zbiorów przybliżonych, która niejako stworzona została do radzenia sobie z ogólnie rozumianym problemem niedostępności informacji (por. [Paw83, Paw91, KPPS99]) oraz jest rozszerzana na przypadek szczególny braku wartości atrybutów. Podejścia do tego problemu można podzielić ze względu na zakładane modyfikacje teorii zbiorów przybliżonych. W pierwszej grupie znajdują się podejścia zakładające, że teoria zbiorów przybliżonych powinna zostać zmodyfikowana w taki sposób, aby w naturalny sposób można było w jej ramach badać problemy występujące w przypadku brakujących wartości atrybutów (np. [GBH01, Kry98, PSZ95, ST01a, Lat05]). W drugiej grupie znajdują się podejścia niezakładające modyfikacji teorii zbiorów przybliżonych, które pozwalają na użycie klasycznych metod zbiorów przybliżonych poprzez omijanie brakujących wartości atrybutów w dodatkowych krokach proponowanych algorytmów (np. [GBGBG99, GBH01, Lat03b]). Na osobną uwagę zasługują prace, które rozszerzają definicję systemu informacyjnego przyjmując tzw. wielowartościowe dziedziny atrybutów (np. atrybut jako funkcja w zbiór podzbiorów dziedziny), uwzględniające również możliwość reprezentacji braku wartości [DO02, SNŚ10, MH16], ale najczęściej jednak stosowane są podejścia zgodne z pierwszą postacią normalną relacyjnych baz danych [Cod70].

Ponieważ teoria zbiorów przybliżonych w pierwotnym sformułowaniu przez Pawlaka bazuje na relacji nierozróżnialności, to większość prac pierwszej grupy koncentrowała się prawie wyłącznie na modyfikacjach tej relacji. Tradycja modyfikacji relacji nierozróżnialności jest dosyć długa i wywodzi się nie tylko z teorii zbiorów przybliżonych, ale również z innych działów matematyki. W algebrze uniwersalnej występują pojęcia częściowych struktur, czy też równości słabych [Bur86, Sta96, Bin00], które zainspirowały badaczy do wprowadzenia relacji tolerancji lub relacji symetrycznego podobieństwa jako modelu relacji nierozróżnialności [GBH01, Kry98, PSZ95, ST01a]. W celu omięcia problemów z zaproponowaną semantyką brakujących wartości, wykazanych m.in. w [ST01a], zostały zaproponowane również inne modele relacji nierozróżnialności, jak na przykład relacja niesymetrycznego podobieństwa, badana w pracach [GMS99, GMSZ99, SV00, ST99, ST01a, ST01b]. W celu osiągnięcia jeszcze większej elastyczności zaproponowano również parametryzowane relacje podobieństwa, czasem również z rozszerzeniem teorii zbiorów przybliżonych o elementy zbiorów rozmytych [GMS00, ST99, ST01b]. Wszystkie te modyfikacje wymuszają jednak szczególną semantykę brakujących wartości. Taka semantyka jest ustalona dla wszystkich zbiorów danych i ich atrybutów (cech opisujących obiekty) w dokładnie ten sam sposób, co może powodować nadmierne obciążenie założeniami modelu. Należy jednak stwierdzić, że te podejścia okazały się pomocne przy wielu zastosowaniach, a z pewnością pozwalają na generowanie lepszych klasyfikatorów niż przy użyciu standardowej relacji nierozróżnialności (po uzupełnieniu braków wymienionych poniżej w opisie wyników moich prac). W ostatnich latach jawne wykorzystanie jednej lub dwu alternatywnych definicji relacji nierozróżnialności (tolerancji/symetryczna i podobieństwa/niesymetryczna) pojawia się w wielu badaniach związanych z teorią zbiorów przybliżonych [UMN<sup>+</sup>15, CGGBM17, HWY17, LWY18, WL18], jednakże w żadnym przypadku nie zakładają one uczenia się semantyki brakujących wartości dopasowanej do danych (tj. wyuczania się relacji nierozróżnialności dopasowanej do danych). Nieco odrębnym nurtem badań od modyfikacji relacji nierozróżnialności jest aktywne poszukiwanie alternatywnych sposobów ujęcia aproksymacji pojęć w przypadku danych z brakującymi wartościami [Rzą09, GB13, CGGBM17, LWY18]. Może to stanowić naturalną drogę dalszego rozwoju prac nad uogólnionymi relacjami nierozróżnialności, oraz znaleźć swoje odzwierciedlenie w implementacji Rseslib 3.

W drugiej grupie podejść znajdują się metody, które nie zakładają modyfikacji relacji nierozróżnialności i teorii zbiorów przybliżonych jako takiej. Zaproponowana przez Grzymałę-Busse modyfikacja algorytmu indukcji reguł decyzyjnych LEM2 zakłada jedynie zmianę metody generowania reguł [GBH01, GBW97]. Zmodyfikowana wersja algorytmu MLEM2 omija przykłady z brakującymi wartościami podczas budowy bloków dla konkretnego atrybutu. W dalszej kolejności proces indukcji reguł przebiega jak w oryginalnej wersji algorytmu LEM2. Prace w tym zakresie kontynuowane są do dnia dzisiejszego (np. [CGGB16, CGGBM17]) i obejmują wykorzystanie danych, w których semantyka brakujących wartości jest zadana za pomocą specjalnego symbolu (? lub \*), lub porównuje się jedną z dwóch semantyk brakujących wartości na całym zbiorze danych. Do tej grupy podejść należy też zaliczyć metodę dekompozycji [Lat03b].

### 3 Główne wyniki prac autora

W swoich badaniach koncentrowałem się na poszukiwaniu i doskonaleniu metod wnioskowania na podstawie danych zawierających obiekty z brakującymi wartościami wykorzystujących teorię zbiorów przybliżonych. Dotyczy to zarówno metod indukcji klasyfikatorów, które nie wymagają ani uzupełniania brakujących danych, ani modyfikacji teorii i algorytmów, jak i rozszerzenia teorii zbiorów przybliżonych tak, aby zapewnić wydobywanie prawdziwej, najbardziej adekwatnej semantyki brakujących wartości i zgodnie z tą semantyką przeprowadzić pełen proces wnioskowania.



Rysunek 1: Metoda dekompozycji

wania na podstawie obiektów z brakującymi wartościami atrybutów. Z uwagi na długi okres czasu w jakim powstawał cykl artykułów (od 2002r.) część wyników moich prac dotyczy również rozwiązań problemów wykraczających nieco poza ściśle zdefiniowany przedmiot badań, ale zdecydowanie związanych z wnioskowaniem na podstawie niekompletnej informacji, które powstały przy pokonywaniu trudności pojawiających się podczas pracy badawczej i licznych eksperymentów komputerowych. W szczególności, dotyczy to systemów zrównoleglania obliczeń w celu efektywnej realizacji eksperymentów, również na danych z brakującymi wartościami.

Wyniki moich prac można scharakteryzować następująco:

- Opracowanie metody indukcji klasyfikatorów na podstawie danych z brakującymi wartościami poprzez dekompozycję danych z brakującymi wartościami na regularne obszary danych kompletnych, indukcję dowolnego klasyfikatora na otrzymanych w ten sposób danych kompletnych oraz łączenie wniosków częściowych z klasyfikatorów na podzbiorach kompletnych w celu uzyskania klasyfikatora złożonego obejmującego oryginalny zbiór danych (por. rys. 1) [Lat02c, Lat02b, Lat02a, Lat03b, BLS06]. W pracach eksperymentalnych przy zastosowaniu indukcji reguł decyzyjnych metodą zbiorów przybliżonych, jak i indukcji drzew decyzyjnych (C4.5), zastosowanie metody dekompozycji dało lepsze wyniki w zakresie dokładności klasyfikacji niż w przypadku bez zastosowanie tej metody (również w porównaniu do metody C4.5). Uzyskanie takich wyników było możliwe przy zespole kilkukilkudziesięciu klasyfikatorów częściowych, a złożoność klasyfikatorów, mierzona średnią sumą liczby reguł lub liczby liści, była czasami mniejsza lub tylko kilkukrotnie większa przy zastosowaniu dekompozycji w porównaniu do klasyfikatorów bez zastosowania dekompozycji. Przewiduję, że dzięki tak ograniczonemu wzrostowi złożoności w zakresie liczby klasyfikatorów częściowych jak i sumarycznej złożoności modelu, metodę dekompozycji będzie można w przyszłości stosować do bardziej intensywnych obliczeniowo i wrażliwych na niekompletność algorytmów indukcji klasyfikatorów.
- Opracowanie metody indukcji reguł decyzyjnych generowanych z danych z brakującymi wartościami poprzez dekompozycję analogiczną do przedstawionej powyżej, w połączeniu z algorytmem skracania i łączenia reguł decyzyjnych na podstawie teorii zbiorów przybliżonych [LM04b, LM04a]. Wyniki eksperymentalne pokazały, że o ile zwiększenie dokładności klasyfikacji nie jest duże (choć istotne statystycznie), to redukcja złożoności klasyfikatora mierzona liczbą reguł decyzyjnych jest średnio trzykrotna. Zatem zaproponowany algorytm umożliwi indukcję klasyfikatora, który stanowi kompaktowy zbiór reguł obejmujących cały system informacyjny z brakującymi wartościami atrybutów, a podczas indukcji reguł zgodnych z teorią zbiorów przybliżonych uniknięto trudności uwzględniania brakujących wartości atrybutów.

- Wprowadzenie uogólnionych relacji nierozróżnialności, które umożliwiają definiowanie szerokiej klasy możliwych semantyk brakujących wartości, wraz z użytecznymi rodzinami parametrycznymi takich relacji: relacje nierozróżnialności ograniczone atrybutowo (dowolna, choć stała semantyka brakujących wartości w obrębie całego atrybutu) oraz relacje nierozróżnialności ograniczone deskryptorem (gdzie semantyka brakujących wartości musi wynikać z opisu obiektów) [Lat05, Lat04]. Parametryczna rodzina relacji ograniczonych atrybutowo została również zaimplementowana, a przeprowadzone wyniki eksperymentalne pokazują, że dzięki doborowi odpowiedniej semantyki brakujących wartości uzyskuje się poprawę dokładności klasyfikacji, która niestety często nie jest istotna statystycznie z uwagi na dużą wariancję wyników przy testach walidacji krzyżowych.
- Wykrycie braku precyzyjnego sformułowania pojęć związanych z indukcją i dopasowaniem reguł decyzyjnych do obiektów w przypadku relacji nierozróżnialności, które nie muszą spełniać warunków symetryczności relacji. O ile w przypadku np. pojęć górnej i dolnej aproksymacji takie doprecyzowanie zostało przedstawione już wcześniej (np. [Kry98, ST01a]), to w przypadku klasyfikacji/rozpoznania/dopasowania reguł do obiektu brak takiego dopasowania powodował trudności w implementacji indukcji i klasyfikacji. Precyzyjne sformułowanie co prawda na ten moment znajduje się w nieopublikowanych skrypcach autora rozprawy, ale zostało zaimplementowane w formie algorytmicznej w bibliotece Rseslib w 2006 roku i jest zamieszczone w instrukcji użytkownika [WLK18, WL18, WL19].
- Rozszerzenie zastosowania systemów informacyjnych z decyzją uogólnioną na przypadek relacji nierozróżnialności, które nie są relacją równoważności. Podobnie do powyższego punktu, indukcja reguł decyzyjnych na podstawie decyzji uogólnionej zdefiniowana została na podstawie klas abstrakcji relacji nierozróżnialności (por. [SP97, Baz98]). Przy implementacji Rseslib 3 skonstruowałem algorytm indukcji reguł z decyzją uogólnioną, który umożliwia wykorzystanie dowolnej relacji nierozróżnialności (w tym niesymetrycznej) dzięki uprzedniemu wyznaczeniu ekwiwalentów klas abstrakcji przy pomocy tranzytywnego domknięcia rozważanej relacji nierozróżnialności na badanym systemie informacyjnym [WLK18, WL18, WL19].
- Wykrycie i oszacowanie załamania „dobrej” złożoności obliczeniowej algorytmów indukcji drzew decyzyjnych w przypadku znacznej ilości brakujących wartości atrybutów, przy stosowaniu popularnej strategii rozdzielania obiektu z brakującą wartością atrybutu do wszystkich dzieci węzła testującego dany atrybut [Lat03a]. Taka strategia stosowana jest w wielu algorytmach np. w implementacjach C4.5 i większości pochodnych [KBR84, Qui89], a z uwagi na brak rozłącznego podziału przy węźle drzewa potrafi doprowadzić do wzrostu złożoności czasowej liniowo-logarytmicznej (głównie z uwagi na liczbę obiektów, ale też atrybutów) do  $n^\lambda$ , gdzie dla danych eksperymentalnych analizowanych w pracy [?] z 75% brakujących  $\lambda = 5$  (dane wizualne ligi symulacyjnej RoboCup [KLR03]).
- Opracowanie i implementacja oprogramowania dla obliczeń rozproszonych z wykorzystaniem paradygmatu obliczeń na siatce małych komputerów osobistych (grid computing). W celu umożliwienia realizacji masowych eksperymentów z wykorzystaniem algorytmów bazujących na teorii zbiorów przybliżonych, w latach 2002-2006 rozwijany był przez autora rozprawy system Distributed EXecutor (DIXER) dla biblioteki RSESlib 2.x [BLS05], a od 2006 roku rozwijany jest Simple Grid Manager dla biblioteki Rseslib 3.x [WL18, WL19]. Oprogramowanie umożliwia efektywne, w pełni skalowalne, wykorzystanie mocy obliczeniowej zarówno sieci laboratoriów komputerowych jak i dowolnej liczby oddalonych węzłów w celu zrównoleglenia masywnych eksperymentów badawczych.

- Wprowadzenie metody wczesnego wykrywania potencjalnego dużego spadku jakości klasyfikacji poprzez testy statystyczne na danych wejściowych do modelu [LG04]. W wielu przypadkach zastosowanie istniejącego modelu klasyfikatora na nowych danych jest obciążone ryzykiem dużego zmniejszenia jakości klasyfikacji w porównaniu np. do danych użytych do indukcji tego klasyfikatora. Metoda wczesnego wykrywania takiego spadku umożliwia przewidzenie takiego zjawiska (np. przed doprowadzeniem do strat materialnych). Metoda ta znalazła swoje zastosowania komercyjne (dzięki wprowadzeniu i popularyzacji przez autora rozprawy) i została zaimplementowana w oprogramowaniu SAS Institute Model Manager oraz Algolytics Model Lifetime Manager (bez udziału autora rozprawy w implementacji).

W przyszłości zamierzam kontynuować badania nad dopracowaniem pojęć teorii zbiorów przybliżonych do przypadku niesymetrycznych relacji nierozróżnialności, co będzie wymagało bardziej precyzyjnego opisanie w postaci oddzielnej pracy, a nie tylko w podręczniku do oprogramowania [WLK18]. Dodatkowo, na podstawie moich zawodowych doświadczeń rozpocząłem prace nad implementacją zupełnie odmiennego algorytmu rozdziału walidacji krzyżowej (Cross-Validation) istotnie zmniejszającym wariancję wyników. Prace te chciałbym kontynuować po złożeniu rozprawy, na którą składa się następujący cykl prac.

## 4 Wykaz prac przedstawionych w ramach rozprawy

- Wprowadzenie metody dekompozycji [Lat02a]  
Rafał Latkowski. Application of data decomposition to incomplete information systems. Mieczysław A. Kłopotek, S. T. Wierzchoń, redaktorzy, *Proceedings of the International Symposium "Intelligent Information Systems XI", June 3–6, 2002, Sopot, Poland*. Physica-Verlag, 2002.
- Rozszerzony opis metody dekompozycji z wykorzystaniem reguł decyzyjnych [Lat03b]  
Rafał Latkowski. On decomposition for incomplete data. *Fundamenta Informaticae*, 54(1):1–16, 2003.
- Metoda dekompozycji wraz z metodą łączenia reguł decyzyjnych [LM04b]  
Rafał Latkowski (50%), Michał Mikołajczyk. Data decomposition and decision rule joining for classification of data with missing values (extended version). James F. Peters, Andrzej Skowron, redaktorzy, *Transactions on Rough Sets I*, wolumen 3100 serii *Lecture Notes in Computer Science*, strony 299–320. Springer, 2004.
- Implementacja metody dekompozycji w oprogramowaniu RSES 2 [BLS06]  
Jan G. Bazan, Rafał Latkowski (60%), Marcin Szczuka. Missing template decomposition method and its implementation in rough set exploration system. *RSCTC 2006 - Japonia*, 2006.
- Załamanie złożoności obliczeniowej drzew typu C4.5 (J48) w przypadku dużej ilości brakujących wartości [Lat03a]  
Rafał Latkowski. High computational complexity of the decision tree induction with many missing attribute values. Ludwik Czaja, redaktor, *Proceedings of Concurrency, Specification and Programming. CS&P'2003, Czarna, September 25-27, Volume 2.*, strony 318–325. Zakłady Graficzne UW, 2003.

- Wprowadzenie uogólnionych relacji nierozróżnialności [Lat05]  
Rafał Latkowski. Flexible indiscernibility relations for missing attribute values. *Fundamenta Informaticae*, 67(1-3):131–147, 2005.
- Implementacja systemu do rozpraszania obliczeń dla RSES 2 [BLS05]  
Jan G. Bazan, Rafał Latkowski (80%), Marcin Szczuka. DIXER — distributed executor for rough set exploration system. Dominik Slezak, J. T. Yao, James F. Peters, Wojciech Ziarko, X. Hu, redaktorzy, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 10th International Conference, RSFDGrC 2005, Regina, Canada, August 31 - September 3, 2005, Proceedings, Part II*, wolumen 3642 serii LNCS, strony 39–47. Springer, 2005.
- Implementacja algorytmów rough-set'owych dla dowolnej relacji nierozróżnialności oraz systemu do rozpraszania obliczeń dla Rseslib 3 [WL19]  
Arkadiusz Wojna, Rafał Latkowski (50%). Rseslib 3: Library of rough set and machine learning methods with extensible architecture. James F. Peters, Andrzej Skowron, redaktorzy, *Transactions on Rough Sets XXI*, wolumen 10810 serii *Lecture Notes in Computer Science*. Springer, 2019.

## 5 Omówienie głównych wyników prac wchodzących w skład rozprawy

### 1. Metoda Dekompozycji

- (a) W pracy [Lat02a] wprowadzono metodę dekompozycji oraz opisano formalizację dekompozycji danych poprzez znajdowanie wzorców wypełnienia. Wspomniano o wynikach prac eksperymentalnych dotyczących niesatysfakcjonujących rezultatów, gdy metodą scalania wyników jest głosowanie oraz, że wzorce o podobnych wymiarach generują bardzo zróżnicowane jakościowo wyniki, co przyczyniło się do poszukiwania zarówno alternatywnych funkcji oceny wzorca, jak i alternatywnych algorytmów wyszukiwania wzorców. W pracy opisano wyniki eksperymentalne, w szczególności porównano wyszukiwanie wzorców za pomocą algorytmu dokładnego z algorytmem genetycznym (dokumentując rozwój algorytmu genetycznego do poziomu, gdzie osiąga zadowalającą jakość w stosunku do algorytmu dokładnego) oraz przedstawiono wyniki jakości klasyfikacji z wykorzystaniem algorytmu C4.5 (wraz z jego wbudowaną metodą radzenia sobie z brakującymi wartościami [Qui89]) pokazujące przewagę skutecznej, ale kosztownej obliczeniowo, funkcji oceny wzorca bazującej na skuteczności klasyfikacji w ramach podsystemu informacyjnego.
- (b) W artykule [Lat03b], dzięki jego większej objętości okazało się możliwe szczegółowe opisanie działania metody dekompozycji, zebranie wszystkich postępów w badaniach opisywanych w pracach [Lat02b, Lat02c] oraz przedstawienie wyników przeprowadzonych prac eksperymentalnych zarówno z wykorzystaniem drzew decyzyjnych (C4.5), reguł decyzyjnych (RSES 2), jak i przy różnej zawartości brakujących wartości. Odnośnie zagadnienia wyszukiwania wzorca wspomniano o wynikach różnych strategii usuwania obiektów w algorytmie generowania pokryć, przypomniano, że jest to problem NP-trudny oraz przedstawiono opis ulepszonego algorytmu genetycznego, który zapewnia szybkość i pewną zbieżność do dobrych wzorców, co jest kluczowe z uwagi na wykorzystanie potencjalnie kosztownych funkcji oceny. Poprzednie wyniki eksperymentów bazowały tylko na drzewach decyzyjnych C4.5, miało to swoje uargumentowanie w porównaniu bardzo skutecznej metody radzenia sobie z brakującymi warto-

ściami (por. [KBR84, Qui89]) z metodą dekompozycji. W omawianej pracy znalazło się już porównanie eksperymentalne wykorzystujące reguły decyzyjne bazujące na zbiorach przybliżonych (tj. optymalne reguły decyzyjne generowane z najkrótszego reduktu), dwie alternatywne, tańsze funkcje oceny wzorca, oraz tzw. eksperymenty kalibrujące pokazujące relatywną niewrażliwość jakości klasyfikacji metody dekompozycji na procent brakujących wartości w danych.

- (c) W rozdziale [LM04b] przedstawiono metodę będącą wynikiem kombinacji metody dekompozycji oraz metody łączenia (klastrowania) reguł decyzyjnych. Dzięki łączeniu reguł z podsystemów informacyjnych uzyskujemy spójny, skuteczny i kompaktowy (tj. o małym rozmiarze) klasyfikator, który umożliwi klasyfikację danych z brakującymi wartościami bez konieczności modyfikacji algorytmu indukcji reguł do przypadku brakujących wartości. Jako ciekawostkę można przytoczyć, że nawet skracanie reguł dla współczynnika  $\alpha < 0.5$  generuje bardzo dobre wyniki skuteczności klasyfikacji przy niewielkich zbiorach reguł.
- (d) W pracy [BLS06] przedstawiono udostępnioną publicznie implementację metody dekompozycji w oprogramowaniu RSES 2. Implementacja udostępnia łącznie 9 parametryzowanych funkcji oceny wzorca.
- (e) W pracy [Lat03a] opisano, wykryte przy okazji eksperymentów z wykorzystaniem C4.5 (Weka J48 [?]), załamanie dobrej złożoności obliczeniowej drzew decyzyjnych z uwagi na złamanie zasady „dziel i rządź”, szczególnie przy dużej zawartości brakujących wartości w danych.

## 2. Uogólnione relacje nierozróżnialności

- (a) W artykule [Lat05] wprowadzono uogólnione relacje nierozróżnialności, które umożliwiają jawną reprezentację różnorodnych semantyk brakujących wartości bezpośrednio w ramach teorii zbiorów przybliżonych. Wprowadzone są też dwie rodziny parametryczne takich relacji (ograniczone atrybutowo i ograniczone deskryptorowo), które umożliwiają zastosowanie metod algorytmicznych w celu poszukiwania najbardziej dopasowanej semantyki brakujących wartości (czyli uogólnionej relacji nierozróżnialności) do analizowanych danych. Takie relacje mają też moc wyrazu niektórych algorytmów uzupełniania brakujących wartości, bez konieczności faktycznej modyfikacji danych. Zaprezentowano również podejście do generowania optymalnych relacji nierozróżnialności w ramach specjalnej pod-rodziny relacji ograniczonych atrybutowo wraz z wynikami eksperymentalnymi tego podejścia.
- (b) W pracy [WL19], oprócz innych aspektów stale rozwijanej biblioteki Rseslib 3, zawarto również implementację rozszerzonych relacji nierozróżnialności. W wyniku prac autora niniejszej rozprawy w bibliotece Rseslib 3 zaimplementowane są obiektowe abstrakty relacji nierozróżnialności i innych pojęć z zakresu teorii zbiorów przybliżonych, algorytmy generowania reduktów (przy wykorzystaniu implikantów pierwszych) oraz indukcji optymalnych reguł decyzyjnych umożliwiające wykorzystanie dowolnej uogólnionej relacji nierozróżnialności. Z uwagi na w pełni obiektową implementację pojęć teorii zbiorów przybliżonych w tej bibliotece, każda zaimplementowana metoda wykorzystująca relacje nierozróżnialności umożliwia wykorzystanie uogólnionych relacji nierozróżnialności.

## 3. Oprogramowanie do rozpraszania obliczeń



- (a) W pracy [BLS05] przedstawiono oprogramowanie DIXER, będące efektem prac autora niniejszej rozprawy, które umożliwia rozpraszanie obliczeń na dowolnej liczbie komputerów (tzw. grid-computing). Dzięki temu oprogramowaniu możliwe było przeprowadzenie wspomnianych wcześniej tak licznych eksperymentów w zakresie metody dekompozycji, z wykorzystaniem kilkunastu zestawów danych, powtórzeń walidacji krzyżowych (Cross-Validation) i różnych ustawień parametrów i wersji metody.
- (b) W pracy [WL19], oprócz innych aspektów stale rozwijanej biblioteki Rseslib 3, przedstawione jest również oprogramowanie Simple Grid Manager, będące efektem prac autora niniejszej rozprawy, które umożliwia efektywne rozpraszanie obliczeń na dowolnej liczbie komputerów podczas przeprowadzania eksperymentów z wykorzystaniem biblioteki Rseslib 3. Dzięki temu oprogramowaniu przeprowadzono również eksperymenty z wykorzystaniem uogólnionych relacji nierozróżnialności.

## Literatura

- [Baz98] Jan G. Bazan. *Metody wnioskowań aproksymacyjnych dla syntezy algorytmów decyzyjnych*. Praca doktorska, Uniwersytet Warszawski, Wydział Matematyki, Informatyki i Mechaniki, 1998.
- [BFOS84] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, Charles J. Stone. *Classification and Regression Trees*. Wadsworth International Group, 1984.
- [Bin00] Grzegorz Bińczak. *Charakteryzacja klas algebr częściowych definiowalnych przez słabe równości*. Praca doktorska, Uniwersytet Warszawski, 2000.
- [BLS05] Jan G. Bazan, Rafał Latkowski, Marcin Szczuka. DIXER — distributed executor for rough set exploration system. Dominik Slezak, J. T. Yao, James F. Peters, Wojciech Ziarko, X. Hu, redaktorzy, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 10th International Conference, RSFDGrC 2005, Regina, Canada, August 31 - September 3, 2005, Proceedings, Part II*, wolumen 3642 serii LNCS, strony 39–47. Springer, 2005.
- [BLS06] Jan G. Bazan, Rafał Latkowski, Marcin Szczuka. Missing template decomposition method and its implementation in rough set exploration system. *RSCTC 2006 - Japonia*, 2006.
- [Bur86] P. Burmeister. *A model — theoretic oriented approach to partial algebras*. Akademie-Verlag, Berlin, 1986, wydanie elektroniczne 2002.
- [CG14] Patrick G. Clark, Jerzy W. Grzymała-Busse. A comparison of two versions of the MLEM2 rule induction algorithm extended to probabilistic approximations. Chris Cornelis, Marzena Kryszkiewicz, Dominik Ślęzak, Ernestina Menasalvas Ruiz, Rafael Bello, Lin Shang, redaktorzy, *Rough Sets and Current Trends in Computing - 9th International Conference, RSCTC 2014, Granada and Madrid, Spain, July 9-13, 2014. Proceedings*, wolumen 8536 serii LNCS, strony 109–119. Springer, 2014.
- [CGGB16] P. G. Clark, C. Gao, J. W. Grzymała-Busse. Rule set complexity for incomplete data sets with many attribute-concept values and „do not care” conditions. Patrz Flores i in. [FGJ<sup>+</sup>16], strony 65–74.

- [CGGBM17] Patrick G. Clark, Cheng Gao, J. W. Grzymała-Busse, Teresa Mroczek. Characteristic sets and generalized maximal consistent blocks in mining incomplete data. Patrz Polkowski i in. [PYA<sup>+</sup>17], strony 477–486.
- [Cod70] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, 1970.
- [DO02] Stéphane Demri, Ewa Orlowska. *Incomplete Information: Structure, Inference, Complexity*. Monographs in Theoretical Computer Science. An EATCS Series. Springer, 2002.
- [FGJ<sup>+</sup>16] Víctor Flores, Fernando Gomide, Andrzej Janusz, Claudio Meneses, Duoqian Miao, Georg Peters, Dominik Ślęzak, Guoyin Wang, Richard Weber, Yiyu Yao, redaktorzy. *International Joint Conference, IJCRS 2016, Santiago de Chile, Chile, October 7–11, 2016, Proceedings*, wolumen 9920 serii LNCS. Springer, 2016.
- [FH02] Yoshikazu Fujikawa, Tu Bao Ho. Cluster-based algorithms for dealing with missing values. Ming-Shan Cheng, Philip S. Yu, Bing Liu, redaktorzy, *Advances in Knowledge Discovery and Data Mining, PAKDD 2002*, wolumen 2336 serii LNCS, strony 549–554. Springer, 2002.
- [Fri77] Jerome H. Friedman. A recursive partitioning decision rule for non-parametric classification. *IEEE Transactions on Computer Science*, 26:404–408, 1977.
- [GB13] J. W. Grzymała-Busse. Generalized probabilistic approximations. James F. Peters, Andrzej Skowron, Sheela Ramanna, Zbigniew Suraj, Xin Wang, redaktorzy, *Transactions on Rough Sets XVI*, strony 1–16. Springer, 2013.
- [GBC16] Ian Goodfellow, Yoshua Bengio, Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [GBGBG99] Jerzy W. Grzymała-Busse, Witold J. Grzymała-Busse, Linda K. Goodwin. A closest fit approach to missing attribute values in preterm birth data. Patrz Zhong i in. [ZSO99], strony 405–413.
- [GBH01] Jerzy W. Grzymała-Busse, Ming Hu. A comparison of several approaches to missing attribute values in data mining. Patrz Ziarko i Yao [ZY01], strony 378–385.
- [GBW97] Jerzy W. Grzymała-Busse, Arthur Y. Wang. Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values. *Proceedings of 5th Workshop on Rough Sets and Soft Computing (RSSC'97) at the 3rd Joint Conference on Information Sciences*, strony 69–72, Research Triangle Park (NC, USA), 1997.
- [GMS99] Salvatore Greco, Benedetto Matarazzo, Roman Słowiński. Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems. Patrz Zhong i in. [ZSO99], strony 146–157.
- [GMS00] Salvatore Greco, Benedetto Matarazzo, Roman Słowiński. Rough sets processing of vague information using fuzzy similarity relations. C. S. Caldue, G. Paun, redaktorzy, *Finite vs. infinite: contribution to an eternal dilemma*, strony 149–173, Berlin, 2000. Springer.

- [GMSZ99] Salvatore Greco, Benedetto Matarazzo, Roman Słowiński, S. Zanakis. Rough set analysis of information tables with missing values. *Proceedings of 5th International Conference Decision Sciences Institute, July 4–7, Athens-Greece*, wolumen 2, strony 1359–1362, 1999.
- [Gra12] John W. Graham. *Missing Data, Analysis and Design*. Springer, 2012.
- [GS10] Iffat A. Gheyas, Leslie S. Smith. A neural network-based framework for the reconstruction of incomplete data sets. *Neurocomputing*, 73(16-18):3039–3065, 2010.
- [HB14] Steven G. Heeringa, Patricia Berglund. *Multiple Imputation of Missing Data Using SAS*. SAS Institute, 2014.
- [HTF01] Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, 2001.
- [HWY17] Jun Hu, Kai Wang, Hong Yu. Attribute reduction on distributed incomplete decision information system. Patrz Polkowski i in. [PYA<sup>+</sup>17], strony 289–305.
- [IL84] Tomasz Imieliński, Witold Lipski, Jr. Incomplete information in relational databases. *Journal of the ACM*, 31(4):761–791, 1984.
- [KBR84] I. Kononenko, I. Bratko, E. Roškar. Experiments in automatic learning of medical diagnostic rules. Technical report, Jozef Stefan Institute, Ljubljana, 1984.
- [KLR03] Gal A. Kaminka, Pedro U. Lima, Raúl Rojas, redaktorzy. *RoboCup 2002: Robot Soccer World Cup VI*, wolumen 2752 serii LNCS. Springer, 2003.
- [KPPS99] Jan Komorowski, Zdzisław Pawlak, Lech Polkowski, Andrzej Skowron. Rough sets: A tutorial. Sankar K. Pal, Andrzej Skowron, redaktorzy, *Rough Fuzzy Hybridization. A New Trend in Decision Making*, strony 3–98, Singapore, 1999. Springer.
- [Kry98] Marzena Kryszkiewicz. Properties of incomplete information systems in the framework of rough sets. Lech Polkowski, Andrzej Skowron, redaktorzy, *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, strony 422–450. Physica-Verlag, 1998.
- [Laa18] Seppo Laaksonen. *Survey Methodology and Missing Data*. Springer, 2018.
- [Lat02a] Rafał Latkowski. Application of data decomposition to incomplete information systems. Mieczysław A. Kłopotek, S. T. Wierzchoń, redaktorzy, *Proceedings of the International Symposium “Intelligent Information Systems XI”, June 3–6, 2002, Sopot, Poland*. Physica-Verlag, 2002.
- [Lat02b] Rafał Latkowski. Incomplete data decomposition for classification. James J. Alpigini, James F. Peters, Andrzej Skowron, Ning Zhong, redaktorzy, *Rough Sets and Current Trends in Computing, Third International Conference, RSCTC 2002, Malvern, PA, USA, October 14-16, 2002, Proceedings*, wolumen 2475 serii LNCS, strony 413–420. Springer, 2002.
- [Lat02c] Rafał Latkowski. On decomposition for incomplete data. Hans-Dieter Burkhard, Ludwik Czaja, Gabriela Lindemann, Andrzej Skowron, Peter Starke, redaktorzy, *Concurrency, Specification and Programming. CS&P’2002, Berlin, October 7-9, Volume 2.*, strony 217–226. HU Informatik-Berichte, 2002.

- [Lat03a] Rafał Latkowski. High computational complexity of the decision tree induction with many missing attribute values. Ludwik Czaja, redaktor, *Proceedings of Concurrency, Specification and Programming. CS&P'2003, Czarna, September 25-27, Volume 2.*, strony 318–325. Zakłady Graficzne UW, 2003.
- [Lat03b] Rafał Latkowski. On decomposition for incomplete data. *Fundamenta Informaticae*, 54(1):1–16, 2003.
- [Lat04] Rafał Latkowski. On indiscernibility relations for missing attribute values. et al. G. Lindemann, redaktor, *Concurrency, Specification and Programming. CS&P'2004, Volume 2. Informatik-Bericht Nr. 170.*, strony 330–335. Humboldt University, 2004.
- [Lat05] Rafał Latkowski. Flexible indiscernibility relations for missing attribute values. *Fundamenta Informaticae*, 67(1-3):131–147, 2005.
- [LG04] Rafał Latkowski, Cezary Glowinski. Classifier monitoring using statistical tests. Barbara Dunin-Keplicz, Andrzej Jankowski, Andrzej Skowron, Marcin S. Szczuka, redaktorzy, *Monitoring, Security, and Rescue Techniques in Multiagent Systems, MSRAS 2004, Plock, Poland, June 7-9, 2004*, wolumen 28 serii *Advances in Soft Computing*, strony 493–500. Springer, 2004.
- [Lip79] Witold Jr. Lipski. On semantic issues connected with incomplete information databases. *ACM Transactions on Database Systems*, 4(3):262–296, 1979.
- [LM04a] Rafał Latkowski, Michał Mikołajczyk. Data decomposition and decision rule joining for classification of data with missing values. Shusaku Tsumoto, Roman Słowiński, Henryk Jan Komorowski, Jerzy W. Grzymała-Busse, redaktorzy, *Rough Sets and Current Trends in Computing, RSCTC 2004*, wolumen 3066 serii *LNCS*, strony 254–263. Springer, 2004.
- [LM04b] Rafał Latkowski, Michał Mikołajczyk. Data decomposition and decision rule joining for classification of data with missing values (extended version). James F. Peters, Andrzej Skowron, redaktorzy, *Transactions on Rough Sets I*, wolumen 3100 serii *LNCS*, strony 299–320. Springer, 2004.
- [LR87] R. J. A. Little, D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley and Sons, 1987.
- [LWY18] Tong-Jun Li, Wei-Zhi Wu<sup>1</sup>, Xiao-Ping Yang<sup>1</sup>. Tolerance relations and rough approximations in incomplete contexts. Patrz Nguyen i in. [NHLP18], strony 533–545.
- [Mar08] Benjamin M. Marlin. *Missing Data Problems in Machine Learning*. Praca doktorska, University of Toronto, 2008.
- [MFK<sup>+</sup>04] Geert Molenberghs, Garrett Fitzmaurice, Michael G. Kenward, Anastasios Tsiatis, Geert Verbeke. *Handbook of Missing Data Methodology*. Handbooks of Modern Statistical Methods. Chapman & Hall / CRC Press, 2004.
- [MGR17] D. P. Mesquita, J. P. Gomes, L. R. Rodrigues. Extreme learning machines for datasets with missing values using the unscented transform. *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*, strony 85–90. IEEE, 2017.

- [MH16] Yiyu Yao Mengjun Hu. Definability in incomplete information tables. Patrz Flores i in. [FGJ<sup>+</sup>16], strony 177–186.
- [NHLP18] Hung Son Nguyen, Quang-Thuy Ha, Tianrui Li, Malgorzata Przybyla-Kasperek, redaktorzy. *Rough Sets - International Joint Conference, IJCRS 2018, Quy Nhon, Vietnam, August 20-24, 2018, Proceedings*, wolumen 11103 serii LNCS. Springer, 2018.
- [Paw83] Zdzisław Pawlak. *Systemy Informacyjne, Podstawy teoretyczne*. WNT, 1983.
- [Paw91] Zdzisław Pawlak. *Rough sets: Theoretical aspects of reasoning about data*. Kluwer, Dordrecht, 1991.
- [PR07] Grzegorz Protaziuk, Henryk Rybiński. Discovering association rules in incomplete transactional databases. *Transactions on Rough Sets*, VI:308–328, 2007.
- [PSŻ95] Lech Polkowski, Andrzej Skowron, Jan M. Żytkow. Tolerance based rough sets. Tsau Young Lin, A. M. Wildberger, redaktorzy, *Soft Computing*, strony 55–58. San Diego Simulation Councils Inc., 1995.
- [PYA<sup>+</sup>17] Lech Polkowski, Yiyu Yao, Piotr Artiemjew, Davide Ciucci, Dun Liu, Dominik Ślęzak, Beata Zielosko, redaktorzy. *International Joint Conference, IJCRS 2017, Olsztyn, Poland, July 3–7, 2017, Proceedings, Part I*, wolumen 10313 serii LNCS. Springer, 2017.
- [Qui89] J. Ross Quinlan. Unknown attribute values in induction. Alberto Maria Segre, redaktor, *Proceedings of the Sixth International Machine Learning Workshop*, strony 31–37. Morgan Kaufmann, 1989.
- [Rub87] D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York, 1987.
- [Rzą09] Wojciech Rząsa. *Globalne i lokalne przybliżenia zbiorów*. Praca doktorska, Instytut Podstaw Informatyki Polskiej Akademii Nauk, 2009.
- [SNŚ10] Hiroshi Sakai, Michinori Nakata, Dominik Ślęzak. Rule generation in lipski’s incomplete information databases. Marcin S. Szczuka, Marzena Kryszkiewicz, Sheela Ramanna, Richard Jensen, Qinghua Hu, redaktorzy, *RSCTC*, wolumen 6086 serii *Lecture Notes in Computer Science*, strony 376–385. Springer, 2010.
- [SP97] Andrzej Skowron, Lech Polkowski. Decision algorithms: A survey of rough set - theoretic methods. *Fundam. Inform.*, 30(3/4):345–358, 1997.
- [SST<sup>+</sup>18] Marek Smieja, Lukasz Struski, Jacek Tabor, Bartosz Zielinski, Przemyslaw Spurek. Processing of missing data by neural networks. Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, Roman Garnett, redaktorzy, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, strony 2724–2734, 2018.
- [ST99] Jerzy Stefanowski, Alexis Tsoukiàs. On the extension of rough sets under incomplete information. Patrz Zhong i in. [ZSO99], strony 73–81.

- [ST01a] Jerzy Stefanowski, Alexis Tsoukiàs. Incomplete information tables and rough classification. *International Journal of Computational Intelligence*, 17(3):545–566, August 2001.
- [ST01b] Jerzy Stefanowski, Alexis Tsoukiàs. Valued tolerance and decision rules. Patrz Ziarko i Yao [ZY01], strony 212–219.
- [Sta96] Bożena Staruch. *Algebraiczna charakteryzacja klas algebr częściowych definiowanych przez silne równości*. Praca doktorska, Uniwersytet Warszawski, 1996.
- [SV00] Roman Słowiński, D. Vanderpooten. A generalized definition of rough approximations based on similarity. *IEEE Transactions on Data and Knowledge Engineering*, 12:331–336, 2000.
- [UMN<sup>+</sup>15] Seiki Ubukata<sup>1</sup>, Taro Miyazaki, Akira Notsu, Katsuhiko Honda, Masahiro Inuiguchi. An ensemble learning approach based on missing-valued tables. Yiyu Yao, Qinghua Hu, Hong Yu, J. W. Grzymała-Busse, redaktorzy, *15th International Conference, RSFDGrC 2015, Tianjin, China, November 20–23, 2015, Proceedings*, wolumen 9437 serii LNCS, strony 310–321. Springer, 2015.
- [vB18] Stef van Buuren. *Flexible Imputation of Missing Data, Second Edition*. CRC/Chapman & Hall, 2018.
- [WF00] Ian H. Witten, Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, 2000.
- [WI00] Sholom M. Weiss, Nitin Indurkha. Lightweight rule induction. *Proceedings of the International Conference on Machine Learning ICML'2000*, 2000.
- [WL18] Arkadiusz Wojna, Rafał Latkowski. Rseslib 3: Open source library of rough set and machine learning methods. Patrz Nguyen i in. [NHLP18], strony 162–176.
- [WL19] Arkadiusz Wojna, Rafał Latkowski. Rseslib 3: Library of rough set and machine learning methods with extensible architecture. James F. Peters, Andrzej Skowron, redaktorzy, *Transactions on Rough Sets XXI*, wolumen 10810 serii LNCS. Springer, 2019.
- [WLK18] Arkadiusz Wojna, Rafał Latkowski, Łukasz Kowalski. *RSESLIB User Guide*, 2009–2018.
- [ZSO99] Ning Zhong, Andrzej Skowron, Setsuo Ohsuga, redaktorzy. *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, 7th International Workshop, RSFDGrC '99, Yamaguchi, Japan, November 9-11, 1999, Proceedings*, wolumen 1711 serii Lecture Notes in Computer Science. Springer, 1999.
- [ZY01] Wojciech Ziarko, Y. Y. Yao, redaktorzy. *Rough Sets and Current Trends in Computing, Second International Conference, RSCTC 2000 Banff, Canada, October 16-19, 2000, Revised Papers*, wolumen 2005 serii LNCS. Springer, 2001.