

Autoreferat

1 Imię i Nazwisko: Małgorzata Przybyła-Kasperek

Instytut Informatyki, Uniwersytet Śląski,
ul. Będzińska 39, 41-200 Sosnowiec
Email: malgorzata.przybyla-kasperek@us.edu.pl
<https://lamp.ii.us.edu.pl/~mpk/>

2 Posiadane dyplomy

- Tytuł magistra matematyki, specjalność informatyczna; dyplom ukończenia wyższych studiów magisterskich; Uniwersytet Śląski, Wydział Matematyki, Fizyki i Chemii, kierunek Matematyka; Katowice, rok 2006; praca magisterska pt. „Twierdzenie Shapley’a” napisana pod kierunkiem prof. dr hab. Władysława Kulpy.
- Stopień doktora nauk technicznych w zakresie informatyki; dyplom ukończenia studiów doktoranckich; Uniwersytet Śląski, Wydział Informatyki i Nauki o Materiałach; Katowice, rok 2011; rozprawa doktorska pt. „Wieloagentowy system decyzyjny – analiza konfliktów” napisana pod kierunkiem prof. dr hab. inż. Alicji Wakulicz-Deja; rozprawa uzyskała wyróżnienie oraz Nagrodę Naukową Rektora Uniwersytetu Śląskiego w Katowicach dla autorów najlepszych prac doktorskich.

3 Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

- 01.10.2011 – nadal — adiunkt: Instytut Informatyki, Wydział Informatyki i Nauki o Materiałach, Uniwersytet Śląski, Sosnowiec
- 01.10.2009 – 30.09.2011 — asystent: Instytut Matematyki, Wydział Matematyki, Fizyki i Chemii, Uniwersytet Śląski, Katowice

4 Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.):

Jako podstawowe osiągnięcie naukowe przedstawiam cykl dwunastu publikacji monotematycznych, na które składają się: 8 wysokopunktowanych artykułów [A1–A8] z listy JCR oraz 4 artykuły [B1–B4] z materiałów konferencyjnych sygnowanych przez wydawnictwo

Springer i indeksowanych w bazie Web of Science. Wszystkie wskazane artykuły zostały zrealizowane i opublikowane po uzyskaniu przeze mnie tytułu doktora. Do osiągnięcia podałam najważniejsze publikacje, chociaż większość mojego dorobku jest związana z przedstawianą tematyką.

a) tytuł osiągnięcia naukowego

Rozproszony system podejmowania decyzji – metody analizy konfliktów i negocjacji

b) (autor/autorzy, tytuł/tytuły publikacji, rok wydania, nazwa wydawnictwa, recenzenci wydawniczy)

Publikacje w czasopismach posiadających współczynnik Impact Factor: (mój procentowy udział w tych pracach przedstawiam w nawiasach kwadratowych)

- [A1] Przybyła–Kasperek M., Wakulicz–Deja A., Application of reduction of the set of conditional attributes in the process of global decision–making, *Fundamenta Informaticae* 122 (4), 327–355, (2013), <http://dx.doi.org/10.3233/FI-2013-793> [80%]
- [A2] Przybyła–Kasperek M., Wakulicz–Deja A., Global decision–making in multi–agent decision–making system with dynamically generated disjoint clusters. *Applied Soft Computing*, 40, 603–615 (2016), <http://dx.doi.org/10.1016/j.asoc.2015.12.016> [90%]
- [A3] Przybyła–Kasperek M., Wakulicz–Deja A., Global decision–making system with dynamically generated clusters, *Information Sciences Volume 270*, 172–191 (2014), <http://dx.doi.org/10.1016/j.ins.2014.02.076> [85%]
- [A4] Przybyła–Kasperek M., Wakulicz–Deja A., A dispersed decision–making system – The use of negotiations during the dynamic generation of a system’s structure, *Information Sciences, Volume 288*, 194–219 (2014), <http://dx.doi.org/10.1016/j.ins.2014.07.032> [90%]
- [A5] Przybyła–Kasperek M., Wakulicz–Deja A., The strength of coalition in a dispersed decision support system with negotiations. *European Journal of Operational Research*, 252, 947–968 (2016), <http://dx.doi.org/10.1016/j.ejor.2016.02.008> [85%]
- [A6] Przybyła–Kasperek M., Selected Methods of Combining Classifiers, when Predictions are Stored in Probability Vectors, in a Dispersed Decision–making System. *Fundamenta Informaticae* 147 (2-3), 353–370 (2016), DOI 10.3233/FI-2016-1402 [100%]
- [A7] Przybyła–Kasperek M., Wakulicz–Deja A., Dispersed decision–making system with fusion methods from the rank level and the measurement level – a comparative study, *Information Systems* 69, 124–154 (2017), <https://doi.org/10.1016/j.is.2017.05.002> [90%]
- [A8] Przybyła–Kasperek M., Wakulicz–Deja A., Comparison of fusion methods from the abstract level and the rank level in a dispersed decision–making system, *International Journal of General Systems* 46 (4), 386–413 (2017), <https://doi.org/10.1080/03081079.2017.1314276> [85%]

Pozostałe publikacje:

- [B1] Przybyła–Kasperek M., Wakulicz–Deja A., Application of decision rules, generated on the basis of local knowledge bases, in the process of global decision–making, Intelligent Decision Technologies Smart Innovation, Systems and Technologies, Vol. 1, Part 2, 375–388, ISBN 978-3-642-29976-6, Springer, (2012), http://dx.doi.org/10.1007/978-3-642-29977-3_38 [85%]
- [B2] Przybyła–Kasperek M., Global Decisions Taking Process, Including the Stage of Negotiation, on the Basis of Dispersed Medical Data, S. Kozielski et al. (Eds.): BDAS 2014, CCIS Communications in Computer and Information Science 424, ISSN 1865-0929, 290–299, (2014), https://doi.org/10.1007/978-3-319-06932-6_28 [100%]
- [B3] Przybyła–Kasperek M., Dispersed decision–making system with selected fusion methods from the measurement level – case study with medical data, Annals of Computer Science and Information Systems, Volume 8, Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, 129–136 (2016) <http://dx.doi.org/10.15439/2016F30> [100%]
- [B4] Przybyła–Kasperek M., Methods Based on Pawlak’s Model of Conflict Analysis – Medical Applications. Rough Sets – International Joint Conference, IJCRS 2017, Olsztyn, Poland, July 3-7, 2017, Proceedings, Part I, Lecture Notes in Computer Science, 249–262, ISBN: 978-3-319-60836-5, Springer, (2017), https://doi.org/10.1007/978-3-319-60837-2_21 [100%]

c) omówienie celu naukowego wymienionych prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania

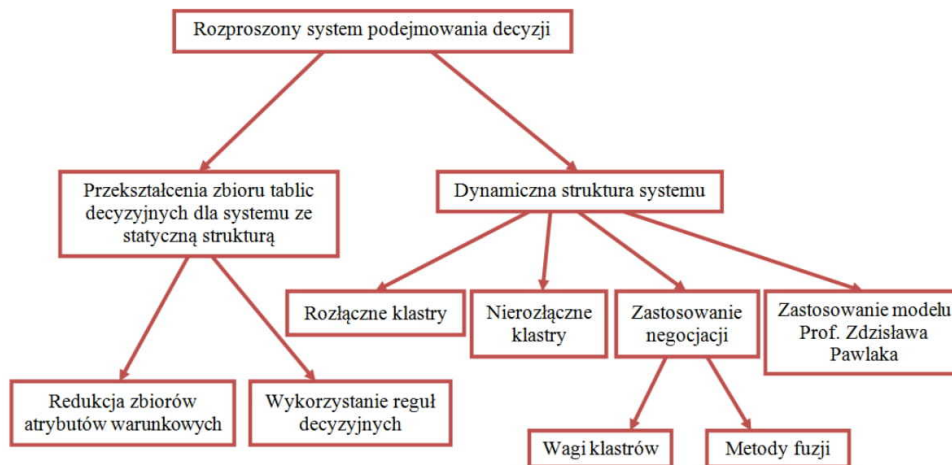
4.1 Cel naukowy

W czasie mojej działalności naukowej zajmowałam się systemami decyzyjnymi, czyli systemami komputerowymi wykorzystującymi wiedzę i wnioskowanie w procesie podejmowania decyzji. W systemach wspomagania decyzji stosowana jest wiedza zapisana w postaci bazy wiedzy. W mojej pracy rozpatrywałam problem, który dotyczył złożonych baz wiedzy, w których istotna jest możliwość współpracy baz lokalnych dla wypracowania wspólnej decyzji. Zakładałam, że wiedza zapisana w bazie wiedzy reprezentowana jest w postaci obiektów, których właściwości opisywane są za pomocą atrybutów warunkowych oraz atrybutów decyzyjnych. Zatem wiedza zapisana jest w postaci tablic decyzyjnych. System decyzyjny działający w oparciu o kilka lokalnych tablic decyzyjnych podejmuje decyzje dla zbioru przesłanek określonych na atrybutach warunkowych wszystkich wykorzystywanych tablic. Takie ogólne ujęcie może być w szczególności sprowadzone do zagadnienia klasyfikacji na podstawie wiedzy zgromadzonej w zbiorze lokalnych tablic decyzyjnych. Wiedzę, która zgromadzona jest w tablicach lokalnych nazywamy wiedzą rozproszoną.

Celem naukowym wymienionych prac był rozwój teorii i zastosowań metod klasyfikacji na podstawie wiedzy zgromadzonej w zbiorze lokalnych tablic decyzyjnych. W mojej pracy wykorzystuję pojęcie decyzji lokalnych, które wynikają z lokalnych tablic decyzyjnych. Z uwagi na to, że na podstawie tablic lokalnych dokonywana jest pewna klasyfikacja (podejmowane są decyzje lokalne), rozważany system może być postrzegany jako pewnego

rodzaju model zespołu klasyfikatorów. W systemie zarówno wiedza zgromadzona w tablicach lokalnych jak i decyzje lokalne mogą być niespójne. Dlatego zajmowałam się problematyką analizy konfliktów, rozpoznawaniem relacji oraz generowaniem koalicji lokalnych tablic decyzyjnych.

Graficzna prezentacja schematu prowadzonych prac została podana na rysunku 1.



Rysunek 1: Schemat prowadzonych prac

Szczegółowe cele prowadzonych prac były następujące:

- poprawa jakości klasyfikacji oraz efektywności rozproszonego systemu ze statyczną strukturą poprzez zastosowanie redukcji zbiorów atrybutów warunkowych lokalnych tablic decyzyjnych (wyniki badań w tym zakresie opisano w rozdziale 4.4.1 oraz w artykule [A1]),
- adaptacja rozproszonego systemu ze strukturą statyczną w celu umożliwienia klasyfikacji na podstawie reguł decyzyjnych (wyniki badań w tym zakresie opisano w rozdziale 4.4.2 oraz w artykule [B1]),
- konstrukcja rozproszonego systemu ze strukturą dynamiczną i rozłącznymi klastrami w celu analizy i identyfikacji relacji zachodzących pomiędzy klasyfikatorami bazowymi oraz poprawy jakości klasyfikacji (wyniki badań w tym zakresie opisano w rozdziale 4.4.3 oraz w artykule [A2]),
- konstrukcja rozproszonego systemu ze strukturą dynamiczną i nierozłącznymi klastrami w celu analizy i identyfikacji relacji zachodzących pomiędzy klasyfikatorami bazowymi oraz poprawy jakości klasyfikacji (wyniki badań w tym zakresie opisano w rozdziale 4.4.4 oraz w artykule [A3]),
- konstrukcja rozproszonego systemu ze strukturą dynamiczną i etapem negocjacji w celu lepszej identyfikacji relacji zachodzących pomiędzy klasyfikatorami bazowymi (wyniki badań w tym zakresie opisano w rozdziale 4.4.5 oraz w artykułach [A4, B2]),

- poprawa jakości klasyfikacji rozproszonego systemu ze strukturą dynamiczną i etapem negocjacji poprzez zastosowanie wag klastrów (wyniki badań w tym zakresie opisano w rozdziale 4.4.6 oraz w artykule [A5]),
- zastosowanie metod fuzji klasyfikatorów bazowych znanych z literatury w rozproszonym systemie ze strukturą dynamiczną i etapem negocjacji. Porównanie jakości klasyfikacji oraz wykorzystanie zespołu metod fuzji w systemie rozproszonym w celu uzyskania jednoznacznych wyników (wyniki badań w tym zakresie opisano w rozdziale 4.4.7 oraz w artykułach [A6–A8, B3]),
- adaptacja modelu konfliktu Profesora Zdzisława Pawłaka do zagadnienia klasyfikacji z wykorzystaniem zespołu klasyfikatorów (wyniki badań w tym zakresie opisano w rozdziale 4.4.8 oraz w artykule [B4]).

Wprowadzenie teoretyczne, podstawowe definicje, prezentacja proponowanych nowych metod wraz z motywacjami oraz omówienie uzyskanych wyników zostało zaprezentowane poniżej.

4.2 Wprowadzenie

W mojej pracy zajmuję się systemami decyzyjnymi, w których wykorzystywana wiedza zgromadzona jest w postaci zbioru lokalnych baz wiedzy, dokładniej w postaci zbioru lokalnych tablic decyzyjnych. W dzisiejszych czasach, w związku z dużym nagromadzeniem wiedzy, problemem staje się wypracowanie decyzji w przypadku takich złożonych systemów wspomagania decyzji. Wiedza gromadzona jest w postaci rozproszonej z różnych powodów. Po pierwsze odrębne jednostki mogą gromadzić wiedzę dotyczącą tej samej dziedziny. Przykładem mogą tutaj być szpitale lub inne centra medyczne, w których tworzone są odrębne bazy wiedzy dotyczące tej samej tematyki. Bazy wiedzy na temat chorób nowotworowych tworzone są w różnych szpitalach na podstawie zdiagnozowanych pacjentów. Po drugie ogromne wolumeny danych, które obecnie są bardzo często spotykane, mogą zostać podzielone na mniejsze części. Następnie takie mniejsze zbiory danych mogą być użyte jednocześnie w procesie wnioskowania. Po trzecie wraz z upływającym czasem możemy mieć dostęp do nowych informacji. Wiedza, która jest aktualnie dostępna może być gromadzona w odrębnej bazie. Wykorzystanie wszystkich takich zbiorów zgromadzonych w przeciągu pewnego czasu powinno zaowocować wzrostem dokładności wnioskowania. Po czwarte wiedza zapisana w postaci rozproszonej może być wynikiem interakcji ze złożonymi zjawiskami (systemami). Fragmenty stanu takiego złożonego systemu są postrzegane w różnych momentach czasu, co dostarcza różnych tablic danych. Na podstawie tablic lokalnych tworzone są lokalne modele, które często okazują się sprzeczne. Prowadzi to do konieczności opracowania metod analiz konfliktów między tymi modelami i ich rozstrzygnięcia. Tablice te są również agregowane i stanowią podstawę do tworzenia aproksymacyjnych bardziej złożonych pojęć/decyzji.

Przegląd literatury oraz wskazanie fundamentalnych różnic dotyczących postaci wykorzystywanej wiedzy i obszaru zastosowań proponowanego przeze mnie systemu, w porównaniu z metodami znanymi z literatury, zostanie przedstawione poniżej.

Problem jednoczesnego wykorzystania wiedzy dostępnej w odrębnych zbiorach danych jest poruszany w kontekście różnych zagadnień informatyki takich jak: multiple classi-

fier systems [14, 24, 36], distributed decision-making [11, 45, 46], group decision-making [2, 16] oraz data science [22, 38]. Moja praca naukowa nie jest jednak bezpośrednio związana z żadnym z tych zagadnień. Oczywiście problematyka dotycząca jednoczesnego zastosowania wiedzy z różnych zbiorów danych jest wspólnym mianownikiem mojej pracy i wymienionych wcześniej podejść, jednak podejścia te różnią się zastosowaniami i założeniami.

Po pierwsze w proponowanym przeze mnie podejściu celem jest wykorzystanie wiedzy, która jest zadana z góry i dostępna w formie rozproszonej – proces rozpraszania wiedzy nie jest jednym z etapów budowania modelu. Zbiór dostępnych baz lokalnych jest gromadzony już w postaci rozproszonej przez różne jednostki, dlatego też wygenerowanie na ich podstawie jednego wspólnego zbioru danych nie jest procesem prostym ani oczywistym. Natomiast w większości podejść znanych z literatury proces podziału jednego zbioru danych na podzbiory jest jednym z etapów budowy modelu. System taki konstruowany jest z klasyfikatorów bazowych w celu poprawy efektywności wnioskowania kosztem zwiększonej złożoności obliczeniowej.

Po drugie z uwagi na wspomniane wyżej zastosowanie w proponowanym przeze mnie podejściu nie zakłada się, że w tablicach lokalnych występują te same obiekty ani, że są one opisywane za pomocą tych samych atrybutów. Zbiory atrybutów oraz obiektów różnych tablic lokalnych nie muszą być ani równe ani rozłączne. Ponadto zakłada się, że w tablicach lokalnych nie występuje ujednolicony sposób identyfikacji obiektów, taki który pozwoliłby na rozpoznawanie czy te same obiekty występują w różnych tablicach lokalnych. Takie założenia musiały zostać przyjęte z uwagi na to, że model ma być wykorzystywany w rzeczywistych zastosowaniach, w których tablice lokalne są gromadzone przez różne jednostki w postaci rozproszonej. W podejściach znanych z literatury zazwyczaj zakłada się, że podział jednej tablicy na zbiór tablic realizowany jest w sposób horyzontalny lub wertykalny [4, 19]. W pierwszym przypadku we wszystkich lokalnych tablicach decyzyjnych występują te same atrybuty warunkowe natomiast różny jest zbiór obiektów. W drugim przypadku we wszystkich lokalnych tablicach decyzyjnych występują te same obiekty ale różny jest zbiór atrybutów. W [27], rozważany jest zespół klasyfikatorów, w którym klasyfikatory bazowe konstruowane są na podstawie tablic lokalnych z różnymi zbiorami obiektów oraz różnymi zbiorami atrybutów. Proponowana metoda jest kombinacją dwóch znanych z literatury metod: Bagging [1] i Random Subspace Method [17]. Jednakże wiedza, która jest dostarczana do systemu jest przechowywana w pojedynczej tablicy – nie ma możliwości wykorzystania zbioru lokalnych tablic decyzyjnych, które są gromadzone niezależnie. W [37], został zaproponowany system z własnością uczenia przyrostowego. W modelu tym przewidziano możliwość nauki nowej wiedzy, gdy nowy zbiór danych jest dostępny. Wykorzystywane zbiory danych, na podstawie których konstruowane są klasyfikatory, mają różne zbiory atrybutów. Jednak w systemie tym nie są analizowane relacje zachodzące pomiędzy dostępnymi zbiorami danych. Nie jest analizowana również spójność wiedzy zawartej w tych zbiorach danych. W podejściu proponowanym w pracy [20] wykorzystywana jest wiedza rozproszona o jeszcze innej postaci. Każdy agent działa w lokalnym świecie, który jest opisany za pomocą c -granuli. Agent kontroluje obliczenia wykonywane przez c -granule w celu osiągnięcia swoich celów. Z uwagi na ograniczoną percepcję agenta tylko częściowa informacja o interakcjach ze światem fizycznym jest dostępna. Koncepcja złożonych granuli poruszana była już w pracy Zadeha [57].

W mojej pracy stosuję termin – wiedza rozproszona (ang. *dispersed knowledge*), podczas gdy w wielu pracach [50, 52] można spotkać termin – wiedza podzielona (ang. *distributed knowledge*). Powodem używania tej terminologii są przyjęte założenia, które zostały wymienione powyżej. Pozwolę sobie przytoczyć je ponownie. Po pierwsze, zarówno obiekty jak i atrybuty mogą być współdzielone pomiędzy lokalnymi tablicami decyzyjnymi. Po drugie, nie ma możliwości sprawdzenia czy dany obiekt występuje w kilku tablicach decyzyjnych, ponieważ w tablicach nie istnieją wspólne i uniwersalne identyfikatory obiektów. Po trzecie, proces podziału wiedzy nie jest częścią pracy systemu, wiedza jest gromadzona już w postaci rozproszonej. Wiemy natomiast, że wiedza zawarta w tablicach lokalnych dotyczy tej samej dziedziny, co jest odzwierciedlone poprzez występowanie wspólnych atrybutów decyzyjnych we wszystkich rozważanych tablicach lokalnych.

Kolejną ważną różnicą pomiędzy podejściami znanymi z literatury, a proponowanym przeze mnie systemem jest jego struktura. W rozważanym systemie analizowane są relacje, które występują między klasyfikatorami bazowymi, podczas podejmowania decyzji dla danego obiektu. Wykorzystywana jest dynamiczna struktura – klasyfikatory są reorganizowane dynamicznie – dla każdego nowego przypadku inny układ klasyfikatorów jest tworzony. Takie podejście jest raczej unikalne i odróżnia system od podejść znanych z literatury. W artykule [55], rozważany jest wybór podzbioru klasyfikatorów bazowych. Jednak ta selekcja jest realizowana tylko raz, a stworzony w ten sposób model jest wykorzystywany dla wszystkich nowych przypadków. Zatem system ma statyczną strukturę. W pracy [26], rozważane są pośrednie tablice decyzyjne zawierające wyniki otrzymane dla klasyfikatorów bazowych. Następnie generowane są reguły decyzyjne, których celem jest opis sposobu łączenia wyników klasyfikatorów. Jednak reguły decyzyjne są generowane tylko raz, a następnie wykorzystywane dla wszystkich nowych przypadków. Zatem system ma również statyczną strukturę. W pracach [53, 51], wykorzystywane są pewne sieci neuronowe w celu łączenia wektorów wygenerowanych przez klasyfikatory bazowe. Jednak w obu przypadkach proces uczenia sieci jest realizowany tylko raz, a otrzymane wyniki są wykorzystywane dla wszystkich nowych przypadków. To również świadczy o statycznej strukturze systemu. Artykuł [54], prezentuje podejście, w którym proces uczenia klasyfikatorów jest realizowany przy przyjęciu pewnych ograniczeń. Część wspólna wyników predykcji klasyfikatorów bazowych powinna być pojedynczą klasą decyzyjną. Ten system jest również przykładem pewnej statycznej metodologii jako, że nie uwzględniono etapu oceny czy wyniki poszczególnych klasyfikatorów są podobne. W pracy [21], rozważane jest grupowanie atrybutów, w celu bardziej efektywnego budowania klasyfikatorów na podstawie podzbiorów cech. System ten ma również statyczną strukturę jako, że proces grupowania jest realizowany tylko raz. Te i wiele innych przykładów świadczą o unikatowości podejścia proponowanego przeze mnie.

Z powodów wymienionych powyżej zaproponowany przeze mnie sposób wykorzystania wiedzy rozproszonej jest nowym podejściem i, według mojej wiedzy, nie był wcześniej rozpatrywany w literaturze.

4.3 Podstawowe pojęcia

W rozdziale tym przedstawione zostaną podstawowe pojęcia wykorzystywane w prowadzonych przeze mnie pracach, jak również krótko omówione zostaną modele analizy kon-

fliktów, znane z literatury, które stanowiły dla mnie inspirację do zaproponowania nowych rozwiązań.

Moje zainteresowania naukowe związane z wykorzystaniem wiedzy rozproszonej stanowią kontynuację pracy zapoczątkowanej w rozprawie doktorskiej [39], w której zaproponowany został rozproszony system ze strukturą statyczną. W systemie tym decyzje lokalne generowane są na podstawie lokalnych tablic decyzyjnych, natomiast w procesie generowania decyzji globalnych bardzo istotnym zagadnieniem są powstające konflikty. Zatem jednym z podstawowych zagadnień rozważanych w mojej pracy jest analiza konfliktów. Model wykorzystywany przeze mnie jak i definicja agenta oparte są na koncepcji analizy konfliktów zaproponowanej przez Profesora Zdzisława Pawłaka.

W 1984 roku, w pracy [28], Profesor Zdzisław Pawlak zaproponował prosty i intuicyjny model analizy konfliktów. Zagadnienie to było następnie rozwijane w pracach [29, 30, 31, 32, 33, 34, 35]. W modelu tym możliwe jest zdefiniowanie relacji pomiędzy jednostkami biorącymi udział w konflikcie, jak również graficzne przedstawienie sytuacji konfliktowej oraz zidentyfikowanie koalicji. W modelu Pawłaka strony biorące udział w konflikcie nazywane są agentami i jest to system wieloagentowy, ale system ten nie ma nic wspólnego z systemami wieloagentowymi znanymi z literatury [13, 25, 44]. Każdy agent wyraża swoją opinię na temat pewnych analizowanych spraw poprzez przyporządkowanie sprawie jednej z trzech wartości: -1 oznacza, że agent jest przeciwny danej sprawie, 0 oznacza neutralność agenta, 1 oznacza, że agent opowiada się za daną sprawą. Wiedzę o sytuacji konfliktowej zapisujemy w postaci systemu informacyjnego $S = (U, A)$. Elementami zbioru uniwersum U są agenci, zbiór atrybutów A jest zbiorem spraw, zaś zbiór wartości atrybutu $a \in A$ jest równy $V^a = \{-1, 0, 1\}$. Wartość $a(x)$, gdzie $x \in U, a \in A$ jest opinią agenta x na temat sprawy a . Dla każdego atrybutu $a \in A$ definiujemy funkcję $\phi_a : U \times U \rightarrow \{-1, 0, 1\}$ określoną następująco:

$$\phi_a(x, y) = \begin{cases} 1 & \text{jeżeli } a(x)a(y) = 1 \text{ lub } x = y, \\ 0 & \text{jeżeli } a(x)a(y) = 0 \text{ oraz } x \neq y, \\ -1 & \text{jeżeli } a(x)a(y) = -1. \end{cases} \quad (1)$$

Następnie w zbiorze $U \times U$ definiujemy trzy relacje: R_a^+ zgody, R_a^0 neutralności, R_a^- konfliktu, które wyrażają związki między agentami:

$$\begin{aligned} R_a^+(x, y) & \text{ wtedy i tylko wtedy, gdy } \phi_a(x, y) = 1, \\ R_a^0(x, y) & \text{ wtedy i tylko wtedy, gdy } \phi_a(x, y) = 0, \\ R_a^-(x, y) & \text{ wtedy i tylko wtedy, gdy } \phi_a(x, y) = -1. \end{aligned}$$

Zatem relacja R_a^+ jest relacją równoważności. Każda klasa abstrakcji relacji R_a^+ jest koalicją agentów określoną przez atrybut a .

W celu wyznaczenia w sposób liczbowy różnicy poglądów dwóch agentów x, y na dany zestaw spraw $B \subseteq A$ definiujemy funkcję odległości $\rho_B^* : U \times U \rightarrow [0, 1]$

$$\rho_B^*(x, y) = \frac{\sum_{a \in B} \phi_a^*(x, y)}{\text{card}\{B\}}, \quad (2)$$

gdzie

$$\phi_a^*(x, y) = \frac{1 - \phi_a(x, y)}{2} = \begin{cases} 0 & \text{jeżeli } a(x)a(y) = 1 \text{ lub } x = y, \\ 0,5 & \text{jeżeli } a(x)a(y) = 0 \text{ oraz } x \neq y, \\ 1 & \text{jeżeli } a(x)a(y) = -1. \end{cases} \quad (3)$$

Funkcję odległości wyznaczoną dla zestawu wszystkich spraw $B = A$ zapisujemy w skrócie ρ^* . Mówimy, że agenci $x, y \in U$ są

- przyjaciółmi $R^+(x, y)$, jeśli $\rho^*(x, y) < 0,5$,
- w konflikcie $R^-(x, y)$, jeśli $\rho^*(x, y) > 0,5$,
- neutralni $R^0(x, y)$, jeśli $\rho^*(x, y) = 0,5$.

Koalicja to podzbiór zbioru agentów $X \subseteq U$ taki, że $\forall x, y \in X R^+(x, y)$. Tak zdefiniowane koalicje nie muszą być zbiorami parami rozłącznymi.

Natężenie konfliktu pomiędzy agentami może być również wyrażone w inny sposób. Definiujemy funkcję konfliktu $\rho_B : U \times U \rightarrow [0, 1]$ dla zestawu spraw $B \subseteq A$ w sposób następujący

$$\rho_B(x, y) = \frac{\text{card}\{\delta_B(x, y)\}}{\text{card}\{B\}}, \quad (4)$$

gdzie $\delta_B(x, y) = \{a \in B : a(x) \neq a(y)\}$.

W modelu Pawlaka wprowadzone zostało również pojęcie siły agentów oraz strategii. Siła agenta może reprezentować jego potęgę militarną lub ekonomiczną. Natomiast strategia określa w jaki sposób agent dzieli swoją siłę na wrogie działania wobec swoich przeciwników.

Okolo roku 1996 zaproponowany został model sytuacji konfliktowej przez Skowrona i Deję, który jest rozszerzeniem modelu Pawlaka. Koncepcja ta została opisana w pracach [6, 7, 8, 9, 47, 10]. Autorzy modelu przytaczają kilka powodów zaproponowania nowego podejścia. Po pierwsze założenie, że wszyscy agenci postrzegają sytuację konfliktową przez pryzmat jednakowych atrybutów powoduje, iż model można zastosować w przypadku niewielu sytuacji rzeczywistych. Ponadto zbiór wartości atrybutów w wielu sytuacjach rzeczywistych może okazać się zbyt ograniczony. Jednak głównym powodem jest to, że w modelu Pawlaka nie docieka się prawdziwych przyczyn konfliktu. Skowron i Deja zauważają, że opinie agentów na dane sprawy są decyzjami podjętymi na podstawie pewnych przesłanek, które są rzeczywistą przyczyną konfliktu.

Za fundamentalną przyczynę konfliktu uznano pośrednio ograniczenia wynikające ze skończoności różnego typu zasobów, natomiast za bezpośrednią przyczynę powstawania konfliktów przyjęto chęć agentów posiadania zasobów. W zaproponowanym modelu uwzględniono również fakt, że każdy agent opisuje sytuację konfliktową po swojemu. Odzwierciedleniem tego założenia w modelu matematycznym jest przyporządkowanie każdemu agentowi odrębnego systemu informacyjnego. Założono, że zbiory atrybutów różnych agentów są parami rozłączne.

Okolo roku 2006 zaproponowany został model sytuacji konfliktowej przez Skowrona, Ramanna i Peters'a, który również jest rozwinięciem modelu Pawlaka. Koncepcja ta została opisana w pracach [40, 41, 42, 43, 49]. W modelu tym, podobnie jak w modelu Pawlaka, poglądy agentów przedstawione są w formie systemu informacyjnego $S = (U, A)$. Autorzy wprowadzają dodatkowy zbiór warunków/atributów do modelu w celu reprezentowania różnych wymiarów złożonej sytuacji konfliktu. Dogłębna analiza skomplikowanej struktury konfliktów jest osiągnięta poprzez zastosowanie przestrzeni przybliżeń. W przestrzeni tej jako miara do określania stopnia konformizmu podczas negocjacji wykorzystywana jest

funkcja pokrycia. W celu rozwiązania konfliktu wykorzystywane jest również wnioskowanie Boolowskie.

W mojej rozprawie doktorskiej zaproponowana została koncepcja agenta, który, podobnie jak w modelu Pawlaka, jest stroną biorącą udział w konflikcie. Ma on dostęp do pewnego fragmentu wiedzy, na podstawie której potrafi podjąć decyzje (dokonać klasyfikacji danego obiektu).

Agent ag należący do zbioru Ag nazywany jest agentem zasobów jeśli ma on dostęp do zasobów zapisanych w tablicy decyzyjnej $D_{ag} := (U_{ag}, A_{ag}, d_{ag})$, gdzie U_{ag} jest zbiorem uniwersum; A_{ag} jest zbiorem atrybutów warunkowych oraz V_{ag}^a jest zbiorem wartości atrybutu a , który zawiera znaki specjalne $*$ oraz $?$; d_{ag} jest atrybutem decyzyjnym. Równanie $a(x) = *$ dla pewnego $x \in U_{ag}$ oznacza, że dla obiektu x , wartość na atrybucie a nie wpływa na wartość atrybutu decyzyjnego. Równanie $a(x) = ?$ oznacza, że dla obiektu x , wartość na atrybucie a jest nieznaną. Jedynym warunkiem jaki musi być spełniony przez tablice decyzyjne agentów zasobów jest występowanie tego samego atrybutu decyzyjnego we wszystkich tablicach decyzyjnych.

Już w pracy doktorskiej przyjęto koncepcję, w której agenci zasobów, podobni w pewnym zadany sensie, będą łączeni w grupę zwaną klastrem. W rozproszonym systemie ze strukturą statyczną przyjęto bardzo proste podejście, w którym agenci zasobów, wnioskujący na podstawie wspólnych atrybutów warunkowych będą uznawani za podobnych i łączeni w grupę. Uzyskane w ten sposób grupy są zbiorami rozłącznymi. Definicja klastra w systemie ze strukturą statyczną jest następująca.

Klastrem nazywamy najmniejszy, ze względu na relację inkluzji, podzbiór zbioru agentów zasobów $C \subseteq Ag$, spełniający następujące dwa warunki

$$\forall ag \in C \exists ag' \in C; ag \neq ag' \quad A_{ag} \cap A_{ag'} \neq \emptyset, \quad (5)$$

$$\forall ag \in Ag \setminus C \quad \forall ag' \in C \quad A_{ag} \cap A_{ag'} = \emptyset. \quad (6)$$

W kolejnych etapach mojej pracy definicja ta była zmieniana tak, aby stworzone klastry lepiej odzwierciedlały faktyczne relacje zachodzące między agentami.

Rozproszony system ma strukturę hierarchiczną. Agenci zasobów znajdują się na najniższym poziomie hierarchii. Dla każdego klastra definiujemy agenta nadrzędnego – agenta syntezy as .

Przez rozproszony system podejmowania decyzji o strukturze statycznej rozumiemy

$$WSD_{Ag} = \langle Ag, \{D_{ag} : ag \in Ag\}, As, \delta \rangle, \quad (7)$$

gdzie Ag jest skończonym zbiorem agentów zasobów; $\{D_{ag} : ag \in Ag\}$ jest zbiorem tablic decyzyjnych agentów zasobów; As jest skończonym zbiorem agentów syntezy; $\delta : As \rightarrow 2^{Ag}$ jest różnowartościową funkcją przyporządkowującą każdemu agentowi syntezy jeden klast. Zbiorem wartości funkcji δ jest zbiór wszystkich klastrów utworzonych zgodnie z przyjętą definicją.

Agent syntezy ma dostęp do zagregowanej tablicy decyzyjnej, która tworzona jest na podstawie tablic decyzyjnych agentów zasobów z klastra podrzędnego. Agregacja ta ma na celu wyeliminowanie niespójności występujących w tablicach lokalnych agentów zasobów z jednego klastra. W literaturze można znaleźć różne metody agregacji hierarchicznej.

W [15], zaprezentowany został przegląd metod agregacji różnych typów danych. W opracowaniu tym rozważane jest, między innymi, podejście wykorzystujące sieci neuronowe w celu hierarchicznej agregacji funkcji fuzji. W [48], rozważana jest sytuacja, w której agenci wykorzystują adaptacyjne osądy w celu oceny ryzyka, analizy kosztów lub zysków. W podejściu tym występuje pewna hierarchia oraz stosowana jest agregacja systemów informacyjnych z uwzględnieniem ograniczeń. Można zauważyć, że hierarchiczna agregacja ma szczególne znaczenie, gdy wykorzystywana jest wiedza rozproszona.

W mojej rozprawie doktorskiej została zaproponowana i zdefiniowana metoda agregacji tablic decyzyjnych. Tablice zagregowane tworzone są zgodnie z poniższą definicją.

Niech $\{ag_{i_1}, \dots, ag_{i_k}\}$ będzie klastrem systemu rozproszonego WSD_{Ag} , oraz niech as będzie agentem syntezy nadrzędnym w stosunku do tego klastra. Wiedza agenta syntezy as zapisana jest w postaci tablicy decyzyjnej

$$D_{as} = (U_{as}, A_{as}, \bar{d}_{as}), \quad (8)$$

gdzie $A_{as} = \{\bar{b} : U_{as} \rightarrow V^b, b \in \bigcup_{j=1}^k A_{ag_{i_j}}\}$ i \bar{d}_{as} jest funkcją $\bar{d}_{as} : U_{as} \rightarrow V^d, V^d = \bigcup_{j=1}^k V_{ag_{i_j}}^d$. Dodatkowo, dla każdego maksymalnego, pod względem zawierania, zbioru $G = \{ag_{i_{j_1}}, \dots, ag_{i_{j_k}}\} \subseteq \{ag_{i_1}, \dots, ag_{i_k}\}$ i zbioru obiektów $\{x_{i_{j_1}}, \dots, x_{i_{j_k}}\}, x_{i_m} \in U_{ag_{i_m}}, j_1 \leq m \leq j_k$, spełniającego następujące dwa warunki:

$$\forall b \in W_{sp_G} \exists v_b \in V^b \forall j_1 \leq m \leq j_k [b \in A_{ag_{i_m}} \implies b(x_{i_m}) = v_b], \quad (9)$$

$$d_{ag_{i_{j_1}}}(x_{i_{j_1}}) = \dots = d_{ag_{i_{j_k}}}(x_{i_{j_k}}), \quad (10)$$

gdzie $W_{sp_G} = \{b : b \in A_{ag_{i_1}} \cap A_{ag_{i_2}} \cap \dots \cap A_{ag_{i_k}}\}$, obiekt $x \in U_{as}$ jest zdefiniowany w sposób następujący:

$$\forall ag_j \in G \forall b \in A_{ag_j} [\bar{b}(x) = b(x_j) \text{ i } \bar{d}_{as}(x) = d_{ag_j}(x_j)], \quad (11)$$

$$\forall b \in \bigcup_{j=1}^k A_{ag_{i_j}} \setminus \bigcup_{m=j_1}^{j_k} A_{ag_{i_m}} \bar{b}(x) = ?. \quad (12)$$

Innymi słowy, obiekty tablicy zagregowanej powstają poprzez połączenie obiektów, z tablic agentów zasobów należących do jednego klastra, które mają takie same wartości na atrybucie decyzyjnym i na wspólnych atrybutach warunkowych. Metoda agregacji tablic decyzyjnych ma wykładniczą złożoność obliczeniową. Dlatego zaproponowano modyfikację polegającą na zastąpieniu, w powyższej definicji, zbiorów uniwersum U_{ag} , zbiorami obiektów relewantnych U_{ag}^{rel} . Zbiór U_{ag}^{rel} definiujemy poprzez wybranie z każdej klasy decyzyjnej $X_v, v \in V^d$, tablicy decyzyjnej D_{ag} najmniejszego zbioru liczącego co najmniej m_2 obiektów, których wartości na atrybutach warunkowych realizują największe podobieństwo do obiektu klasyfikowanego. Metoda ta nazwana została aproksymacyjną metodą agregacji tablic decyzyjnych. Oprócz oczywistej zalety, że ma ona mniejszą złożoność obliczeniową i można ją zastosować nawet dla dużych zbiorów danych, to jeszcze równoważy ona licznosc klas decyzyjnych. Czyli nawet dla zbiorów danych z dużą dysproporcją liczności poszczególnych klas decyzyjnych, tablice zagregowane agentów syntezy są konstruowane na podstawie obiektów relewantnych, które są równoliczne pomiędzy klasami decyzyjnymi. W mojej pracy stosuję miarę podobieństwa Gowera, której definicję można znaleźć

między innymi w [A2]. Głównymi zaletami tej miary jest możliwość zastosowania jej do atrybutów mieszanych oraz to, że uwzględnia ona występowanie wiedzy niepełnej.

Na podstawie tablic zagregowanych agentów syntezy generowane są decyzje, które zapisywane są w postaci wektorów wartości. Dla j -tego agenta syntezy generowany jest wektor wartości $\mu_j(x) = [\mu_{j,1}(x), \dots, \mu_{j,c}(x)]$ o wymiarze c równym liczbie klas decyzyjnych. W pracy doktorskiej współrzędna tego wektora $\mu_{j,i}(x)$ definiowana była jako maksymalna wartość miary podobieństwa obiektów z klasy decyzyjnej X_{v_i} , tablicy decyzyjnej agenta syntezy as_j , do obiektu klasyfikowanego x . Takie podejście jest też stosowane w pracach [A1–A5] oraz [B1, B2, B4]. Natomiast w pracach [A6–A8] oraz [B3] współrzędna $\mu_{j,i}(x)$ definiowana jest jako średnia wartość miary podobieństwa obiektu klasyfikowanego do obiektów relewantnych wybranych z klasy decyzyjnej X_{v_i} , tablicy decyzyjnej agenta syntezy as_j . W celu wyboru obiektów relewantnych z tablic zagregowanych wykorzystywany jest parametr m_3 . Wartość tego parametru określa wielkość zbioru obiektów relewantnych.

Następnie realizowana jest kolejna agregacja, która tym razem dotyczy wektorów wartości wygenerowanych przez agentów syntezy. Ostatecznie podejmowana jest decyzja globalna. W mojej pracy doktorskiej, w tym celu, zaproponowana została metoda algorytmu gęstościowego, która jest wykorzystywana w pracach [A1–A5] oraz [B1, B2, B4].

Najpierw generowany jest jeden wektor wartości odzwierciedlający decyzje podjęte przez wszystkich agentów syntezy. Jest to realizowane przez przekształcenie liniowe na wektorach agentów syntezy. Mając jeden wektor wartości, najprostszym podejściem dla zdefiniowania ostatecznej decyzji byłoby wybranie decyzji, dla której maksymalna wartość współczynnika wektora została przyporządkowana. Jednakże, aby wybrać zbiór decyzji, który zawiera nie tylko decyzje z maksymalnym wsparciem, ale również te decyzje, dla których wsparcie jest relatywnie duże, zastosowano algorytm gęstościowy. Wówczas nie mamy problemu ze zdefiniowaniem pojęcia „relatywnie duże wsparcie”, które może oznaczać coś innego dla różnych zbiorów danych. Zatem każdy współczynnik wektora traktowany jest jako oddzielny punkt. Następnie wykorzystując algorytm DBSCAN [12] wybierany jest zbiór decyzji, dla których odpowiadające współczynniki wektora są gęsto ułożone wokół współczynnika z maksymalną wartością.

W następnych rozdziałach opisane zostaną rozważane problemy, zaproponowane nowe rozwiązania oraz wybrane otrzymane wyniki.

Należy jeszcze zauważyć, że w mojej pracy wykorzystywałam sztucznie rozprasane zbiory danych, ponieważ nie miałam dostępu do rzeczywistych, rozproszonych zbiorów danych. Niestety w wielu dziedzinach zastosowań, dostęp do zbiorów danych jest ograniczony. To ograniczenie jest skutkiem wielu czynników, włączając prawo ochrony danych, które zabrania klinikom i naukowcom swobodnego dzielenia się danymi pacjentów lub poufnymi informacjami biznesowymi. Dostępność zbiorów danych w postaci rozproszonej (w postaci zbiorów tablic decyzyjnych) jest jeszcze gorsza. Problem dostępności danych, z którym się zetknęłam, jest powszechny. W literaturze nie znalazłam przypadku, w którym wykorzystywane byłyby rzeczywiste dane rozproszone. Podobnie jak ja, inni autorzy prac [18], sztucznie rozpraszają dane dostępne w repozytoriach.

W mojej pracy eksperymenty były wykonywane na rozproszonych zbiorach danych stworzonych na podstawie kilku zbiorów wybranych z repozytorium UCI. W większości prac były to zbiory: Soybean, Vehicle Silhouettes, Dermatology i Landsat Satellite. Zostały one wybrane tak, aby miały różne typy atrybutów (ilościowe, jakościowe i binarne), zna-

cząco różne liczności zbiorów obiektów oraz różną liczbę klas decyzyjnych. Bardzo ważnym było też, aby zbiory te miały dużą liczbę atrybutów warunkowych (tak, aby sensowne było rozpraszanie takich danych). W ten sposób, proponowane podejścia mogły być testowane pod różnym kątem. Dla każdego zbioru danych rozważano pięć wersji rozproszeń – podział na różną liczbę tablic decyzyjnych – trzy, pięć, siedem, dziewięć i jedenaście tablic decyzyjnych. Uznano, że trzy lokalne tablice decyzyjne to najmniejsza liczba, dla której jest sens stosować system rozproszony – rozpatrywać relacje między tablicami i tworzyć klastry. Największą rozważaną liczbą tablic lokalnych było jedenaście, ponieważ przy dostępnych zbiorach atrybutów warunkowych, jeszcze drobniejszy podział spowodowałby, że lokalne tablice miałyby za mało atrybutów warunkowych lub większość atrybutów w tablicach byłaby wspólna. Rozważano zatem różne liczby tablic decyzyjnych z zakresu od trzech do jedenastu, przy zastosowaniu skoku równego dwa. Zauważmy jeszcze, że podział zbioru danych na lokalne tablice nie miał na celu poprawy jakości wnioskowania systemu. Proces ten był przeprowadzany tylko raz, w sposób losowy (szczegóły można znaleźć w pracach [A1, A2, A4]), i nie był optymalizowany w żaden sposób.

Eksperymenty realizowane były głównie z wykorzystaniem zbioru treningowego i testowego (ang. *train-and-test*) (dla zbiorów danych Soybean oraz Landsat Satellite zbiory te są bezpośrednio dostępne w repozytorium UCI). Jest to odpowiednia metoda dla testowania danych rozproszonych, ponieważ metoda krosvalidacji (ang. *cross-validation*) jest niemożliwa do zastosowania bez utraty ogólności. Jak zostało wspomniane wcześniej, w systemie rozproszonym zbiory atrybutów warunkowych i zbiory obiektów różnych lokalnych tablic decyzyjnych nie muszą być ani równe, ani rozłączne. Dodatkowo, identyfikatory obiektów nie są uniwersalne dla wszystkich tablic decyzyjnych. Oznacza to, że nie ma możliwości sprawdzenia czy te same obiekty występują w różnych lokalnych tablicach decyzyjnych. Zatem wykorzystanie k-krotnej krosvalidacji jest niemożliwe, przy zachowaniu ogólności. Bardziej szczegółowo, wynika to z założenia, że wiedza przechowywana jest w formie kilku tablic lokalnych, które są zadane z góry. Podczas, gdy zbiór testowy jest przechowywany w jednej tablicy decyzyjnej, w której zbiór atrybutów warunkowych jest równy sumie zbiorów atrybutów warunkowych lokalnych tablic decyzyjnych. Gdybyśmy chcieli wykorzystać metodę krosvalidacji, to z każdej tablicy lokalnej musiałaby zostać wybrana próbka testowa. Przy utrzymaniu ogólności, nie ma możliwości zweryfikowania czy te same obiekty zostały wylosowane z różnych tablic lokalnych. Dlatego, wygenerowanie jednej tablicy decyzyjnej ze zbiorów wybranych próbek testowych nie jest możliwe. Jednakże, z uwagi na kompletność eksperymentów, w pracach [A2, A3, A6] zrezygnowano z ogólności założeń i zrealizowano testy metodą 10-krotnej krosvalidacji. Uzyskane wyniki potwierdziły hipotezy, które zostały sformułowane na podstawie wyników uzyskanych metodą *train-and-test*.

System rozproszony generuje zbiór decyzji, a nie jedną decyzję. Dlatego dla oceny jakości klasyfikacji wykorzystano następujące miary

- estymator błędu klasyfikacji na zbiorze testowym niezależnym od zbioru treningowego

$$e = \frac{1}{\text{card}\{U_{test}\}} \sum_{x \in U_{test}} I(d(x) \notin \hat{d}_{WSD_{Ag}}(x)), \quad (13)$$

gdzie $I(d(x) \notin \hat{d}_{WSD_{Ag}}(x)) = 1$, gdy $d(x_i) \notin \hat{d}_{WSD_{Ag}}(x)$ i $I(d(x) \notin \hat{d}_{WSD_{Ag}}(x)) = 0$,

gdy $d(x) \in \hat{d}_{WSD_{Ag}}(x)$; zbiór testowy jest zapisany w postaci tablicy decyzyjnej $D_{test} = (U_{test}, \bigcup_{ag \in Ag} A_{ag}, d)$; $\hat{d}_{WSD_{Ag}}(x)$ jest zbiorem decyzji wygenerowanym przez system rozproszony dla obiektu testowego x ,

- estymator błędu jednoznaczności klasyfikacji na zbiorze testowym niezależnym od zbioru treningowego

$$e_{ONE} = \frac{1}{card\{U_{test}\}} \sum_{x \in U_{test}} I(d(x) \neq \hat{d}_{WSD_{Ag}}(x)), \quad (14)$$

gdzie $I(d(x) \neq \hat{d}_{WSD_{Ag}}(x)) = 1$, gdy $\{d(x)\} \neq \hat{d}_{WSD_{Ag}}(x)$ i $I(d(x) \neq \hat{d}_{WSD_{Ag}}(x)) = 0$, gdy $\{d(x)\} = \hat{d}_{WSD_{Ag}}(x)$;

- średnia liczebność zbiorów decyzji wygenerowanych dla obiektów testowych

$$\bar{d}_{WSD_{Ag}} = \frac{1}{card\{U_{test}\}} \sum_{x \in U_{test}} card\{\hat{d}_{WSD_{Ag}}(x)\}. \quad (15)$$

4.4 Rozważane problemy, zaproponowane nowe rozwiązania oraz wybrane otrzymane wyniki

W pierwszym etapie prac zastosowałam pewne przekształcenia na tablicach decyzyjnych agentów zasobów nie zmieniając głównych założeń systemu z rozprawy doktorskiej. Natomiast w drugim etapie prac zmodyfikowałam całkowicie strukturę systemu wykorzystującego wiedzę rozproszoną. W dalszym ciągu zostały podtrzymane założenia dotyczące gromadzenia wiedzy w kilku tablicach decyzyjnych oraz o strukturze hierarchicznej systemu. Jednak głównym moim celem było zastosowanie metod analizy konfliktów w systemie z wiedzą rozproszoną. Przy przyjętych założeniach, które zostały wymienione powyżej, niemożliwym było zastosowanie żadnego z modeli znanych z literatury.

Zaproponowałam różne podejścia do tworzenia systemu wykorzystującego wiedzę rozproszoną. Moje główne osiągnięcia w tej dziedzinie polegają na zaadoptowaniu metod analizy konfliktów do budowy struktury systemu rozproszonego. Zaproponowałam modyfikacje relacji między agentami, funkcji odległości agentów i koalicji. Wykorzystałam elementy negocjacji w celu stworzenia klastrów, które lepiej odzwierciedlają poglądy agentów na sytuację konfliktową. Rozpatrywałam możliwość wykorzystania wag agentów, uwzględniających różnorodne aspekty, podczas generowania wspólnej decyzji na podstawie wiedzy rozproszonej. Wykorzystałam piętnaście różnych metod fuzji predykcji klasyfikatorów bazowych, znanych z literatury, do systemu z wiedzą rozproszoną. Rozważałam również podejście, w którym wiele różnych metod fuzji jest wykorzystywanych w jednym procesie podejmowania decyzji. Porównałam otrzymane wyniki z wynikami uzyskanymi bez stosowania systemu z wiedzą rozproszoną oraz z wynikami generowanymi przez inne metody.

Wszystkie wspomniane wyżej metody, wraz z implementacjami oraz ich zastosowania, określają mój wkład do habilitacji przedstawionej do oceny.

4.4.1 Poprawa jakości klasyfikacji oraz efektywności rozproszonego systemu ze statyczną strukturą poprzez zastosowanie redukcji zbiorów atrybutów warunkowych lokalnych tablic decyzyjnych [A1]

W pracy [A1] zaproponowałam nowe podejście, które polega na zastosowaniu metody redukcji wielowymiarowości w rozproszonym systemie ze strukturą statyczną, który został zdefiniowany w rozprawie doktorskiej [39]. Wprowadzona przeze mnie modyfikacja polega na redukcji zbiorów atrybutów warunkowych tablic decyzyjnych agentów zasobów. Redukcja wielowymiarowości jest bardzo istotnym i szeroko omawianym w literaturze [3, 5] zagadnieniem. Dwa główne podejścia to ekstrakcja cech i selekcja cech. W mojej pracy zastosowałam podejście polegające na selekcji atrybutów, które dostarczają najwięcej informacji, oraz usuwaniu atrybutów nieistotnych i nadwymiarowych. Oczywiście jest, że występowanie atrybutów redundantnych w zbiorze danych zwiększa czas działania algorytmów ale, w wielu przypadkach, wpływa również na obniżenie jakości klasyfikacji. Te nadwymiarowe atrybuty, które są usuwane, często powodują zdeformowanie i zaburzenie wyników klasyfikacji, szczególnie gdy miara podobieństwa lub odległości jest wykorzystywana w procesie klasyfikacji (jak to ma miejsce w przypadku systemu rozproszonego).

Zastosowana przeze mnie metoda redukcji atrybutów wykorzystuje zagadnienia teorii zbiorów przybliżonych i realizowana była w następujący sposób. Dla każdej lokalnej tablicy decyzyjnej wyznaczane były reduktu zbioru atrybutów warunkowych. W tym celu wykorzystany został program RSES [58]. Spośród wszystkich wygenerowanych reduktów wybierany był redukt z najmniejszą liczbą atrybutów. Następnie tworzona była podtablica zawierająca zbiór atrybutów warunkowych ograniczony do tego reduktu. Rozproszony system wykorzystywał te zmodyfikowane tablice lokalne, w celu wygenerowania ostatecznej decyzji.

W pracy [A1] rozważałam różne metody eliminacji niespójności wiedzy i metody analizy konfliktów. Wykorzystywałam metodę agregacji tablic decyzyjnych oraz jej aproksymacyjną wersję i metodę algorytmu gęstościowego, które zostały omówione w rozdziale 4.3. Ale również metodę boolowską analizy niespójności, metodę agregacji tablic decyzyjnych poprzedzoną losową edycją i kondensacją Harta oraz metodę głosowania ważonego, które zostały zaproponowane w mojej pracy doktorskiej. W metodzie boolowskiej analizy niespójności sprawdzane jest czy dla każdego dwóch agentów zasobów z jednego klastra oraz każdej klasy decyzyjnej istnieją obiekty relewantne z danej klasy decyzyjnej, które mają zgodne wartości na wspólnych atrybutach warunkowych. W metodzie agregacji tablic decyzyjnych poprzedzonej losową edycją i kondensacją Harta zaproponowano zastosowanie metod znanych z literatury: losowej edycji oraz metody Harta [23], w celu wyboru z każdej tablicy decyzyjnej agenta zasobów niewielkiego podzbioru obiektów reprezentujących wiedzę zgromadzoną w tablicy decyzyjnej. Następnie realizowana jest agregacja tablic decyzyjnych z wykorzystaniem wybranych podzbiorów reprezentatywnych. W pracy [A1] porównałam wyniki uzyskane przy zastosowaniu redukcji zbiorów atrybutów warunkowych z wynikami uzyskanymi na podstawie niezredukowanych danych. Mały fragment tych wyników, ograniczony tylko do podania błędu klasyfikacji dla systemu, w którym zastosowano jedynie aproksymacyjną metodę agregacji tablic decyzyjnych oraz metodę algorytmu gęstościowego, został przedstawiony w tabeli 1. W tabeli nie podano wyników dla wersji rozproszenia na siedem, dziewięć i jedenaście tablic decyzyjnych zbioru danych Landsat Satellite, gdyż zastosowana metoda redukcji zbiorów atrybutów warunkowych

Tabela 1: Błędy klasyfikacji – redukcja

Tablice lokalne	Zbiór danych Soybean		Zbiór danych Dermatology		Zbiór danych Landsat Satellite	
	Redukcja	Bez redukcji	Redukcja	Bez redukcji	Redukcja	Bez redukcji
	3	0,024	0,027	0,009	0,009	0,021
5	0,037	0,035	0,036	0,055	0,008	0,012
7	0,008	0,008	0,027	0,027		
9	0,058	0,058	0,018	0,018		
11	0,029	0,029	0,009	0,009		

nie przyniosła żadnych zmian dla tych danych. Jak można zauważyć zastosowanie selekcji atrybutów w niektórych przypadkach wpłynęło na poprawę jakości klasyfikacji. Na podstawie wszystkich wyników, przedstawionych w pracy [A1], sformułowałam pewne wnioski ogólne. Zastosowanie reduktu w większości przypadków wpłynęło na poprawę jakości wnioskowania; dla niektórych analizowanych zbiorów danych nie odnotowano większych zmian. Stwierdziłam, że w przypadku większych zbiorów danych zastosowanie reduktu daje większą poprawę efektywności wnioskowania (pod względem czasu działania algorytmu).

4.4.2 Adaptacja rozproszonego systemu ze strukturą statyczną w celu umożliwienia klasyfikacji na podstawie reguł decyzyjnych [B1]

W publikacji [B1] przedstawiłam nowe podejście, polegające na wykorzystaniu rodziny zbiorów reguł decyzyjnych w rozproszonym systemie ze strukturą statyczną, który został zaproponowany w rozprawie doktorskiej [39]. Przyjęłam założenie, że wiedza dostępna jest w postaci rodziny zbiorów reguł decyzyjnych (każdy z tych zbiorów reguł może być wygenerowany przez odrębne źródło wiedzy, w którym wykorzystywany jest dowolny zbiór atrybutów warunkowych). Zatem każdy z agentów zasobów ma teraz dostęp do zbioru reguł decyzyjnych.

Wprawdzie w pracy [B1] założyłam, że każdy agent zasobów ma pierwotnie dostęp do lokalnej tablicy decyzyjnej, na podstawie której generowane były reguły decyzyjne. W tym celu wykorzystywał program RSES [58]. Dla każdej tablicy decyzyjnej najpierw generowany był zbiór reduktów względem atrybutu decyzyjnego i obiektu, a następnie generowany był zbiór reguł metodą lokalną na podstawie wygenerowanych wcześniej reduktów. Przyjęłam takie podejście w celu wykorzystania danych rozproszonych, które zostały już zastosowane w poprzednich pracach. Dzięki temu możliwe było porównanie otrzymanych wyników z wynikami z poprzednich prac. Jednak definicje i metody zaproponowane przeze mnie w pracy [B1] można zastosować do dowolnej rodziny zbiorów reguł decyzyjnych.

W celu wykorzystania reguł lokalnych w systemie rozproszonym zaproponowałam metodę agregacji reguł. W wyniku jej zastosowania każdy z agentów syntezy ma dostęp do zbioru zagregowanych reguł decyzyjnych, które są konstruowane na podstawie reguł decyzyjnych agentów zasobów z klastra podrzędnego. Aby opisać metodę tworzenia reguł zagregowanych, najpierw należy podać pewne wstępne definicje.

Mówimy, że obiekt x spełnia formułę $\alpha = (a_1, v_1) \wedge \dots \wedge (a_k, v_k)$, gdzie $\{a_1, \dots, a_k\}$ jest podzbiorem zbioru atrybutów warunkowych oraz $v_i \in V^{a_i}$, $i \in \{1, \dots, k\}$ wtedy i tylko wtedy, gdy $a_i(x) = v_i$.

Regułą decyzyjną nazywamy wrazenie postaci $\alpha \rightarrow (d, v)$, gdzie α jest formułą, d jest

Tabela 2: Błędy klasyfikacji – reguły decyzyjne

Tablice lokalne	Zbiór danych Soybean		Zbiór danych Dermatology		Zbiór danych Landsat Satellite	
	Reguły	Tablice	Reguły	Tablice	Reguły	Tablice
	3	0,093	0,027	0,045	0,009	
5	0,048	0,035	0,018	0,055	0,059	0,012
7	0,077	0,008	0,009	0,027	0,035	0,012
9	0,072	0,058	0,018	0,018	0,036	0,020
11	0,027	0,029	0,009	0,009	0,031	0,010

atrybutem decyzyjnym, a v jest wartością tego atrybutu. Formułę α nazywamy częścią warunkową reguły decyzyjnej, a wyrażenie (d, v) nazywamy konkluzją reguły decyzyjnej.

Mówimy, że reguły decyzyjne $\alpha \rightarrow (d, v_1)$ i $\beta \rightarrow (d, v_2)$ są niesprzeczne wtedy i tylko wtedy, gdy $v_1 = v_2$ oraz jeżeli w formułach α i β istnieją formuły atomowe dla tego samego atrybutu warunkowego, wówczas wartości atrybutów w tych formułach atomowych są równe. Innymi słowy

$$(\alpha \equiv (a, v_1^a) \wedge \alpha' \text{ oraz } \beta \equiv (a, v_2^a) \wedge \beta') \implies (v_1^a = v_2^a), \quad (16)$$

gdzie α', β' są formułami.

W metodzie agregacji reguł dla każdego maksymalnego, ze względu na relację zawierania, zbioru niesprzecznych reguł decyzyjnych różnych agentów zasobów z jednego klastra, definiowana jest nowa reguła decyzyjna. Przez zbiór reguł decyzyjnych różnych agentów zasobów rozumiemy zbiór, który nie zawiera dwóch różnych reguł decyzyjnych jednego agenta zasobów. Reguła zagregowana jest definiowana w następujący sposób: część warunkowa reguły jest koniunkcją części warunkowych rozważanych reguł decyzyjnych agentów zasobów, natomiast konkluzja jest taka sama jak konkluzja rozważanych reguł decyzyjnych agentów zasobów (rozważane reguły mają tą samą konkluzję, ponieważ są niesprzeczne).

Następnie dla każdego agenta syntezy, ze zbioru jego reguł zagregowanych, wybierane są reguły, których część warunkowa jest spełniona przez obiekt klasyfikowany. Zazwyczaj dla obiektu klasyfikowanego istnieje wiele takich reguł i często mają one różne konkluzje. Dlatego dla każdego j -tego agenta syntezy definiowany jest wektor $[\mu_{j,1}(x), \dots, \mu_{j,c}(x)]$ o wymiarze równym liczbie klas decyzyjnych. Współrzędna tego wektora $\mu_{j,i}(x)$ jest równa liczbie reguł decyzyjnych wybranych ze zbioru reguł j -tego agenta syntezy, których w konkluzji występuje wartość v_i atrybutu decyzyjnego. Ostateczne decyzje podejmowane są z wykorzystaniem tych wektorów oraz metody algorytmu gęstościowego lub metody głosowania ważonego.

W pracy [B1], porównałam wyniki otrzymane przy zastosowaniu reguł decyzyjnych z wynikami uzyskanymi przy zastosowaniu, w systemie ze strukturą statyczną, czterech różnych metod eliminacji niespójności wiedzy (wymienionych w poprzednim podrozdziale). Mały fragment tych wyników, ograniczony tylko do podania błędu klasyfikacji dla systemu, w którym zastosowano jedynie aproksymacyjną metodę agregacji tablic decyzyjnych oraz metodę algorytmu gęstościowego, został przedstawiony w tabeli 2. Nie podano wyników dla rozproszonego, na trzy tablice decyzyjne, zbioru danych Landsat Satellite, gdyż program RSES nie wygenerował reguł decyzyjnych dla tych danych w akceptowalnym czasie. Jak można zauważyć zastosowanie reguł decyzyjnych wygenerowanych na podstawie

tablic agentów zasobów w większości przypadków nie poprawia efektywności wnioskowania. Jednak jak wiadomo wykorzystanie reguł decyzyjnych wpływa na znaczącą poprawę czytelności systemu i zrozumienia na jakiej podstawie podejmowane są decyzje. Ponadto podejście takie daje szerszy zakres zastosowań systemu. System można wykorzystać, gdy jest dostępna rodzina reguł decyzyjnych zamiast zbioru tablic decyzyjnych. Dlatego też to podejście jest istotnym i użytecznym rozwiązaniem.

4.4.3 Rozproszony system ze strukturą dynamiczną i rozłącznymi klastrami – poprawa identyfikacji relacji zachodzących pomiędzy klasyfikatorami bazowymi oraz jakości klasyfikacji [A2]

W pracy [A2] zaproponowałam nową metodę, która polega na całkowitej zmianie podejścia do tworzenia struktury systemu rozproszonego. Poprzednio system miał strukturę statyczną – tworzoną raz i wykorzystywaną dla wszystkich obiektów klasyfikowanych, zależną od zbiorów atrybutów warunkowych lokalnych tablic decyzyjnych. Jednakże występowanie wspólnych atrybutów w tablicach lokalnych nie zawsze związane jest z tym, że decyzje podejmowane na podstawie tych tablic są zgodne dla wszystkich rozpoznawanych obiektów. Być może nie jest to wystarczające uzasadnienie na to, aby tablice znalazły się w jednym klastrze.

Zaproponowałam model dynamiczny systemu, który polega na tym, że dla każdego obiektu klasyfikowanego definiowana jest nowa struktura systemu. Struktura ta jest zależna od klasyfikacji, która jest dokonywana na podstawie tablic lokalnych. W podejściu tym dąży się do wyznaczenia grup agentów zasobów, którzy są zgodni co do klasyfikacji danego obiektu testowego do klas decyzyjnych. Wykorzystałam model analizy konfliktów w celu zdefiniowania klastrów. Jak wspomniano we wprowadzeniu, podejście takie jest unikatowe i odróżnia system od podejść znanych z literatury.

Zaproponowałam definicje relacji przyjaźni oraz konfliktu, jak również sposób określania natężenia konfliktu, dla których inspiracją był model analizy konfliktów Profesora Pawlaka. Relacje między agentami są zdefiniowane poprzez ich pogląd na klasyfikację obiektu testowego do klas decyzyjnych. W pierwszym kroku, dla każdego agenta zasobów $ag_i \in Ag$, generowany jest wektor prawdopodobieństw $\bar{\mu}_i(x) = [\bar{\mu}_{i,1}(x), \dots, \bar{\mu}_{i,c}(x)]$ odzwierciedlający klasyfikację obiektu testowego x na podstawie tablicy decyzyjnej. Wymiar wektora jest równy liczbie klas decyzyjnych. Współrzędna wektora $\bar{\mu}_{i,j}(x)$ jest równa średniej wartości podobieństwa obiektu klasyfikowanego x do obiektów relewantnych wybranych z klasy decyzyjnej v_j tablicy decyzyjnej agenta ag_i . Wartość $\bar{\mu}_{i,j}(x)$ wyraża się następującym wzorem

$$\bar{\mu}_{i,j}(x) = \frac{\sum_{y \in U_{ag_i}^{rel} \cap X_{v_j}^{ag_i}} s(x, y)}{\text{card}\{U_{ag_i}^{rel} \cap X_{v_j}^{ag_i}\}}, \quad i \in \{1, \dots, n\}, j \in \{1, \dots, c\}, \quad (17)$$

gdzie c jest liczbą klas decyzyjnych, $U_{ag_i}^{rel}$ jest zbiorem obiektów relewantnych z tablicy decyzyjnej agenta ag_i , czyli z każdej klasy decyzyjnej wybieranych jest co najmniej m_1 obiektów realizujących największe podobieństwo do obiektu klasyfikowanego x , $X_{v_j}^{ag_i}$ jest klasą decyzyjną z tablicy decyzyjnej agenta ag_i ; $s(x, y)$ jest wartością miary podobieństwa dla obiektów x i y .

Następnie na podstawie wektora $\bar{\mu}_i(x)$ generowany jest wektor rang $r_i(x) = [r_{i,1}(x), \dots, r_{i,c}(x)]$. Dla obiektu testowego x oraz każdej wartości decyzji v_j definiujemy funkcję $\phi_{v_j}^x : Ag \times Ag \rightarrow \{0, 1\}$ określoną następująco

$$\phi_{v_j}^x(ag_i, ag_k) = \begin{cases} 0 & \text{gdy } r_{i,j}(x) = r_{k,j}(x) \\ 1 & \text{gdy } r_{i,j}(x) \neq r_{k,j}(x) \end{cases} \quad (18)$$

gdzie $ag_i, ag_k \in Ag$. Następnie określamy natężenie konfliktu pomiędzy agentami za pomocą funkcji odległości agentów. Dla obiektu testowego x definiujemy odległość agentów $\rho^x : Ag \times Ag \rightarrow [0, 1]$ w następujący sposób

$$\rho^x(ag_i, ag_k) = \frac{\sum_{v_j \in V^d} \phi_{v_j}^x(ag_i, ag_k)}{\text{card}\{V^d\}}, \quad (19)$$

gdzie $ag_i, ag_k \in Ag$ oraz V^d jest zbiorem wartości decyzji.

Mówimy, że agenci $ag_i, ag_k \in Ag$ są w relacji przyjaźni względem klasyfikacji obiektu x , co zapisujemy $R^+(ag_i, ag_k)$, wtedy i tylko wtedy, gdy $\rho^x(ag_i, ag_k) < 0,5$. Agenci $ag_i, ag_k \in Ag$ są w relacji konfliktu względem klasyfikacji obiektu x , co zapisujemy $R^-(ag_i, ag_k)$, wtedy i tylko wtedy, gdy $\rho^x(ag_i, ag_k) \geq 0,5$. W systemie rozproszonym tworzone są grupy agentów pozostających w relacji przyjaźni.

W pracy [A2] zaproponowałam podejście, w którym tworzone są rozłączne grupy agentów. Proces tworzenia klastrow jest bardzo podobny do hierarchicznej aglomeracyjnej metody grupowania i przebiega w sposób następujący. Początkowo każdy agent jest traktowany jako oddzielny klastrow. Poniższe dwa kroki są wykonywane dopóki warunek stopu, który jest podany w pierwszym kroku, nie będzie spełniony.

1. Wybieramy parę różnych klastrow, dla których funkcja odległości przyjmuje wartość minimalną. Jeśli wybrana wartość funkcji odległości jest mniejsza niż 0,5, to łączymy agentów z wybranej pary klastrow w jeden nowy klastrow. W przeciwnym wypadku kończymy proces tworzenia klastrow.
2. Po zdefiniowaniu nowego klastra przeliczamy ponownie wartość funkcji odległości pomiędzy klastrami. Niech D_i będzie klastrem powstałym z połączenia dwóch klastrow $D_i = D_{i,1} \cup D_{i,2}$ oraz niech D_j będzie dowolnym klastrem, wówczas

$$\rho^x(D_i, D_j) = \begin{cases} \frac{\rho^x(D_{i,1}, D_j) + \rho^x(D_{i,2}, D_j)}{2} & \text{jeżeli } \rho^x(D_{i,1}, D_j) < 0,5 \\ & \text{oraz } \rho^x(D_{i,2}, D_j) < 0,5 \\ \max\{\rho^x(D_{i,1}, D_j), \rho^x(D_{i,2}, D_j)\} & \text{jeżeli } \rho^x(D_{i,1}, D_j) \geq 0,5 \\ & \text{lub } \rho^x(D_{i,2}, D_j) \geq 0,5 \end{cases} \quad (20)$$

W pracy [A2] podałam algorytm realizujący powyższą metodę tworzenia klastrow. Omówiłam również złożoność obliczeniową tego podejścia. Czas działania algorytmu jest liniowo zależny od liczby obiektów w tablicach lokalnych i wartości parametru m_1 oraz wielomianowo zależny od liczby agentów zasobów.

Ostatecznie decyzje generowane są przy wykorzystaniu aproksymacyjnej metody agregacji tablic decyzyjnych oraz metody algorytmu gęstościowego lub metody głosowania ważonego, które zostały omówione wcześniej.

Tabela 3: Błędy klasyfikacji – dynamiczna struktura systemu i rozłączne klastry

Tablice lokalne	Zbiór danych Soybean			Zbiór danych Vehicle			Zbiór danych Landsat Satellite		
	Dynamiczna		Statyczna	Dynamiczna		Statyczna	Dynamiczna		Statyczna
	TT	10 CV	TT	TT	10 CV	TT	TT	10 CV	TT
3	0,026	0,009	0,027	0,138	0,138	0,106	0,033	0,093	0,042
5	0,019	0,006	0,035	0,181	0,142	0,154	0,012	0,014	0,012
7	0,016	0,006	0,008	0,130	0,124	0,161	0,015	0,012	0,012
9	0,043	0,018	0,058	0,177	0,140	0,189	0,013	0,020	0,020
11	0,037	0,012	0,029	0,138	0,155	0,146	0,011	0,012	0,010

W pracy [A2], porównałam wyniki uzyskane przy zastosowaniu dynamicznej struktury systemu i rozłącznych klastrów z wynikami uzyskanymi przy zastosowaniu statycznej struktury systemu. Zrealizowałam eksperymenty przy wykorzystaniu metody train-and-test (TT) oraz metody 10-krotnej krosvalidacji (10 CV). W tabeli 3 podano błędy klasyfikacji (otrzymane dla większej średniej liczebności generowanych zbiorów decyzji). Porównując wyniki uzyskane dla dwóch metod testowania (train-and-test oraz 10-krotnej krosvalidacji) oraz podejścia z dynamiczną strukturą można zauważyć, że często metoda 10-krotnej krosvalidacji generuje nawet mniejsze wartości błędów klasyfikacji (oczywiście wyniki te obciążone są pewną zmiennością). Porównując wyniki uzyskane przy zastosowaniu dwóch podejść, dynamicznego i statycznego, można zauważyć, że w większości przypadków zastosowanie struktury dynamicznej wpłynęło na poprawę jakości wnioskowania. W pracy [A2], na podstawie wszystkich wyników, sformułowałam pewne wnioski ogólne. System z dynamiczną strukturą i rozłącznymi klastrami generuje zazwyczaj lepsze wyniki niż system ze statyczną strukturą. Im większa jest liczba lokalnych tablic decyzyjnych, tym większa poprawa jakości wnioskowania przy zastosowaniu struktury dynamicznej z rozłącznymi klastrami.

4.4.4 Rozproszony system ze strukturą dynamiczną i nierozłącznymi klastrami – identyfikacja relacji zachodzących pomiędzy klasyfikatorami bazowymi oraz poprawa jakości klasyfikacji [A3]

Praktycznie równoległe z poprzednim podejściem zaproponowałam kolejne podejście do generowania dynamicznej struktury systemu. Przy wykorzystaniu definicji relacji przyjaźni, konfliktu oraz funkcji odległości agentów, klastry mogą być generowane w inny sposób. Tym razem, klastry również będą grupami agentów pozostającymi w relacji przyjaźni. Jednak zaproponowałam nową definicję klastra, zgodnie z którą tworzone grupy agentów są zbiorami nierozłącznymi.

Klastrem jest maksymalny, pod względem zawierania, zbiór agentów zasobów spełniający warunek

$$\forall ag_i, ag_k \in Ag \quad R^+(ag_i, ag_k) \quad (21)$$

Metoda ta eliminuje wadę poprzedniego podejścia, która może być opisana poniższym przykładem. Załóżmy, że mamy trzech agentów zasobów ag_1, ag_2, ag_3 oraz, że agent ag_1 jest w relacji przyjaźni zarówno z agentem ag_2 , jak i ag_3 . Natomiast agenci ag_2 i ag_3 są w relacji konfliktu. Wówczas, w poprzednim podejściu, stworzony zostanie jeden klastrow zawierający dwóch agentów (ag_1 oraz jednego z pozostałych agentów) oraz jeden klastrow z pojedynczym

Tabela 4: Błędy klasyfikacji – dynamiczna struktura systemu i nierozłączne klastry

Tablice lokalne	Zbiór danych Soybean		Zbiór danych Vehicle		Zbiór danych Landsat Satellite	
	Nierozłączne klastry		Nierozłączne klastry		Nierozłączne klastry	
	TT	10 CV	TT	10 CV	TT	10 CV
	3	0,019	0,013	0,126	0,123	0,021
5	0,008	0,006	0,181	0,151	0,019	0,013
7	0,016	0,010	0,130	0,114	0,012	0,014
9	0,027	0,012	0,177	0,145	0,016	0,022
11	0,035	0,035	0,146	0,146	0,016	0,017

agentem (tym, który nie został uwzględniony w pierwszym klastrze). Jeśli dodatkowo założymy, że wartości funkcji odległości agentów dla par (ag_1, ag_2) oraz (ag_1, ag_3) są równe, to wybór agenta, który znajdzie się w jednym klastrze z agentem ag_1 będzie całkiem losowy i nieuzasadniony żadnymi logicznymi przesłankami. W nowym podejściu zostaną stworzone dwa klastry $\{ag_1, ag_2\}$ oraz $\{ag_1, ag_3\}$, co dużo lepiej odzwierciedla rozważaną sytuację. Zatem nowy sposób generowania klastrów poprawia zdolność systemu do analizy i identyfikacji relacji między agentami.

W pracy [A3] podałam algorytm realizujący powyższą metodę tworzenia klastrów. Idea algorytmu jest następująca. W zbiorze agentów zasobów szukamy pary agentów w relacji konfliktu. Jeśli taka para istnieje, to tworzymy dwa nowe zbiory poprzez usunięcie z rozważanego zbioru najpierw jednego, a następnie drugiego z agentów pozostających w konflikcie. Postępujemy w ten sposób tak długo, aż w rozważanym zbiorze nie będzie już agentów pozostających w relacji konfliktu. Algorytm ten ma wykładniczą złożoność obliczeniową ze względu na liczbę agentów zasobów. W najgorszym przypadku, w zbiorze agentów nie będzie istniała para agentów w relacji przyjaźni. Wówczas wszystkie podzbiory zbioru agentów zostaną wygenerowane. Jednakże z uwagi na to, że liczba agentów zasobów w systemie jest raczej niewielka, czas działania algorytmu też nie jest zbyt duży.

Ostatecznie decyzje generowane są przy wykorzystaniu aproksymacyjnej metody agregacji tablic decyzyjnych oraz metody algorytmu gęstościowego lub metody głosowania ważonego, które zostały omówione wcześniej.

W pracy [A3] porównałam wyniki otrzymane przy zastosowaniu podejścia dynamicznego z nierozłącznymi klastrami z wynikami uzyskanymi dla systemu z rozłącznymi klastrami oraz systemu ze strukturą statyczną. Zrealizowałam eksperymenty przy wykorzystaniu metody train-and-test (TT) oraz metody 10-krotnej krosvalidacji (10 CV). W tabeli 4 podano błędy klasyfikacji (otrzymane dla większej średniej liczebności generowanych zbiorów decyzji). Na podstawie tych wartości można stwierdzić, że metoda 10-krotnej krosvalidacji daje nie gorsze wyniki niż metoda train-and-test. Porównując wyniki zawarte w tabelach 3 i 4 można powiedzieć, że w przeważającej większości przypadków podejście z dynamicznymi nierozłącznymi klastrami generuje lepsze wyniki niż podejście z rozłącznymi klastrami oraz podejście statyczne. W pracy [A3] sformułowałam pewne wnioski ogólne. Najlepsze wyniki generuje system z dynamiczną strukturą oraz nierozłącznymi klastrami. Podejście z rozłącznymi klastrami generuje wyniki lepsze niż podejście ze statyczną strukturą.

4.4.5 Rozproszony system ze strukturą dynamiczną i etapem negocjacji – poprawa identyfikacji relacji zachodzących pomiędzy klasyfikatorami bazowymi [A4, B2]

Po pewnym czasie zauważyłam potrzebę rozszerzenia proponowanych wcześniej podejść dynamicznych. We wcześniejszych podejściach, w zależności od wartości funkcji odległości rozpoznaje się relację konfliktu lub przyjaźni między agentami (definicje relacji podane zostały w rozdziale 4.4.3). Jednak zależność ta jest mało elastyczna, polega na sprawdzeniu czy wartość funkcji odległości przekracza 0,5 czy też nie. Ponadto relacje między agentami będą lepiej identyfikowane, jeśli podczas generowania klastrów wykorzystany zostanie trzeci typ relacji – relacja neutralności między agentami. Zaproponowałam nowe podejście, w ramach którego opracowałam dwuetapowy proces tworzenia dynamicznej struktury systemu, podałam nowe definicje relacji, zaproponowałam etap negocjacji, w którym nатеżenie konfliktu między agentami wyrażane jest za pomocą uogólnionej funkcji odległości. Wszystko to miało na celu, aby opracowana metoda była dużo bardziej czuła i dokładna w rozpoznawaniu relacji między agentami.

Przyjęłam następujące definicje relacji między agentami. Niech dana będzie pewna liczba rzeczywista p należąca do przedziału $[0;0,5)$. Mówimy, że agenci $ag_i, ag_k \in Ag$ są w relacji przyjaźni względem klasyfikacji obiektu x , co zapisujemy $R^+(ag_i, ag_k)$, wtedy i tylko wtedy, gdy $\rho^x(ag_i, ag_k) < 0,5 - p$. Agenci $ag_i, ag_k \in Ag$ są w relacji konfliktu względem klasyfikacji obiektu x , co zapisujemy $R^-(ag_i, ag_k)$, wtedy i tylko wtedy, gdy $\rho^x(ag_i, ag_k) > 0,5 + p$. Natomiast agenci $ag_i, ag_k \in Ag$ są w relacji neutralności względem klasyfikacji obiektu x , co zapisujemy $R^0(ag_i, ag_k)$, wtedy i tylko wtedy, gdy $0,5 - p \leq \rho^x(ag_i, ag_k) \leq 0,5 + p$.

W pierwszym kroku tworzenia struktury systemu generowane są klastry wstępne zawierające agentów pozostających w relacji przyjaźni. Klastrem wstępnym nazywamy maksymalny, pod względem zawierania, zbiór agentów zasobów spełniający warunek

$$\forall ag_i, ag_k \in Ag \quad R^+(ag_i, ag_k) \quad (22)$$

Jak można zauważyć, definicja klastra wstępnego jest tożsama z definicją klastra w podejściu dynamicznym z nierozłącznymi klastrami. Jednak tym razem otrzymamy zupełnie inne zbiory, gdyż dodatkowo wykorzystywana jest relacja neutralności. Po realizacji pierwszego etapu tworzenia struktury systemu otrzymujemy zbiór klastrów wstępnych i zbiór agentów, którzy nie zostali przyłączeni do żadnego klastra. W drugim etapie tworzenia struktury systemu – etapie negocjacji – próbuje się dołączyć agentów bezkoalicyjnych do istniejących klastrów wstępnych. Aby, na samym początku, nie osiągnąć sytuacji patowej, należy złagodzić warunki przyłączania agentów do klastrów. Jest to strategia, którą stosuje się w negocjacjach. W drugim etapie tworzenia struktury systemu wykorzystywana jest uogólniona funkcja odległości agentów, która wyznacza odległość wyłącznie na podstawie najbardziej znaczących decyzji.

Dla obiektu klasyfikowanego x definiujemy funkcję $\phi_G^x : Ag \times Ag \rightarrow [0, \infty)$

$$\phi_G^x(ag_i, ag_j) = \frac{\sum_{v_l \in Sign_{i,j}} |r_{i,l}(x) - r_{j,l}(x)|}{card\{Sign_{i,j}\}} \quad (23)$$

gdzie $ag_i, ag_j \in Ag$ oraz $Sign_{i,j}$ jest zbiorem wartości atrybutu decyzyjnego, które są, w zadanym sensie, znaczące dla pary agentów ag_i, ag_j . W zbiorze $Sign_{i,j}$ uwzględnione

zostaną decyzje, którym agenci ag_i, ag_j przyznali najwyższe rangi. W pracy [A4] przyjąłam, że będą to decyzje z rangą 1.

Natężenie konfliktu pomiędzy dwoma zbiorami agentów jest określone za pomocą uogólnionej odległości. Dla obiektu klasyfikowanego definiujemy funkcję $\rho_G^x : 2^{Ag} \times 2^{Ag} \rightarrow [0, \infty)$

$$\rho_G^x(X, Y) = \begin{cases} 0 & \text{jeżeli } \text{card}\{X \cup Y\} \leq 1 \\ \frac{\sum_{ag, ag' \in X \cup Y} \phi_G^x(ag, ag')}{\text{card}\{X \cup Y\} \cdot (\text{card}\{X \cup Y\} - 1)} & \text{w przeciwnym razie} \end{cases} \quad (24)$$

gdzie $X, Y \subseteq Ag$. Jak można zauważyć, wartość $\rho_G^x(X, Y)$ jest równa średniej wartości funkcji ϕ_G^x dla każdej pary agentów należących do zbioru $X \cup Y$. Czyli jest to średnia różnica rang przypisanych najbardziej znaczącym decyzjom dla agentów ze zbioru $X \cup Y$.

Dla każdego agenta bezkoalicyjnego ag obliczana jest wartość uogólnionej odległości $\rho_G^x(\{ag\}, Y)$, gdzie Y jest klastrem wstępnym takim, że agent ag nie jest w relacji konfliktu z żadnym agentem należącym do tego klastra lub Y jest jednoelementowym zbiorem składającym się z innego agenta bezkoalicyjnego, z którym agent ag nie jest w relacji konfliktu. Następnie agent ag jest przyłączany do wszystkich klastrów wstępnych, dla których uogólniona odległość nie przekracza pewnej wartości progowej. Ponadto tworzone są nowe klastry składające się z agentów bezkoalicyjnych, dla których również uogólniona odległość nie przekracza wartości progowej. W pracy [A4] zaproponowałam wzór dla ustalenia wartości progowej

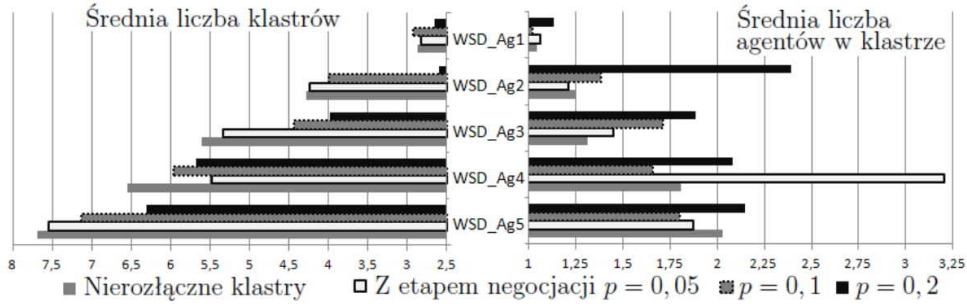
$$\lfloor \log_3(\text{card}\{V^d\}) \rfloor + 1, \quad (25)$$

gdzie V^d jest zbiorem wartości atrybutu decyzyjnego takim, że $\text{card}\{V^d\} \geq 2$. Sprawdziłam eksperymentalnie, że taka wartość progowa daje dobre wyniki.

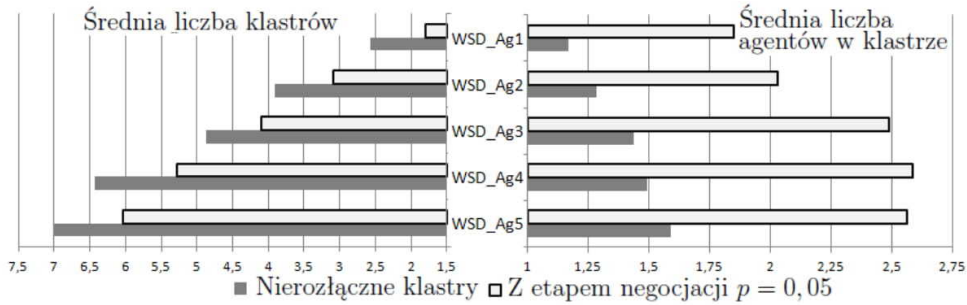
W pracy [A4] podałam algorytm realizujący metodę tworzenia klastrów z etapem negocjacji. Algorytm generowania klastrów wstępnych ma wykładniczą złożoność obliczeniową ze względu na liczbę agentów zasobów (jak to zostało uzasadnione w poprzednim podrozdziale). Algorytm realizujący drugi etap tworzenia klastrów również ma wykładniczą złożoność obliczeniową ze względu na liczbę agentów zasobów. W najgorszym przypadku, po realizacji pierwszego etapu, otrzymamy tylko agentów bezkoalicyjnych. Wtedy, może się zdarzyć, że będziemy analizować wszystkie możliwe podzbiory zbioru agentów zasobów. Jednak, jak to już zostało wspomniane wcześniej, liczba agentów zasobów w systemie jest raczej niewielka, zatem czas działania algorytmu też nie jest zbyt duży.

Ostatecznie decyzje generowane są przy wykorzystaniu aproksymacyjnej metody agregacji tablic decyzyjnych oraz metody algorytmu gęstościowego lub metody głosowania ważonego, które zostały omówione wcześniej.

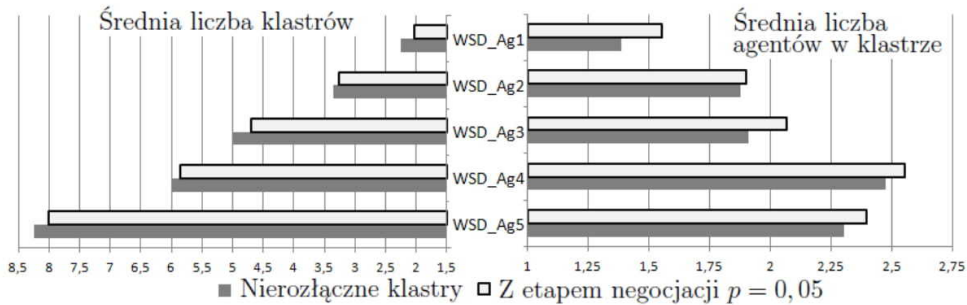
W pracy [A4] porównałam strukturę klastrów tworzonych przy wykorzystaniu nowego dynamicznego podejścia z etapem negocjacji ze strukturą klastrów uzyskanych przy zastosowaniu podejścia z nierozłącznymi klastrami. Na rysunkach 2, 3 oraz 4 przedstawiono średnią liczbę klastrów oraz średnią liczbę agentów zasobów w klastrze dla zbiorów danych Soybean, Vehicle i Landsat Satellite odpowiednio. Jak można zauważyć na podstawie graficznej reprezentacji, większa liczba klastrów jest tworzona przy zastosowaniu podejścia



Rysunek 2: Porównanie struktury klastrów tworzonych dla podejść z etapem negocjacji oraz nierozłącznych klastrów (Soybean)



Rysunek 3: Porównanie struktury klastrów tworzonych dla podejść z etapem negocjacji oraz nierozłącznych klastrów (Vehicle Silhouettes)



Rysunek 4: Porównanie struktury klastrów tworzonych dla podejść z etapem negocjacji oraz nierozłącznych klastrów (Landsat Satellites)

z nierozłącznymi klastrami. Jednak klastry te są mniejsze – składają się z mniejszej liczby agentów zasobów. Wykonałam również bardziej szczegółowe analizy (dotyczące poszczególnych klastrów), okazało się, że dla nierozłącznych klastrów zostało stworzonych wiele

klastrów składający się tylko z jednego agenta. Podejście z etapem negocjacji pozwala na stworzenie mniejszej liczby, jednak bardziej złożonych i większych klastrów. Ponadto zauważyłam, że przy wykorzystaniu nowej metody większa jest liczba agentów, którzy jednocześnie zaliczeni zostali do kilku klastrów. Jest to konsekwencja zastosowania etapu negocjacji, w którym agenci neutralni są przyłączani do wszystkich klastrów wstępnych, z którymi ich relacje są wystarczająco dobre. Stwierdziłam, że w nowym podejściu klastry są bardziej złożone i lepiej rekonstruują i ilustrują poglądy agentów na klasyfikację.

Jednak, mimo znaczącej różnicy w strukturze systemu nie uzyskałam równie znaczącej zmiany w jakości wnioskowania. Stwierdziłam, że prawdopodobnie przyczyną takiego stanu rzeczy jest stosowana metoda analizy konfliktów, w której nie jest brana pod uwagę siła klastrów ani ich struktura. Zatem, zarówno małe jak i duże klastry mają taki sam wpływ na ostateczną decyzję. Ponadto nie uwzględnia się stopnia zaangażowania agenta, który należy jednocześnie do kilku klastrów.

4.4.6 Poprawa jakości klasyfikacji rozproszonego systemu ze strukturą dynamiczną i etapem negocjacji poprzez zastosowanie wag klastrów [A5]

W kolejnym etapie mojej pracy zaproponowałam metodę analizy konfliktu, która uwzględniłaby złożoną strukturę klastrów generowanych przy zastosowaniu podejścia z etapem negocjacji i przełożyłaby zalety tej metody na jakość klasyfikacji.

Rozważałam system z dynamiczną strukturą i etapem negocjacji, który został omówiony w poprzednim podrozdziale. W systemie tym wykorzystywana jest metoda aproksymacyjnej agregacji tablic decyzyjnych. Na podstawie każdej j -tej tablicy zagregowanej generowane są wektory wartości $\mu_j(x)$, które zostały zdefiniowane w rozdziale 4.3. Następnie, tak jak poprzednio wykonywane jest przekształcenie liniowe na tych wektorach, jednak przy zastosowaniu pewnych wag. Ostatecznie realizowana jest metoda algorytmu gęstościowego.

Proponowana nowa metoda polega na opracowaniu sposobu wyliczania wag klastrów. Zaproponowałam cztery różne metody wyliczania wag, każda z nich uwzględnia inny aspekt budowy i struktury klastrów.

Metoda 1 – siła klastrów (ang. *cluster strength*). W metodzie tej waga klastra jest wyrażona liczbą agentów wchodzących w jego skład. Jednak, z uwagi na to, że jeden agent może być jednocześnie uwzględniony w kilku klastrach, wylicza się współczynnik zaangażowania agenta ag w tworzenie klastra m_{ag}^x . Współczynnik ten jest odwrotnie proporcjonalny do liczby klastrów, do których należy dany agent. Waga przypisana do klastra jest równa sumie wartości współczynników zaangażowania agentów zasobów należących do tego klastra. Dzięki temu duże klastry mają większy wpływ na ostateczną decyzję, podczas gdy wpływ małych klastrów jest zredukowany.

Metoda 2 – zdecydowanie agentów (ang. *decisiveness of agents*). W metodzie tej waga klastra jest zależna do jednoznaczności decyzji podjętych przez agentów zasobów w danym klastrze. Decyzje i -tego agenta zasobów są wyrażone w postaci wektora $\bar{\mu}_i(x)$, zdefiniowanego za pomocą formuły 17. Jeśli współrzędne tego wektora mają bardzo zbliżone wartości, oznacza to, że agent zasobów nie jest zdecydowany co do klasyfikacji obiektu. Nie potrafi wskazać jednej wartości decyzji lecz ocenia, że prawdopodobieństwo zaklasyfikowania obiektu do kilku różnych klas decyzyjnych jest praktycznie takie samo. Zatem wiedza jaką

Tabela 5: Porównanie liczby najlepszych wyników

Zbiór danych	Metoda 1	Metoda 2	Metoda 3	Metoda 4	Dynamiczne klastry z etapem negocjacji
Soybean	4	4	2	0	0
Vehicle Silhouettes	1	2	5	1	1
Landsat Satellite	4	0	1	4	0
Σ	9	6	8	5	1

dostarcza ten agent do systemu jest mało wartościowa. Natomiast, gdy wartości współczynników wektora są mocno zróżnicowane, oznacza to, że agent jest pewien tego, w jaki sposób zaklasyfikować obiekt. Taka wiedza jest dużo bardziej cenna. Zatem dla każdego agenta zasobów ag_i , wyliczane jest odchylenie standardowe wartości współczynników wektora $\bar{\mu}_i(x)$. Waga przypisana do klastra jest równa średniej wartości odchyleń standardowych wektorów agentów zasobów należących do tego klastra. Dzięki temu wzrasta wpływ tych agentów na postać ostatecznej decyzji, którzy są pewni i przekonani co do klasyfikacji obiektu do jednej z klas decyzyjnych.

Metoda 3 – połączenie siły klastrów i zdecydowania agentów (ang. *mixed cluster strength and decisiveness of agents*). Ta metoda jest połączeniem dwóch powyższych metod. Dla każdego agenta zasobów wyliczany jest współczynnik, który jest iloczynem zaangażowania agenta (Metoda 1) oraz odchylenia standardowego wektora $\bar{\mu}$ (Metoda 2). Waga przypisana do klastra jest równa średniej wartości współczynników agentów zasobów należących do tego klastra. W ten sposób wagi uwzględniają nie tylko wielkość klastra, lecz również to, czy agenci należący do tego klastra mają umiejętność podejmowania jednoznacznych decyzji.

Metoda 4 – siła klastra oparta na jego zdecydowaniu (ang. *decisiveness-based cluster strength*). W metodzie tej dla każdego j -tego agenta syntezy obliczane jest odchylenie standardowe współczynników wektora $\mu_j(x)$ (definicja wektora podana została we wprowadzeniu) wygenerowanego na podstawie tablicy zagregowanej. Uzasadnienie tej metody jest analogiczne jak dla Metody 2. W ten sposób grupa agentów (klastr), która podejmuje kilka różnych decyzji z podobnym współczynnikiem pewności ma mniejszy wpływ na ostateczną decyzję. Natomiast wpływ klastrów, które podejmują bardziej jednoznaczne decyzje, zwiększa się.

W pracy [A5] porównałam wyniki uzyskane przy zastosowaniu czterech metod wyliczania wag klastrów z wynikami otrzymanymi bez wyliczania wag klastrów (podejście dynamiczne z etapem negocjacji). Nie przedstawiam tutaj szczegółowych rezultatów, lecz pewne podsumowanie ograniczone do liczby najlepszych wyników, jakie zostały wygenerowane przy zastosowaniu poszczególnych podejść (tabela 5). Można jednoznacznie stwierdzić, że zastosowanie wag wpłynęło na poprawę jakości klasyfikacji. Na podstawie informacji zamieszczonych w tabeli 5 można stwierdzić, że najlepsze wyniki uzyskano dla metod 1 oraz 3. Jednakże proporcje dla poszczególnych zbiorów danych są bardzo różne. To skłoniło mnie do przeprowadzenia bardziej szczegółowej analizy, polegającej na dokładnym sprawdzeniu struktury klastrów generowanych dla tych różnych zbiorów danych. Na podstawie tej analizy, w pracy [A5], podałam pewne ogólne wytyczne określające przypadki, w których należy zastosować jedną z zaproponowanych metod wyliczania wag klastrów. Główne wytyczne są następujące:

- Metoda 1 – stosujemy, gdy w rozproszonym zbiorze danych występuje duże zróżnicowanie liczby agentów w generowanych klastrach (liczność klastrów jest bardzo różna) oraz, gdy agenci są jednocześnie uwzględniani w wielu klastrach.
- Metoda 2 – stosujemy, gdy w rozproszonym zbiorze danych występują agenci niezdecydowani (ang. *dummy agents*), to znaczy tacy, którzy nie potrafią podjąć jednoznacznej decyzji.
- Metoda 3 – stosujemy, gdy zastosowanie powyższych dwóch metod nie jest oczywiste, to znaczy, gdy klastry nie są aż tak bardzo zróżnicowane pod względem wielkości oraz, gdy odchylenie standardowe współczynników wektora agentów nie jest aż tak znaczące.
- Metoda 4 – stosujemy, gdy w rozproszonym zbiorze danych, w zbiorze agentów syntezy, występują agenci niezdecydowani.

4.4.7 Zastosowanie metod fuzji klasyfikatorów bazowych znanych z literatury w rozproszonym systemie ze strukturą dynamiczną i etapem negocjacji [A6–A8, B3]

W omówionych powyżej podejściach, w ostatnim etapie generowania decyzji, stosowano metodę algorytmu gęstościowego, która została zaproponowana w mojej rozprawie doktorskiej [39]. Celem kolejnego etapu mojej pracy było zaadoptowanie metod fuzji, znanych z literatury, do systemu rozproszonego. Zaproponowałam nowe podejście polegające na wprowadzeniu licznych modyfikacji do metod znanych z literatury tak, aby można było je zastosować w systemie rozproszonym. Ponadto zaproponowałam metodę, w której wiele różnych metod fuzji jest stosowanych jednocześnie, a ostateczne decyzje podejmowane są poprzez głosowanie zespołu metod.

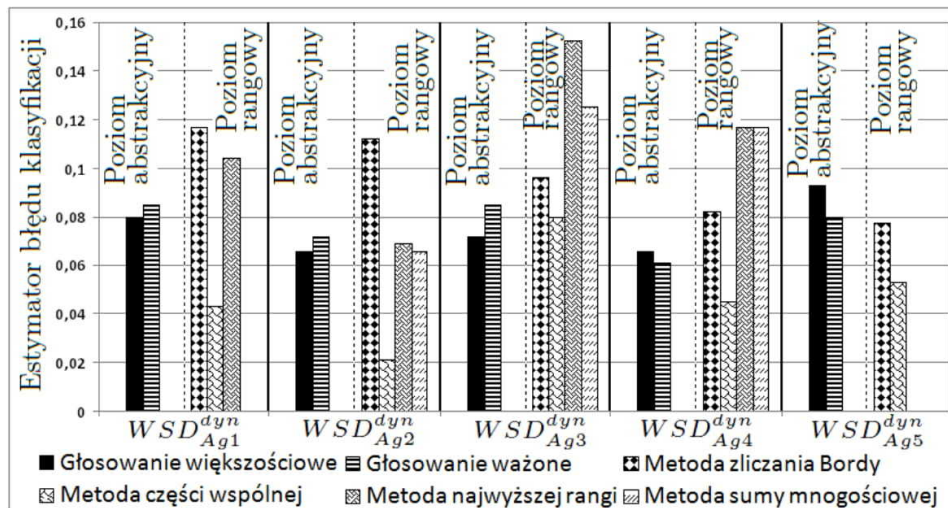
Różne metody łączenia wyników predykcji klasyfikatorów bazowych są wykorzystywane w zależności od tego, w jakiej postaci klasyfikatory bazowe generują decyzje. W literaturze [14, 24, 56], wyróżnia się trzy typy wyników predykcji. Typ 1 – poziom abstrakcyjny (ang. *abstract level*), w którym klasyfikator bazowy generuje jedną wartość decyzji, do której należy klasyfikowany obiekt. Typ 2 – poziom rangowy (ang. *rank level*), w którym klasyfikator bazowy generuje zbiór decyzji uporządkowany zgodnie z prawdopodobieństwem, że decyzja jest poprawną klasą decyzyjną dla obiektu. Typ 3 – poziom pomiarowy (ang. *measurement level*), w którym klasyfikator bazowy generuje wektor wartości reprezentujący prawdopodobieństwo klasyfikowania obiektu do danej klasy decyzyjnej.

Zastosowałam metody fuzji w systemie rozproszonym z etapem negocjacji, który został omówiony w rozdziale 4.4.5. W systemie tym wykorzystywana jest aproksymacyjna metoda agregacji tablic decyzyjnych. Na podstawie każdej j -tej tablicy decyzyjnej generowany jest wektor wartości $\mu_j(x) = [\mu_{j,1}(x), \dots, \mu_{j,c}(x)]$, gdzie c jest liczbą klas decyzyjnych, a współrzędna $\mu_{j,i}(x)$ jest średnią wartością miary podobieństwa obiektu klasyfikowanego do m_3 obiektów relewantnych wybranych z klasy decyzyjnej X_{v_i} , tablicy decyzyjnej agenta syntezy as_j . Ostatecznie wektory $\mu_j(x)$ podlegają agregacji z wykorzystaniem metod fuzji i generowane są decyzje globalne.

Zatem otrzymujemy wyniki predykcji typu pomiarowego. Jednak w prosty sposób mogą one zostać sprowadzone do typu rangowego lub typu abstrakcyjnego. W pracach [A6–A8,

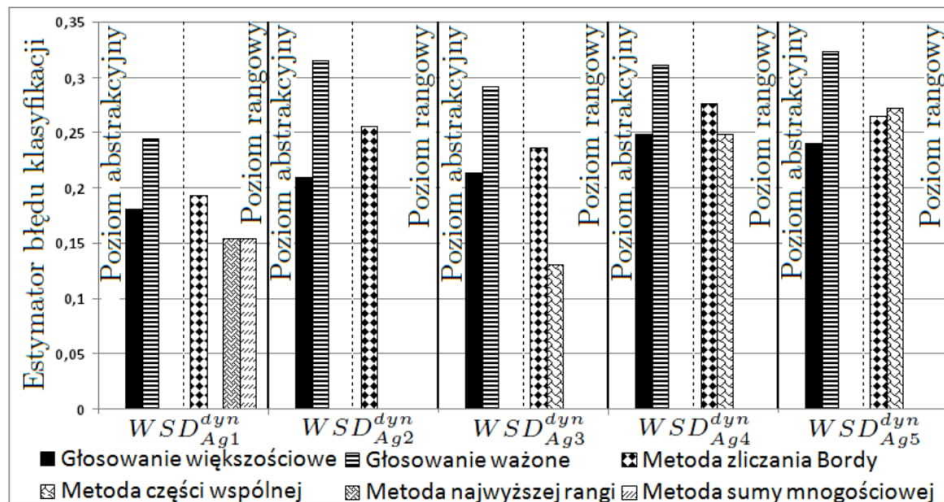
B3] rozważałam w sumie piętnaście metod fuzji z wszystkich trzech typów. Dwie metody z poziomu abstrakcyjnego – głosowanie większościowe (ang. *majority vote*) oraz głosowanie ważone (ang. *weighted majority vote*). Cztery metody z poziomu rangowego – metoda zliczania Bordy (ang. *Borda count*), metoda części wspólnej (ang. *intersection method*), metoda najwyższej rangi (ang. *highest rank method*) oraz metoda sumy mnogościowej (ang. *union method*). Dziewięć metod z poziomu pomiarowego – reguła maksimum, minimum, mediany, sumy, iloczynu (ang. *maximum, minimum, median, sum and product rule*), metoda średniej ważonej (ang. *weighted average method*), metoda iloczynu probabilistycznego (ang. *probabilistic product method*), metoda szablonów decyzyjnych (ang. *method using decision templates*) oraz metoda oparta na teorii przekonania (ang. *method that is based on the theory of evidence*).

W pracy [A8] podałam kompleksowe porównanie metod poziomu abstrakcyjnego i poziomu rangowego. Na rysunkach 5, 6 i 7 przedstawiono porównanie wartości błędu klasyfikacji otrzymane dla zbiorów danych Soybean, Vehicle oraz Landsat Satellite odpowiednio. Czasem metody fuzji, dla niektórych wersji rozproszenia, generowały tak niejednoznaczne wyniki (duża średnia liczebność generowanych zbiorów decyzji), że były one bezużyteczne. W tych przypadkach, na wykresach nie zaznaczono wartości błędu klasyfikacji. Na

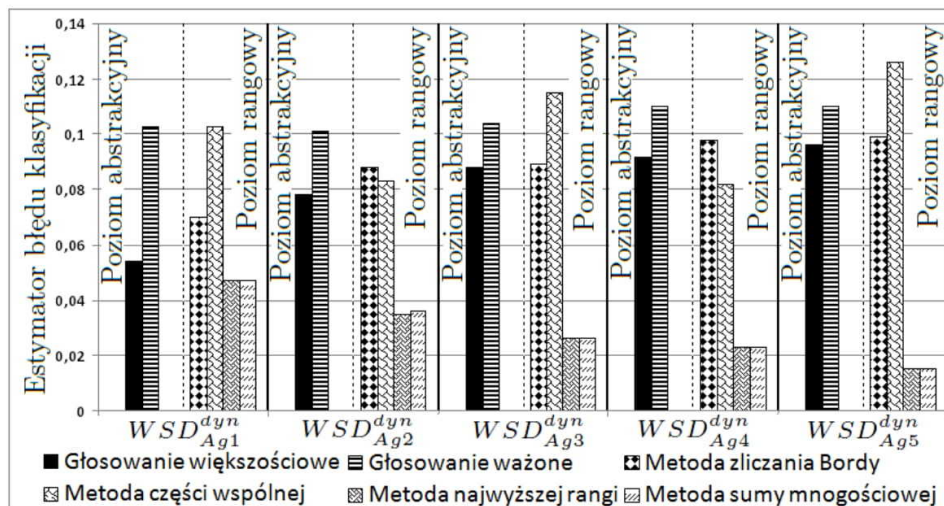


Rysunek 5: Porównanie błędu klasyfikacji – poziom abstrakcyjny i poziom rangowy (Soybean)

podstawie zaprezentowanych wyników, można stwierdzić, że czasem najprostsza metoda – głosowania większościowego – z poziomu abstrakcyjnego osiąga najlepsze wyniki. Jednakże, szczególnie dla dużych zbiorów danych (Landsat Satellites), metody z poziomu rangowego (metoda najwyższej rangi oraz metody sumy mnogościowej) generują lepsze wyniki. W przypadku zbiorów danych, w których jest dużo niezdecydowanych (ang. *dummy agents*) agentów (Vehicle Silhouettes), metoda części wspólnej, metoda najwyższej rangi oraz metoda sumy mnogościowej mają tendencję do generowania niejednoznacznych



Rysunek 6: Porównanie błędu klasyfikacji – poziom abstrakcyjny i poziom rangowy (Vehicle Silhouettes)

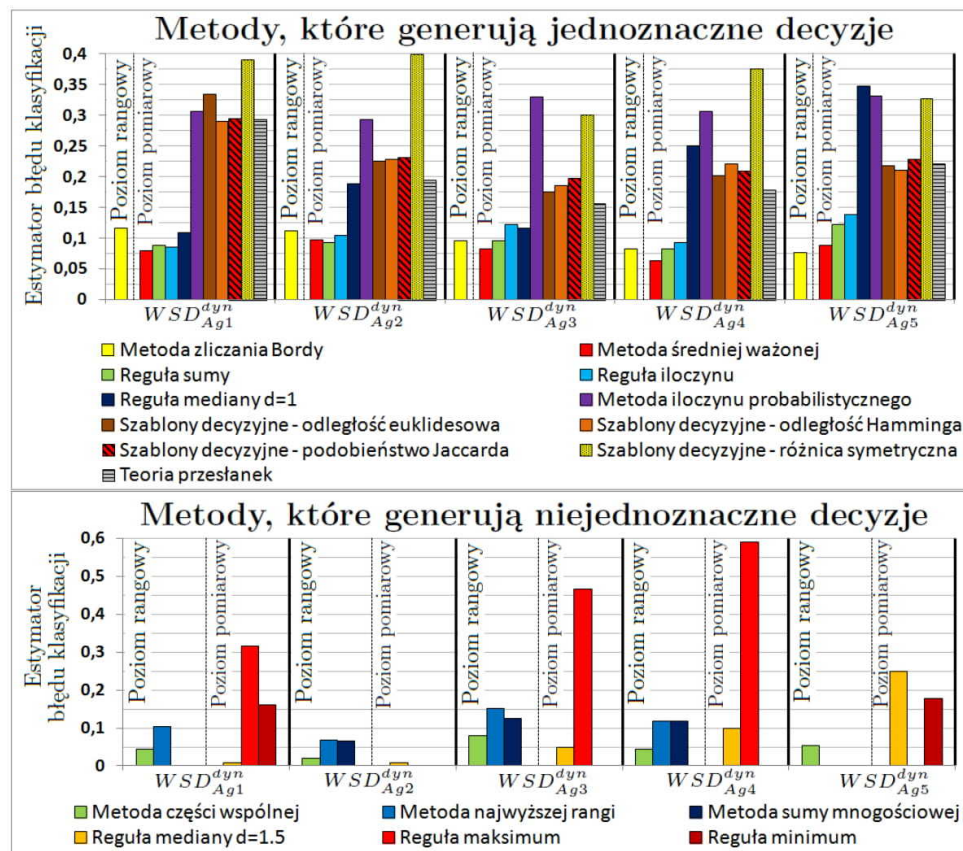


Rysunek 7: Porównanie błędu klasyfikacji – poziom abstrakcyjny i poziom rangowy (Land-sat Satellites)

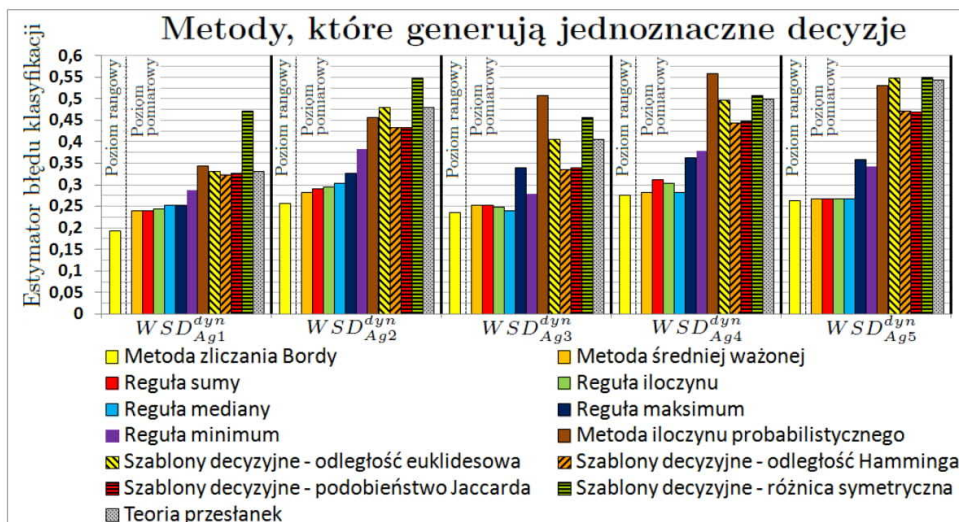
wyników. Dla takich zbiorów danych metody te są bezużyteczne. W tej sytuacji, proste metody fuzji dają zaskakująco dobre wyniki. W pracy [A8] analizowałam również wyniki przy zastosowaniu metod fuzji bezpośrednio do wektorów wartości generowanych na pod-

stawie tablic lokalnych (czyli bez rozpatrywania relacji zachodzących między tablicami lokalnymi i stosowania struktury hierarchicznej). Stwierdziłam, że w większości przypadków zastosowanie systemu rozproszonego wpłynęło na poprawę jakości klasyfikacji.

W publikacjach [A6] i [B3] porównałam pewne wstępne wyniki, uzyskane przy zastosowaniu siedmiu metod z poziomu pomiarowego. W pracy [A7] wykonałam kompleksowe porównanie metod poziomu rangowego i poziomu pomiarowego. Na rysunkach 8, 9 oraz 10 przedstawiono wartości błędu klasyfikacji otrzymane dla zbiorów danych Soybean, Vehicle oraz Landsat Satellite odpowiednio. Dla zbiorów danych Soybean oraz Landsat Satellite, rezultaty zostały zaprezentowane w dwóch częściach – dla jednoznacznych wyników (średnia liczebność generowanych zbiorów decyzji jest równa jeden) – dla niejednoznacznych wyników (średnia liczebność generowanych zbiorów decyzji jest większa niż jeden, ale wielkość ta jest akceptowalna). Dla zbioru danych Vehicle generowane były jednoznaczne wyniki lub niejednoznaczność była tak duża, że wyników nie rozważano.

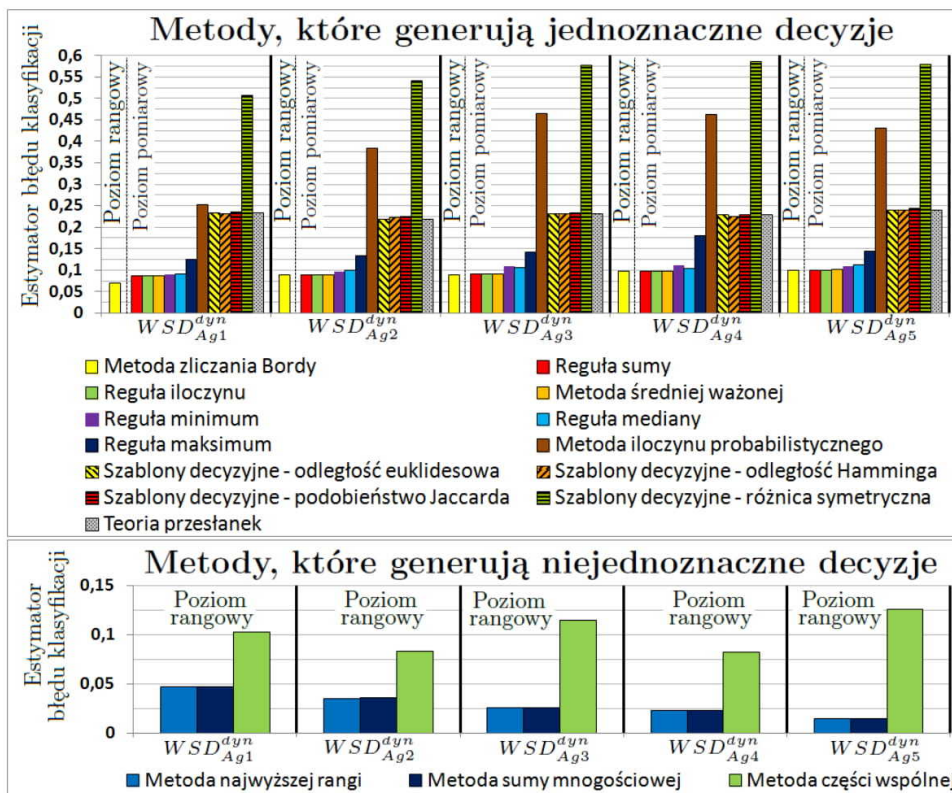


Rysunek 8: Porównanie błędu klasyfikacji – poziom rangowy i poziom pomiarowy (Soybean)



Rysunek 9: Porównanie błędu klasyfikacji – poziom rangowy i poziom pomiarowy (Vehicle Silhouettes)

Na podstawie otrzymanych rezultatów wyciągnęłam następujące wnioski ogólne. Metody z poziomu pomiarowego generują, w większości przypadków, jednoznaczne wyniki. Metody z poziomu rangowego generują raczej niejednoznaczne wyniki (z wyjątkiem metody zliczania Bordy). Ta niejednoznaczność jest, w niektórych przypadkach, tak duża, że rezultaty są bezużyteczne. Jeśli w zbiorze danych nie ma niezdecydowanych agentów, to najlepszą metodą, spośród metod, które generują jednoznaczne wyniki, jest metoda zliczania Bordy. Jednakże, tylko odrobinę gorsze wyniki otrzymujemy przy zastosowaniu reguły sumy, reguły iloczynu oraz metody średniej ważonej. Spośród metod generujących niejednoznaczne wyniki, najlepsze są metoda najwyższej rangi i metoda sumy mnogościowej. Jeżeli w zbiorze danych występują niezdecydowani agenci, najlepszą spośród metod generujących jednoznaczne wyniki jest metoda średniej ważonej. Reguła mediany lub metoda części wspólnej są najlepsze spośród metod generujących niejednoznaczne wyniki. W pracy [A7] rozważałam również podejście, w którym wiele różnych metod fuzji jest wykorzystywanych jednocześnie, a ostateczna decyzja jest wyznaczana poprzez proste głosowanie tych metod. Stosowałam metody w dwóch grupach: z poziomu rangowego lub z poziomu pomiarowego. Stwierdziłam, że lepsze wyniki są generowane przy jednoczesnym wykorzystaniu metod z poziomu rangowego. Ponadto uzyskane w ten sposób wyniki są lepsze niż najgorsze i średnie wyniki otrzymane dla poszczególnych metod. W pracy analizowałam również wyniki przy zastosowaniu metod fuzji bezpośrednio do wektorów wartości generowanych na podstawie tablic lokalnych. Ponownie stwierdziłam, że w większości przypadków zastosowanie systemu rozproszonego wpłynęło na poprawę jakości klasyfikacji.



Rysunek 10: Porównanie błędu klasyfikacji – poziom rangowy i poziom pomiarowy (Landsat Satellites)

4.4.8 Adaptacja modelu konfliktu Profesora Zdzisława Pawłaka do zagadnienia klasyfikacji z wykorzystaniem zespołu klasyfikatorów [B4]

Jak już wspomniano wcześniej, jedną z motywacji do podjęcia problemu analizy konfliktów w zagadnieniach klasyfikacji na podstawie wiedzy rozproszonej był model zaproponowany przez Profesora Pawłaka. Powyżej opisano podejścia, które znacząco różnią się od tego modelu zarówno pod względem stosowanych sposobów obliczania natężenia konfliktu, jak i samej definicji klastra (koalicji, klik). W kolejnym etapie mojej pracy zastosowałam bezpośrednio model Profesora Pawłaka (opisany w rozdziale 4.3) do zagadnienia klasyfikacji na podstawie wiedzy rozproszonej. Pierwotny zakres zastosowań modelu nie obejmował problematyki z zakresu zespołu klasyfikatorów, i według mojej najlepszej wiedzy, do tej pory nie podjęto próby wykorzystania modelu Pawłaka w tym celu. Zaproponowałam nowe podejście, które obejmuje zdefiniowanie dwóch sposobów przekształcenia wyników predykcji klasyfikatorów bazowych w system informacyjny wykorzystywany w modelu Pawłaka. Ponadto, zaproponowałam trzy różne metody generowania grup klasyfikatorów. Techniki

te powstały poprzez połączenie sposobów definiowania systemu informacyjnego z dwoma różnymi metodami określania natężenia konfliktów – za pomocą funkcji odległości (wzór 3) lub funkcji konfliktu (wzór 4). Grupy klasyfikatorów, zdefiniowane przy zastosowaniu powyższych metod, są następnie klastrami w systemie rozproszonym. Na podstawie wiedzy klastrów generowane są tablice zagregowane, przy wykorzystaniu aproksymacyjnej metody agregacji tablic decyzyjnych. Ostateczne decyzje są podejmowane przy użyciu algorytmu gęstościowego.

Przejdziemy teraz do dokładniejszego omówienia zaproponowanych metod. Podobnie jak poprzednio zakładamy, że dana jest pewna wiedza rozproszona w postaci zbioru tablic decyzyjnych. Dla danego obiektu klasyfikowanego, na podstawie każdej z tablic generowany jest wektor rang. W ogólności proponowane podejście może być zastosowane do dowolnej rodziny klasyfikatorów, w której każdy z nich generuje wektor rang. W moich rozważaniach przyjąłam, że wektory te są definiowane tak jak to opisano w rozdziale 4.4.3. Następnie przekształcamy wektory rang w system informacyjny $S = (U, A)$, gdzie U jest zbiorem agentów zasobów (w ogólnym przypadku zbiorem klasyfikatorów), A jest zbiorem wartości atrybutu decyzyjnego. Zaproponowałam dwie metody, których celem jest odzworowanie poglądów agentów, które reprezentowane są w postaci wektorów rang, na system informacyjny, w którym wykorzystujemy jedynie trzy wartości atrybutów, aby przedstawić opinie agentów (-1 agent jest przeciwny, 0 agent jest neutralny, 1 agent jest za). W pierwszej metodzie definiujemy atrybuty następująco, dla każdego $a \in A$, $a : U \rightarrow \{-1, 0, 1\}$

$$a(ag) = \begin{cases} 1 & \text{jeżeli } r_{ag,a}(x) = 1 \\ -1 & \text{jeżeli } r_{ag,a}(x) > 1 \end{cases} \quad (26)$$

W drugiej metodzie definiujemy $a : U \rightarrow \{-1, 0, 1\}$ dla każdego $a \in A$

$$a(ag) = \begin{cases} 1 & \text{jeżeli } r_{ag,a}(x) = 1 \\ 0 & \text{jeżeli } r_{ag,a}(x) = 2 \\ -1 & \text{jeżeli } r_{ag,a}(x) > 2 \end{cases} \quad (27)$$

W pierwszej metodzie, agenci są za wartością atrybutu decyzyjnego jedynie, w przypadku, gdy decyzja ta otrzymała najwyższą rangę. Agent jest przeciwko wszystkim pozostałym decyzjom. W drugiej metodzie, opinie agentów zostały złagodzone. Tym razem, agenci są neutralni wobec decyzji, które otrzymały rangę 2.

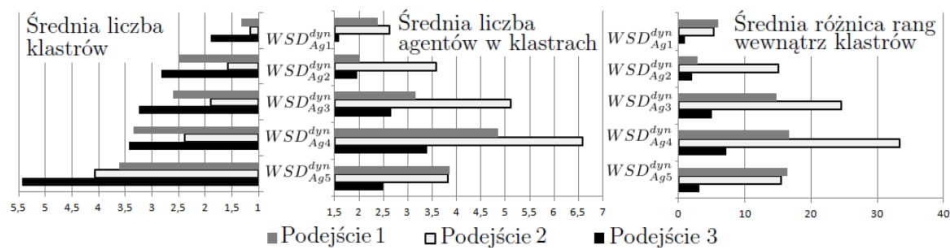
W celu wyznaczenia klastrów względem klasyfikacji dokonanej przez agentów należy uwzględnić wszystkie wartości atrybutu decyzyjnego, czyli zbiór A . W tym celu może zostać wykorzystana jedna z dwóch funkcji – funkcja odległości (wzór 3) lub funkcja konfliktu (wzór 4). Główna różnica pomiędzy tymi funkcjami dotyczy przypisywania wartości agentom neutralnym. Wytlumaczmy to na przykładzie. Niech dany będzie atrybut $a \in A$ oraz trzech agentów $ag_1, ag_2, ag_3 \in U$. Załóżmy, że $a(ag_1) = 1, a(ag_2) = 0, a(ag_3) = -1$. Wtedy $\rho_{\{a\}}^*(ag_1, ag_2) = 0,5$ oraz $\rho_{\{a\}}^*(ag_1, ag_3) = 1$. Zatem odległość agentów, którzy są neutralni jest mniejsza niż odległość agentów, którzy są w konflikcie. Natomiast w przypadku funkcji konfliktu mamy $\rho_{\{a\}}(ag_1, ag_2) = 1$ oraz $\rho_{\{a\}}(ag_1, ag_3) = 1$. Zatem odległość agentów neutralnych i agentów w konflikcie jest taka sama.

W pierwszej metodzie definiowania systemu informacyjnego nie występuje wartość 0 , zatem nie ma agentów neutralnych. To oznacza, że obie funkcje przyjmują te same wartości.

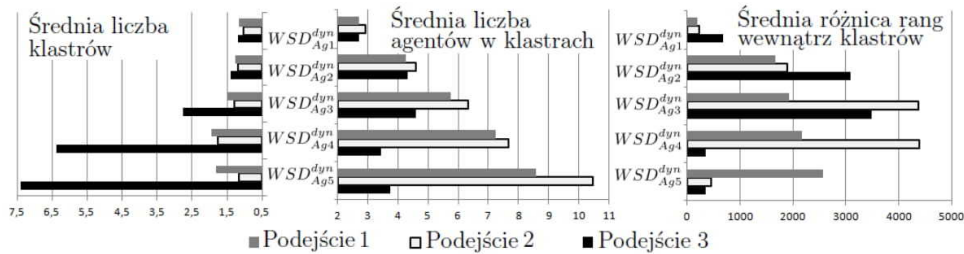
W drugiej metodzie definiowania systemu informacyjnego uwzględniono już agentów neutralnych, więc funkcje mogą przyjmować różne wartości. Jak wyjaśniono powyżej, funkcja odległości będzie mniej restrykcyjna dla agentów neutralnych – przyporządkowuje agentom neutralnym mniejszą wartość – niż funkcja konfliktu. Na podstawie wartości jednej z funkcji definiujemy klastery jako podzbiór agentów zasobów taki, że dla każdej pary agentów wartość funkcji jest mniejsza od 0,5. Ostatecznie rozważałam trzy różne podejścia do generowania klastrów:

- *Podejście 1* – wykorzystywana jest pierwsza metoda definiowania systemu informacyjnego oraz funkcja odległości (która w tym przypadku jest tożsama z funkcją konfliktu),
- *Podejście 2* – wykorzystywana jest druga metoda definiowania systemu informacyjnego oraz funkcja odległości,
- *Podejście 3* – wykorzystywana jest druga metoda definiowania systemu informacyjnego oraz funkcja konfliktu.

W pracy [B4], badałam strukturę klastrów tworzonych przy wykorzystaniu trzech powyższych podejść. W celu porównania jakości klastrów zastosowałam miarę opartą na średniej różnicy wartości rang przypisanych decyzjom przez agentów z jednego klastra. Na rysunkach 11 oraz 12 przedstawiono średnie liczby klastrów, liczby agentów w klastrach oraz średnią różnicę rang wewnątrz klastrów otrzymane dla rozproszonych danych utworzonych na podstawie zbiorów Lymphography i Primary Tumor z repozytorium UCI. Jak można zauważyć, podejście 3 zazwyczaj generuje większą liczbę klastrów, które są mniej liczne. Ponadto to podejście zazwyczaj generuje klastry z mniejszą wartością różnicy rang wewnątrz klastra. W przypadku danych z dużą liczbą klas decyzyjnych (Primary Tumor), podejścia 1 i 2 tracą zdolność do identyfikowania relacji między agentami, ponieważ podejścia te generują praktycznie jeden klastery składający się z wszystkich agentów. Ostatecznie stwierdziłam, że podejście 3 najlepiej identyfikuje relacje zachodzące między klasyfikatorami.



Rysunek 11: Porównanie struktury klastrów tworzonych dla trzech podejść opartych na modelu Pawlaka (Lymphography)



Rysunek 12: Porównanie struktury klastrow tworzonych dla trzech podejść opartych na modelu Pawlaka (Primary Tumor)

5 Omówienie pozostałych osiągnięć naukowo – badawczych

W większości moich publikacji poruszana jest tematyka głównego osiągnięcia badawczego. Moje pozostałe osiągnięcia są związane z moimi głównymi zainteresowaniami naukowymi i również dotyczą wykorzystania wiedzy rozproszonej.

Poniżej zostaną wymienione prace, które wykazują w zakresie dodatkowych osiągnięć naukowych. Wszystkie wskazane artykuły zostały zrealizowane i opublikowane po uzyskaniu przeze mnie tytułu doktora.

W następujących pracach

- [D1] Wakulicz–Deja A., Przybyła–Kasperek M., Podejmowanie decyzji globalnych z zastosowaniem reguł decyzyjnych generowanych na podstawie tablic zagregowanych, Systemy Wspomagania Decyzji, Wakulicz–Deja A. (red.), 109–130, ISBN 978-83-62462-05-6, Instytut Informatyki Uniwersytetu Śląskiego, Katowice, (2012)
- [D2] Przybyła–Kasperek M., Ocena efektywności wnioskowania wieloagentowego systemu decyzyjnego działającego w oparciu o reguły decyzyjne, wygenerowane na podstawie lokalnych baz wiedzy, ZN Pol. Śl. Studia Informatica Vol. 33, No. 2A (105), 419–437, (2012)

analizowałam wykorzystanie reguł decyzyjnych w systemie rozproszonym ze strukturą statyczną. Rozważałam dwa możliwe podejścia. W pierwszym podejściu reguły decyzyjne generowane były na podstawie tablic lokalnych, w drugim podejściu reguły generowane były na podstawie zagregowanych tablic decyzyjnych. Stwierdziłam, że lepsze wyniki gwarantuje podejście, w którym reguły generowane są na podstawie tablic lokalnych.

W publikacjach

- [D3] Przybyła–Kasperek M., Wieloagentowy system decyzyjny z dynamicznie generowanymi rozłącznymi klastrami, ZN Pol. Śl. Studia Informatica Vol. 34, No. 2A (111), 275–294, (2013)
- [D4] Wakulicz–Deja A., Nowak–Brzezińska A., Przybyła–Kasperek M., Complex decision systems and conflicts analysis problem, Fundamenta Informaticae 127 (1–4), 341–356, (2013)

- [D5] Przybyła–Kasperek M., Wieloagentowy system decyzyjny z dynamicznie generowanymi klastrami, Systemy Wspomagania Decyzji, Wakulicz–Deja A. (red.), Instytut Informatyki Uniwersytetu Śląskiego, Katowice, 65–82, ISBN: 978-83-62462-08-7, (2013)
- [D6] Przybyła–Kasperek M., Podejmowanie decyzji na podstawie wiedzy rozproszonej – podejście dynamiczne do tworzenia struktury systemu, rozdział w monografii jubileuszowej dedykowanej prof. dr hab. inż. Alicji Wakulicz–Deji, Uniwersytet Śląski, 335–348, ISBN: 978-83-62462-09-4, (2013)
- [D7] Przybyła–Kasperek M., Wakulicz–Deja A., Global decisions taking on the basis of dispersed medical data, Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Lecture Notes in Computer Science Volume 8170, 355–365, (2013)
- [D8] Przybyła–Kasperek M., Rozproszony system wspomagania decyzji z zastosowaniem negocjacji, Systemy Inteligencji Obliczeniowej, Instytut Informatyki Uniwersytetu Śląskiego, 155–166, Sosnowiec, (2014), ISBN: 978-83-62462-11-7

rozważałam rozproszone systemy decyzyjne z dynamiczną strukturą. Prace te stanowią kontynuację prac [A2–A4] oraz [B2] wchodzących w skład mojego głównego osiągnięcia naukowego. Przedstawiłam w nich analizy uzupełniające oraz, jak w pracy [D7], wyniki eksperymentów na dodatkowych zbiorach danych związanych z wykorzystaniem wiedzy rozproszonej w medycynie.

W artykule

- [D9] Przybyła–Kasperek M., Wakulicz–Deja A., Methods of calculating the strength of coalition in a dispersed decision support system with the stage of negotiations – a study of medical data, Concurrency, Specification and Programming CS&P 2014, Informatik–Bericht Nr. 245, 208–219, Berlin (2014)

dokonałam wstępnej analizy wykorzystania metody analizy konfliktów z wagami klastrów w systemie ze strukturą dynamiczną i etapem negocjacji. Stwierdziłam, że zastosowanie wag wpływa na poprawę jakości klasyfikacji systemu. Szczegółowe wyniki przedstawiłam w pracy [A5] wchodzącej w skład mojego głównego osiągnięcia naukowego.

W poniższym zbiorze publikacji

- [D10] Przybyła–Kasperek M., Selected methods of combining classifiers, when predictions are stored in probability vectors, in a dispersed decision–making system, 24th International Workshop Concurrency, Specification and Programming, CS&P 2015, University of Rzeszow, Suraj Z., Czaja L. (Eds.), 211–222, (2015), ISBN: 978-83-7996-181-8
- [D11] Przybyła–Kasperek M., The Borda Count, the Intersection and the Highest Rank Method in a Dispersed Decision–Making System. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing – 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20–23, 2015, Proceedings, Lecture Notes in Computer Science, 298–309, (2015)

- [D12] Przybyła–Kasperek M., Two Methods of Combining Classifiers, Which are Based on Decision Templates and Theory of Evidence, in a Dispersed Decision–Making System, Beyond Databases, Architectures and Structures. *Advanced Technologies for Data Mining and Knowledge Discovery – 12th International Conference, BDAS 2016*, Ustroń, Poland, May 31 – June 3, 2016, 109–119 (2016), http://dx.doi.org/10.1007/978-3-319-34099-9_7
- [D13] Przybyła–Kasperek M., Comparison of Four Methods of Combining Classifiers on the Basis of Dispersed Medical Data, *Intelligent Decision Technologies 2016*, Volume 57 of the series Smart Innovation, Systems and Technologies, Proceedings of the 8th KES International Conference on Intelligent Decision Technologies (KES–IDT 2016), Springer, 3–13 (2016)
- [D14] Przybyła–Kasperek M., Nowak–Brzezińska A., Intersection Method, Union Method, Product Rule and Weighted Average Method in a Dispersed Decision–Making System – a Comparative Study on Medical Data, *Computational Collective Intelligence – 8th International Conference, ICCCI 2016*, Halkidiki, Greece, September 28–30, 2016, Proceedings, Part II, Lecture Notes in Artificial Intelligence, 451–461 (2016)
- [D15] Przybyła–Kasperek M., Nowak–Brzezińska A., Simiński R., Decision Fusion Methods in a Dispersed Decision System – A Comparison on Medical Data. In: Nguyen N., Papadopoulos G., Jędrzejowicz P., Trawiński B., Vossen G. (eds) *Computational Collective Intelligence. ICCCI 2017. Lecture Notes in Computer Science*, vol 10449. Springer, Cham, 139–149, (2017)

analizowałam zastosowanie różnych metod fuzji w systemie rozproszonym z etapem negocjacji. Wyniki zaprezentowane w tych pracach stanowią ujęcie fragmentaryczne, które zostało uzupełnione i kompleksowo zaprezentowane w pracach [A7, A8] wchodzących w skład mojego głównego osiągnięcia naukowego. W powyższych pracach zaprezentowałam wyniki eksperymentów na dodatkowych, medycznych zbiorach danych.

Tematyka prac

- [D16] Wakulicz–Deja A., Przybyła–Kasperek M., Pawlak’s Conflict Model: Directions of Development, *Annals of Computer Science and Information Systems*, Volume 8, Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, 191–197, (2016), <http://dx.doi.org/10.15439/2016F003>
- [D17] Przybyła–Kasperek M., Directions of use of the Pawlak’s approach to conflict analysis, *Thriving Rough Sets: 10th Anniversary - Honoring Professor Zdzisław Pawlak’s Life and Legacy and 35 Years of Rough Sets*, Wang, G., Skowron, A., Yao, Y., Ślęzak, D., Polkowski, L. (Eds.), 329–351, Springer, ISBN 978-3-319-54966-8, 329–351, (2017)

dotyczy modelu konfliktu zaproponowanego przez Profesora Pawlaka. W publikacjach tych przedstawiłam rozwinięcia modelu Pawlaka znane z literatury oraz kierunki zastosowań tego modelu. W pracy [D17] przedstawiłam możliwość wykorzystania modelu Pawlaka do zagadnienia zespołu klasyfikatorów. Zastosowanie to zaprezentowałam na podstawie

przykładu. Kontynuacją tych artykułów jest praca [B4], która wchodzi w skład mojego głównego osiągnięcia naukowego.

Dodatkowo wykazuję publikacje, które są mniej związane z głównymi osiągnięciami, a dotyczą:

- wykorzystania zagadnień z teorii gier w systemie z wiedzą rozproszoną (rozdział 5.1),
- zastosowania metody selekcji atrybutów opartej na zbiorach przybliżonych w rozproszonym systemie ze strukturą dynamiczną (rozdział 5.2).

[D18] Przybyła–Kasperek M., Application of the Shapley–Shubik power index in the process of decision making on the basis of dispersed medical data, S. Kozielski et al. (Eds.): BDAS 2015, CCIS 521, 277–287, Springer International Publishing Switzerland (2015)

[D19] Przybyła–Kasperek M., Attribute Reduction in a Dispersed Decision–Making System with Negotiations, Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation – 13th International Conference, BDAS 2017, Ustroń, Poland, May 30 – June 2, 2017, Proceedings, 79–88, Springer (2017)

[D20] Przybyła–Kasperek M., Dispersed System with Dynamically Generated Non–disjoint Clusters – Application of Attribute Selection, Intelligent Decision Technologies 2017, the series Smart Innovation, Systems and Technologies, Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017), Springer, (2018)

[D21] Przybyła–Kasperek M., Feature selection based on the rough set theory and dispersed system with dynamically generated disjoint clusters, IEEE International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2017, Gdynia, Poland, July 3–5, 223–228 (2017)

[D22] Przybyła–Kasperek M., Attribute selection in a dispersed decision-making system, Advances in Feature Selection for Data and Pattern Recognition Springer, Stańczyk, U., Zielosko, B., Jain, L.C., (Eds.), Intelligent Systems Reference Library, 138, 133–162, Springer, (2018)

5.1 Wykorzystanie zagadnienia z teorii gier w systemie z wiedzą rozproszoną

Tak jak wspomniano wcześniej, system rozproszony z etapem negocjacji (rozdział 4.4.5) ma zdolność do identyfikowania skomplikowanych relacji zachodzących pomiędzy agentami. Jednak bez zastosowania wag klastrów ta zaleta systemu nie przekłada się na jakość klasyfikacji. W rozdziale 4.4.6 opisano jedno z podejść uwzględnienia struktury klastrów podczas podejmowania ostatecznej decyzji. W mojej pracy zastosowałam również całkiem odmienne rozwiązanie, w którym zagadnienia związane z wiedzą rozproszoną zostały połączone z pojęciami z teorii gier. W pracy [D18] zaproponowałam zastosowanie indeksu Shapley’a–Shubika w systemie z wiedzą rozproszoną. Indeks Shapley’a–Shubika

jest powszechnym narzędziem do wyliczania siły agentów w grach prostych. Zastosowanie tej miary w systemie rozproszonym było naturalnym i oczywistym rozwiązaniem. Razem z indeksem siły zaproponowałam wykorzystanie nowej metody wyznaczania decyzji lokalnych o zasięgu jednego klastra. Metoda ta polega na wykonaniu przekształceń liniowych na wektorach wartości wygenerowanych na podstawie tablic lokalnych. Złożoność obliczeniowa algorytmu realizującego tę technikę jest znacznie niższa niż złożoność algorytmu agregacji tablic decyzyjnych. Zatem w pracy [D18] zaproponowałam rozwiązanie realizujące dwa cele – poprawa jakości klasyfikacji systemu oraz zmniejszenie jego złożoności obliczeniowej.

5.2 Zastosowanie metody selekcji atrybutów opartej na zbiorach przybliżonych w rozproszonym systemie ze strukturą dynamiczną

Problem występowania atrybutów zbędnych, nadwymiarowych w zbiorze danych jest szeroko omawiany w literaturze [3, 5]. Zagadnienie to ma negatywny wpływ na proces klasyfikacji pod dwoma względami. Po pierwsze wpływa na zwiększenie czasu działania algorytmu. Ponadto często nadwymiarowe atrybuty zaburzają opis rozważanego zagadnienia, co powoduje obniżenie jakości klasyfikacji.

Celem tego etapu mojej pracy było zastosowanie metody selekcji atrybutów opartej na zbiorach przybliżonych (opisanej w rozdziale 4.4.1) w systemie z dynamiczną strukturą. W pracy [D21] zastosowałam metodę selekcji w systemie rozproszonym z rozłącznymi klastrami (opisany w rozdziale 4.4.3). W pracy [D20], wykorzystałam redukt w systemie rozproszonym z nierozłącznymi klastrami (opisany w rozdziale 4.4.4). W pracy [D19], omówiłam wyniki uzyskane przy zastosowaniu selekcji atrybutów w systemie z etapem negocjacji (opisany w rozdziale 4.4.5). W pracy [D22] dokonałam kompleksowego porównania wyników otrzymanych przy zastosowaniu redukcji atrybutów we wszystkich typach systemów. Na podstawie otrzymanych rezultatów wyciągnęłam pewne wnioski ogólne. W większości przypadków zastosowanie selekcji atrybutów opartej na zbiorach przybliżonych generuje lepszą jakość klasyfikacji. Największą poprawę otrzymano dla systemu z etapem negocjacji oraz systemu z nierozłącznymi klastrami.

Literatura

- [1] Breiman, L.: *Bagging predictors*, Machine Learning, 24 (2), 123–140 (1996)
- [2] Cabrerizo, F. J., Herrera-Viedma, E., Pedrycz, W.: *A method based on PSO and granular computing of linguistic information to solve group decision making problems defined in heterogeneous contexts*, European Journal of Operational Research, 230 (3), 624–633 (2013)
- [3] Cano, A., Ventura, S., Cios, K.: *Multi-objective genetic programming for feature extraction and data visualization*, Soft Computing, 21(8), 2069–2089 (2017)
- [4] Caragea, D.: *Learning Classifiers from Distributed, Semantically Heterogeneous, Autonomous Data Sources*, PhD thesis, Iowa State University, (2004)
- [5] Cichocki, A., Mandic, D., Phan, A.H., Caiafa, C., Zhou, G., Zhao, Q., De Lathauwer, L.: *Tensor decompositions for signal processing applications from two-way to multiway component analysis*, CoRR, abs/1403.4462, (2014)
- [6] Deja, R.: *Conflict analysis*, Proceedings of the fourth international workshop on rough sets, fuzzy sets and machine discovery, The University of Tokyo, 6-8 November, 118–124 (1996)

- [7] Deja, R.: *Conflict analysis*, Proceedings of the 7th European Congress on Intelligent Techniques & Soft Computing, Aachen, Germany, September 13-16, (1999)
- [8] Deja, R.: *Conflict analysis*, Rough Sets; New Developments. In: Polkowski L. (eds.), Studies in Fuzzy Business and Soft Computer Science, Physica-Verlag, (2000)
- [9] Deja, R.: *Zastosowanie teorii zbiorów przybliżonych w analizie konfliktów*, Instytut Podstaw Informatyki Polskiej Akademii Nauk, Rozprawa doktorska, promotor: A. Skowron (2000)
- [10] Deja, R.: *Conflict analysis*, Int. J. Intell. Syst., 17, 2, 235–253 (2002)
- [11] Delimata, P., Suraj, Z.: *Feature selection algorithm for multiple classifier systems: A hybrid approach*, Fundamenta Informaticae 85 (1-4), Amsterdam: IOS Press, 97–110 (2008)
- [12] Ester, M., Kriegel, H., Sander, J., Xu, X.: *A density-based algorithm for discovering clusters in large spatial databases with noise*, In Proceedings of 2nd international conference on knowledge discovery and data mining, Portland, 226–231 (1996)
- [13] Franklin, S., Graesser, A.: *Is it an Agent, or just a Program - A Taxonomy for Autonomous Agents*, Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, Berlin: Springer-Verlag, 21–35 (1996)
- [14] Gatnar, E.: *Podejscie wielomodelowe w zagadnieniach dyskryminacji i regresji*, Warsaw: PWN, (2008)
- [15] Gagolewski, M.: *Data fusion, Theory, Methods, and Applications*, Institute of Computer Science Polish Academy of Sciences, (2015)
- [16] Greco, S., Matarazzo, B., Slowiński, R.: *Rough sets theory for multicriteria decision analysis*, European Journal of Operational Research, 129 (1), 1–47 (2001)
- [17] Ho, T.K.: *The random subspace method for constructing decision forests*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (8), 832–844 (1998)
- [18] Hu, J., Wang, K., Yu, H.: *Attribute Reduction on Distributed Incomplete Decision Information System*, In: Polkowski, L., Yao, Y., Artiemjew, P., Ciucci, D., Liu, D., Ślęzak, D., Zielosko, B. (eds) Rough Sets: International Joint Conference, IJCRS 2017, Olsztyn, Poland, July 3–7, 2017, Proceedings, Part I, Springer International Publishing, 289–305 (2017)
- [19] Jiang, W., Zhang, X., Cohen, A., Ras, Z.: *Multiple classifiers for different features intimbre estimation*, Adv. Intell. Inform. Syst., 335–356 (2010)
- [20] Jankowski, A., Skowron, A., Swiniarski, R. W.: *Interactive complex granules*. Fundamenta Informaticae, 133(2–3), 181–196 (2014)
- [21] Janusz, A., Ślęzak, D.: *Rough Set Methods for Attribute Clustering and Selection*, Applied Artificial Intelligence, 28(3), 220–242 (2014)
- [22] Kanter, J. M., Veeramachaneni, K.: *Deep feature synthesis: Towards automating data science endeavors*. In Data Science and Advanced Analytics (DSAA), IEEE International Conference, 1–10 (2015)
- [23] Koronacki, J., Ćwik, J.: *Statystyczne systemy uczące się*, EXIT, Warszawa, (2008)
- [24] Kuncheva, L.: *Combining pattern classifiers methods and algorithms*, John Wiley & Sons (2004)
- [25] Maes, P.: *Artificial Life Meets Entertainment: Life like Autonomous Agents*, Communications of the ACM (38), 108–114 (1995)
- [26] Nguyen, S.H., Bazan, J.G., Skowron, A., Nguyen, H.S.: *Layered learning for concept synthesis*. In: Transactions on Rough Sets I. LNCS, vol. 3100, Springer, Berlin, 187–208 (2004)
- [27] Panov, P., Dżeroski, S.: *Combining Bagging and Random Subspaces to Create Better Ensembles*, In: R. Berthold M., Shawe-Taylor J., Lavra N. (eds) Advances in Intelligent Data Analysis VII. IDA 2007. LNCS, vol 4723. Springer. (2007)
- [28] Pawlak, Z.: *On conflicts*. Int. J. of Man-Machine Studies 21, 127–134 (1984)
- [29] Pawlak, Z.: *About conflicts* (in Polish), Polish Scientific Publishers, Warsaw, 1–72 (1987)
- [30] Pawlak, Z.: *Anatomy of conflict*. Bulletin of the European Association for Theoretical Computer Science, 50, 234–247 (1993)

- [31] Pawlak, Z.: On some issues connected with conflict analysis. Institute of Computer Science Reports, 37/93, Warsaw University of Technology 1993.
- [32] Pawlak, Z.: An Inquiry Anatomy of Conflicts. *Journal of Information Sciences* 109, 65–78 (1998)
- [33] Pawlak, Z.: Some remarks on conflict analysis. *European Journal of Operational Research* 166, 649–654 (2005)
- [34] Pawlak, Z.: Conflicts and Negotiations. In: Wang, G.-Y., Peters, J.F., Skowron, A., Yao, Y. (eds.) RSKT 2006. LNCS (LNAI), vol. 4062, Springer, Heidelberg 12–27 (2006)
- [35] Pawlak, Z., Skowron, A.: Rough Sets and Conflict Analysis. *E-Service Intelligence: Methodologies, Technologies and Applications*, 35–74 (2007)
- [36] Polikar, R.: *Ensemble based systems in decision making*, *IEEE Circuits and Systems Magazine*, 6, 21–45 (2006)
- [37] Polikar, R., Parikh, D., Mandayam, S.: *Multiple classifier systems for multisensor data fusion*, In: Proceedings of the 2006 IEEE Sensors Applications Symposium, 180–184 (2006)
- [38] Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59 (2013)
- [39] Przybyła-Kasperek, M.: *Wieloagentowy system decyzyjny - analiza konfliktów*, Rozprawa doktorska, Uniwersytet Śląski
- [40] Ramanna, S., Peters, J.F., Skowron, A.: *Generalized Conflict and Resolution Model with Approximation Spaces*. Rough Sets and Current Trends in Computing, 5th International Conference, RSCTC 2006, Kobe, Japan, November 6-8, 2006, Proceedings, 274–283 (2006)
- [41] Ramanna, S., Peters, J.F., Skowron, A.: *Approaches to Conflict Dynamics Based on Rough Sets*. *Fundam. Inform.* 75(1-4) 453–468 (2007)
- [42] Ramanna, S., Skowron, A.: Requirements *Interaction an Conflicts A Rough Set Approach*. Proceedings of the IEEE Symposium on Foundations of Computational Intelligence, FOCI 2007, part of the IEEE Symposium Series on Computational Intelligence 2007, Honolulu, Hawaii, USA, 1-5 April 2007, 308–313 (2007)
- [43] Ramanna, S., Skowron, A., Peters, J.F.: *Approximation Space-Based Socio-Technical Conflict Model*. Rough Sets and Knowledge Technology, Second International Conference, RSKT 2007, Toronto, Canada, May 14-16, 2007, Proceedings 476–483 (2007)
- [44] Russell, S., Norvig, P.: *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs, New Jersey, (1995)
- [45] Schneeweiss, C.: *Distributed decision making*, Berlin: Springer (2003)
- [46] Schneeweiss, C.: *Distributed decision making-a unified approach*, *European Journal of Operational Research*, 150 (2), 237–252, (2003)
- [47] Skowron, A., Deja, R.: *On Some Conflict Models and Conflict Resolutions*, *Romanian Journal of Information Science and Technology* 3(1-2), 69–82 (2002)
- [48] Skowron, A., Jankowski, A., Dutta, S.: *Interactive granular computing*, *Granul. Comput.* 1 (2), 95–113 (2016)
- [49] Skowron, A., Ramanna, S., Peters, J.F.: *Conflict Analysis and Information Systems: A Rough Set Approach*. Rough Sets and Knowledge Technology, First International Conference, RSKT 2006, Chongqing, China, July 24-26, 2006, Proceedings, 233–240 (2006)
- [50] Skowron, A., Wang, H., Wojna, A., Bazan, J.: *Multimodal Classification: Case Studies*, T. Rough Sets: 224–239 (2006)
- [51] Sosnowski, L.: Framework of compound object comparators, *Intelligent Decision Technologies*, 9(4), 343–363 (2015)
- [52] Suraj, Z., Gayar, Neamat El, Delimata, P.: *A Rough Set Approach to Multiple Classifier Systems*, *Fundamenta Informaticae* 72 (1-3), IOS Press, Amsterdam, 393–406 (2006)
- [53] Szczuka, M.S., Ślęzak, D.: Feedforward neural networks for compound signals, *Theor. Comput. Sci.*, 412(42), 5960–5973 (2011)

- [54] Ślęzak, D.: On Generalized Decision Functions: Reducts, Networks and Ensembles, Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings, 13-23 (2015)
- [55] Wróblewski, J.: Ensembles of Classifiers Based on Approximate Reducts, *Fundam. Inform.*, 47 (3-4), 351-360 (2001)
- [56] Xu, L., Krzyzak, A., Suen, C. Y.: *Methods of combining multiple classifiers and their application to handwriting recognition*, *IEEE Transactions on Systems, Man, and Cybernetics*, 22:418-435 (1992)
- [57] Zadeh, L. A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems* 90 111-127 (1997)
- [58] <http://logic.mimuw.edu.pl/rses/>

Małgorzata
Przybyła-Kasperek