

POLITECHNIKA ŚLĄSKA
WYDZIAŁ AUTOMATYKI, ELEKTRONIKI I INFORMATYKI

STRESZCZENIE ROZPRAWY DOKTORSKIEJ

Indukcja reguł akcji na podstawie metody sekwencyjnego pokrywania

Paweł Matyszok

Promotor: dr hab. Marek Sikora, prof. Pol. Śl.

Promotor pomocniczy: dr inż. Łukasz Wróbel

Gliwice 2022

1 Wprowadzenie

Praca poświęcona jest zastosowaniu metod indukcji reguł do odkrywania użytecznej wiedzy w postaci reguł akcji [10]. Wiedzę określamy użyteczną wtedy, gdy jesteśmy w stanie na jej podstawie osiągnąć pewną korzyść. W kontekście reguł akcji najczęściej przez taką korzyść rozumieć należy zmianę klasyfikacji danego obiektu z klasy niepożądaney (zwanej źródłową) do klasy pożądaney (docelowej). Reguły akcji wskazują, jak należy zmienić wartości atrybutów charakteryzujących obiekt, aby można było oczekiwać pożądaney zmiany jego klasyfikacji. Reguły akcji są szczególnym przypadkiem reguł logicznych.

Reguły logiczne formalizują odkrytą wiedzę w postaci wyrażeń warunkowych opisujących lokalne zależności między wartościami atrybutów w pewnym zbiorze danych. Wśród wszystkich atrybutów (cech) wyróżnia się jeden atrybut zwany decyzyjnym i oznaczający tę cechę opisywanych obiektów, która jest interesująca ze względu na cel prowadzonego modelowania. Reguły znajdują swoje zastosowanie w zadaniach klasyfikacji (przypisywania obiektów do właściwej dla nich klasy), analizy asocjacji (odkrywanie zestawów cech często występujących razem), regresji (przewidywania wartości liczbowej atrybutu decyzyjnego) i analizie przeżycia (przewidywanie czasu, jaki upłynie do zajścia pewnego zdarzenia) [15].

Odkrywanie użytecznej wiedzy (ang. *Actionable Knowledge*) poprzez badanie zależności prowadzących do zmiany wartości atrybutu decyzyjnego jest stosunkowo nowym obszarem zastosowań reguł. Narzędziem umożliwiającym opis zbioru danych pod tym kątem są reguły akcji. Reguły akcji zostały po raz pierwszy opisane przez Rasia i Wierzchowską w roku 2000 [10]. Reguła akcji w swej budowie podobna jest do reguły decyzyjnej, przy czym zamiast wiązać atrybut decyzyjny z jedną wartością, wskazuje na wartość źródłową i docelową. Podobna zmiana zachodzi w części warunkowej reguły. Zmienia się także znaczenie reguły. O ile regułę decyzyjną należy rozumieć jako stwierdzenie „Jeżeli spełnione są pewne warunki dotyczące cech obiektu, to zachodzi podana decyzja”, to reguła akcji odpowiada na pytanie: „Jak muszą zmienić się cechy obiektu, by zaszła zmiana w decyzji?”.

Reguły akcji mogą stanowić narzędzie opisujące zależności między atrybutami warunkowymi a atrybutem decyzyjnym, ale mogą one też pełnić rolę bardzo praktycznej rekomendacji, ponieważ wdrażając pewne operacje, jesteśmy w stanie zmienić stan obiektu zgodnie z sugestiami podanymi w regule akcji i oczekiwać odpowiedniej zmiany wartości atrybutu decyzyjnego. Proces odkrywania fizycznych działań prowa-

dzących do pożądaných zmian w wartościach atrybutów nazywa się odkrywaniem meta-akcji [9]. Przez *akcję* należy więc rozumieć żądanie zmiany wartości pewnego atrybutu, natomiast termin *meta-akcja* określa działanie, dające w efekcie taką zmianę. Przykładowo, akcją określimy stwierdzenie „obniżyć temperaturę pacjenta poniżej 39°C”. Komplementarną meta-akcją mogłoby być stwierdzenie „umieścić pacjenta w wannie z lodem” lub „podać pacjentowi leki obniżające gorączkę”. W omawianej pracy skupiono się tylko na regułach akcji.

Bezpośrednie wdrażanie metod odkrywania użytecznej wiedzy do podejmowania pewnych decyzji wiąże się z ryzykiem popełnienia przez nie błędu. Minimalizacja takiego błędu jest bardzo istotna. Z tego względu pożądanym jest, by wyniki odkrywania użytecznej wiedzy były prezentowane w sposób klarowny i łatwy do zrozumienia przez użytkownika systemu jak i eksperta domenowego, który jest w stanie określić, czy wyniki działania systemu są wartościowe. Przeglądając się opublikowanym w literaturze wdrożeniom metod odkrywania użytecznej wiedzy w zagadnieniach medycznych [14], optymalizacji procesów obsługi klienta [13] czy zwiększania przychodów przedsiębiorstw [7] potwierdza się przesłankę, że wyniki odkrywania użytecznej wiedzy są raczej poddawane dalszej analizie przez człowieka, a nie bezpośrednio wdrażane w życie. W takim zastosowaniu reprezentacja wiedzy w postaci reguł jest szczególnie przydatna.

1.1 Motywacja i cel pracy

W literaturze podano już wiele algorytmów odkrywania reguł akcji, wykorzystujących stosowane już wcześniej w indukcji reguł decyzyjnych podejścia, takie jak analiza tablicy decyzyjnej [10], algorytm A’piori [6] czy konwersja drzew decyzyjnych do postaci reguł [16].

Do tej pory jednak nie zastosowano paradygmatu sekwencyjnego pokrywania, uważanego za szczególnie efektywny w zadaniu indukcji reguł decyzyjnych [4, 5]. Metoda ta pozwala nadzorować proces indukcji reguły miarami oceny jakości zadawanymi przez użytkownika, co pozwala na łatwe dostosowywanie algorytmu do specyficznych potrzeb wynikających z rozwiązywanego zadania. Dodatkowo, algorytmy oparte o metodę sekwencyjnego pokrywania pozwalają na indukcję reguł w zbiorach danych zawierających zarówno atrybuty kategorię jak i numeryczne, nie wymagając uprzedniej dyskretyzacji. Możliwa jest także obsługa wartości brakujących w analizowanych zbiorach.

Idąc dalej, dzięki modyfikacjom pokryciowego algorytmu indukcji reguł decyzyjnych, można uzyskać metodę odkrywania reguł regresyjnych (czyli takich, w których atrybut decyzyjny jest ciągły) oraz przeżycia (w których zmienna zależna określa czas, jaki upłynął do pewnego zdarzenia).

Rzadko poruszonym zagadnieniem jest również proces wyboru reguły akcji, która miałaby być zastosowana na pewnym przykładzie. Przez zastosowanie reguły akcji będziemy rozumieć wdrożenie działań mających na celu zrealizowanie zmian wartości atrybutów opisanych w jej przesłance. Problem ten jest istotny w obliczu zbiorów reguł akcji, w których jeden przykład może być pokryty przez więcej niż jedną regułę.

Większość opublikowanych dotychczas prac na temat indukcji reguł akcji wprowadza pewną metodę odkrywania reguł akcji i przedstawia studium przypadku jej wdrożenia. Niewiele jest prac zestawiających ze sobą wyniki działania kilku metod na szerokim zestawie zbiorów danych, wreszcie brak jest powszechnie uznanej metody oceniania skuteczności reguł akcji. Oczywiście jedynym niezawodnym sposobem oceny jakości reguł akcji jest ich zastosowanie na pewnej grupie obiektów, a następnie śledzenie ich dalszych losów w celu określenia, czy zaszedł oczekiwany efekt zmiany cechy uznanej za decyzyjną, czy też nie. Działanie takie może być jednak kosztowne, trudne, czy wręcz niemożliwe w realizacji. Stąd przydatnym byłby sposób na oszacowanie skuteczności zbioru reguł akcji na etapie modelowania. Skuteczność zbioru reguł akcji będzie definiowana wskaźnikami pozwalającymi określić, jak często zastosowanie reguł akcji pochodzących z danego zbioru prowadzi do uzyskania oczekiwanego efektu zmiany atrybutu decyzyjnego.

Niewykorzystany jeszcze w zadaniu odkrywania reguł akcji potencjał paradygmatu sekwencyjnego pokrywania stanowi motywację tej pracy. Popularność tego podejścia na polu indukcji reguł decyzyjnych rodzi nadzieję na uzyskanie nowych algorytmów indukcji reguł akcji, charakteryzującymi się możliwością indukcji reguł wprost z danych, bez preprocessingu, obsługą wartości ciągłych, dyskretnych i brakujących oraz możliwością sterowania procesem indukcji przez zadanie funkcji oceny jakości wytwarzanych reguł.

Celem pracy jest próba wypełnienia luki w obszarze indukcji reguł akcji i opracowanie nowych metod eksploracji danych prowadzących do uzyskania reguł akcji ze zbiorów opisujących dane klasyfikacyjne i regresyjne. Oprócz tego, praca skupia się na rozwiązaniu problemu dobrania najlepszej reguły akcji dla danego przykładu oraz indukcji specjalizowanych reguł akcji, zwanych rekomendacjami. Rekomendacje tworzone są tak, by podać możliwie najlepsze akcje dla danego przykładu. Skutecz-

ność rekomendacji oraz zbiorów reguł akcji jest oceniana przez zaproponowany w pracy schemat eksperymentu, który może znaleźć zastosowanie w ocenie różnych metod uzyskiwania reguł akcji.

1.2 Teza pracy

Zastosowanie w odkrywaniu reguł akcji paradygmatu sekwencyjnego pokrywania i odpowiednio dobranych kryteriów sterowania algorytmem indukcji reguł pozwala na uzyskanie modeli o dobrych zdolnościach prognostycznych i objaśniających.

1.2.1 Tezy pomocnicze

- Metodyka wymieniona w głównej tezie pracy może być zastosowana do indukcji reguł akcji dla danych klasyfikacyjnych i opisujących problemy regresyjne.
- W zależności od sposobu prowadzenia indukcji (rozważane podejścia to indukcja z punktu widzenia klasy źródłowej i indukcja z punktu widzenia klasy docelowej) uzyskujemy różne zbiory reguł, o różnych zdolnościach predykcyjnych i opisowych.
- Agregacja wiedzy zawartej w wyznaczonych regułach akcji pozwala na opracowanie algorytmu rekomendacyjnego, zdolnego do wskazania dla zadanego obiektu w jaki sposób należy zmienić wartości jego atrybutów, aby osiągnąć zamierzoną wartość atrybutu decyzyjnego.

1.3 Układ pracy

Praca składa się z 6 rozdziałów. Rozdział 1 i 6 to odpowiednio wprowadzenie i podsumowanie pracy. Rozdział 3 jest wprowadzeniem do tematyki indukcji reguł akcji, począwszy od indukcji reguł decyzyjnych wraz z opisem budowy klasyfikatora regułowego i kilku popularnych podejść do indukcji reguł decyzyjnych, w tym metody sekwencyjnego pokrywania. Omówione są również miary jakości reguł klasyfikacyjnych. Następnie zdefiniowane zostają reguły akcji oraz przedstawione jest studium literatury na temat metod ich odkrywania.

Rozdział 3.1 przedstawia pokryciowy algorytm indukcji reguł akcji w zadaniu klasyfikacji w dwóch wariantach: w przód (F) i w tył (B). Indukcja reguł w proponowanym algorytmie nadzorowana jest miarami oceny jakości klasycznych reguł

decyzyjnych [1, 15]. W rozdziale 3.3.1 podano modyfikację algorytmu pozwalającą indukować także reguły akcji w zadaniu regresji. Rozdział 3.3.2 krótko omawia reguły akcji w zadaniu analizy przeżycia, bez podawania szczegółów implementacyjnych.

W rozdziale 4 omówiono problemy pojawiające się przy próbach zastosowania reguł akcji, a w szczególności problem wyboru reguły akcji spośród wielu reguł pokrywających pewien przykład. Jako odpowiedź na ten problem podany został algorytm indukcji rekomendacji, będący specjalną formą reguły akcji indukowanej na potrzeby zadanego przykładu (obiektu). Indukcja rekomendacji odbywa się na podstawie struktury danych skupiającej w sobie wiedzę odkrytą przez cały zbiór reguł akcji nazywanej meta-tablicą. Tym samym rodzina algorytmów, nazywana w pracy modelami akcyjnymi, zostaje rozszerzona o kolejną metodę, dla której dane wejściowe stanowią wyniki działania algorytmów zaproponowanych w rozdziale 3.1. Ostatecznie, w podrozdziale 4.3 wprowadzono sposób weryfikacji reguł akcji, tj. metody pozwalającej ocenić różne modele akcji w kontekście ich przydatności przy przenoszeniu przykładów z klasy niepożądaney do klasy docelowej. Wraz z opisem metody prowadzona jest dyskusja na temat jej istotnych parametrów i interpretacji wyników. Podany sposób oceny jakości modeli akcji jest odpowiedni zarówno dla zadania klasyfikacji jak i regresji.

Opisany w rozdziale 4 sposób oceny zostaje zastosowany w rozdziale 5 do oceny metod opisanych w rozdziałach 3.1 i 4. W podrozdziale 5.1 omówiono parametry algorytmów wykorzystane w badaniach oraz konwencję nazewnictwa badanych wariantów metod. Sekcja 5.2 omawia metody analizy statystycznej stosowane w ocenie algorytmów klasyfikujących, które znajdują zastosowanie w dalszej części rozdziału. Wyniki doświadczenia prezentowane są osobno dla zadania klasyfikacji (podrozdział 5.3) i regresji (5.4). W ramach każdego z tych podrozdziałów najpierw omówiono charakterystykę zbiorów danych testowych, następnie opisano uzyskane wyniki ilościowo i jakościowo, uwzględniając m.in. nie tylko skuteczność modeli akcji ale także ich postać (np. długość i liczbę reguł), by ostatecznie przeprowadzić analizę statystyczną uzyskanych wyników. W rozdziale tym zawarto także parę poglądowych sposobów wizualizacji wyników, które mogą pozwolić łatwiej wybrać najlepszy wariant algorytmu indukcji reguł akcji do danego zastosowania. Celem autora nie jest wykazanie, która metoda jest uniwersalnie najlepsza, ale pokazanie w jaki sposób należy się kierować przy wyborze metody dla własnych celów.

2 Indukcja reguł akcji

W pracy zaproponowano cztery algorytmy indukcji reguł akcji: dwa warianty algorytmu indukcji reguł akcji z danych klasyfikacyjnych, jeden algorytm indukcji reguł z danych regresyjnych oraz algorytm rekomendacyjny, pozwalający uzyskiwać wyspecjalizowane reguły akcji dla obu typów danych. Każdy z nich generuje reguły o następującej postaci:

$$\text{Jeżeli } (A_1, V_{s_1} \rightarrow V_{t_1}) \wedge (A_2, V_{s_2} \rightarrow V_{t_2}) \wedge \dots \wedge (A_n, V_{s_n} \rightarrow V_{t_n}) \text{ to } C = c_s \rightarrow c_t \quad (1)$$

Przesłanka reguły składa się z koniunkcji akcji elementarnych. Każda akcja elementarna postaci $(A_i, V_{s_i} \rightarrow V_{t_i})$ dotyczy pewnego atrybutu warunkowego A_i oraz wskazuje zmianę z wartości źródłowej V_{s_i} na wartość docelową V_{t_i} , przy czym wartości V_{s_i}, V_{t_i} pochodzą z dziedziny atrybutu A_i . Konkluzja takiej reguły zawiera akcję elementarną dotyczącą atrybutu decyzyjnego w danym zbiorze. W przypadku danych klasyfikacyjnych konkluzja reguły będzie wskazywać na oczekiwaną zmianę przypisania klasy, w przypadku danych regresyjnych - zmianę wartości ciągłej zmiennej objaśnianej.

Wprowadzone algorytmy bazują na strategii sekwencyjnego pokrywania. Zaproponowano dwa warianty algorytmu dla uzyskiwania reguł akcji z danych klasyfikacyjnych o nazwach F (od ang. *Forward*) i B (ang. *Backward*). Algorytmy różnią się klasą, od której rozpoczyna się tworzenie reguł akcji. Ogólny schemat pokryciowego algorytmu indukcji reguł akcji zawarto w Algorytmie 1.

Jego analiza pokazuje, że w porównaniu do klasycznego algorytmu pokryciowej indukcji reguł decyzyjnych, indukcja reguł kończy się gdy pokrytych zostanie dostatecznie dużo przykładów klasy źródłowej (oznaczanej s), a nie gdy pokrytych zostanie odpowiednio wiele przykładów wszystkich klas.

W kroku specjalizacji reguły akcji, najpierw wyszukiwany jest najlepszy warunek elementarny opisujący przykłady klasy s . Wybór warunków elementarnych odbywa się z zastosowaniem zadanej przez użytkownika miary oceny jakości q . Następnie, do znalezionej już najlepszej reguły dla klasy źródłowej, szuka się odpowiedniego, najlepszego warunku elementarnego opisującego klasę docelową (t), ale z tym ograniczeniem, że poszukiwanie odbywa się jedynie na atrybucie użytym już w warunku elementarnym opisującym klasę s . Ostatecznie tworzona jest akcja elementarna, która dołączana jest do przesłanki budowanej reguły.

Warianty F i B algorytmu powstają poprzez sterowanie przypisaniem klas. Przykładowo, załóżmy, że pewnym zbiorze danych występują dwie klasy „+” i „-”, gdzie „+”

Algorytm 1 Pokryciowy algorytm indukcji reguł akcji

Wejście: $D(A, C)$ - zbiór przykładów (zbiór atrybutów A oraz atrybut decyzyjny C)

Wejście: q - miara oceny jakości reguł klasyfikacyjnych

Wejście: $mincov$ - wartość minimalnego pokrycia zbioru przykładów

Wejście: s - klasa źródłowa, $s \in V_C$

Wejście: t - klasa docelowa, $t \in V_C$

Wyjście: R - zbiór reguł akcji

```
1: function GENERUJREGUŁYAKCJI( $D, q, mincov, s, t$ )
2:    $R \leftarrow \emptyset$ 
3:    $S \leftarrow \{d : d \in D \wedge d[C] \equiv s\}$ 
4:    $D_t \leftarrow D \setminus S$ 
5:   while  $D \setminus D_t \neq \emptyset$  do
6:      $D_s \leftarrow S \cap D$ 
7:      $r \leftarrow$  SPECIALIZUJREGUŁĘAKCJI( $D, D_s, s, t, q, mincov$ )
8:      $r \leftarrow$  PRZYTNIJREGUŁĘAKCJI( $r, D, q$ )
9:      $R \leftarrow R \cup r$ 
10:     $D_s \leftarrow$  POKRYCIE( $r, D$ )
11:     $D \leftarrow D \setminus D_s$ 
12:   end while
13:   return  $R$ 
13: end function
```

jest klasą pożądaną, a „-” niepożądaną. Naturalnym będzie wyszukiwanie reguł akcji mających w konkluzji akcję $- \rightarrow +$. Aby uzyskać takie reguły, możemy skorzystać z przypisania $s = -$ i $t = +$. Wówczas powiemy, że algorytm 1 pracuje w wariancie F. Można jednak odwrócić przypisanie klas, tj. $s = +$ i $t = -$ i wzbogacić algorytm o krok końcowego przetwarzania, w ramach którego każda z odkrytych reguł zostanie odwrócona, tj. we wszystkich jej warunkach elementarnych wartości źródłowe zostaną zamienione miejscami z wartościami docelowymi. W efekcie uzyskamy nowy algorytm, nazywany wariantem B.

Warianty F i B reprezentują, odpowiednio, indukcję reguł akcji z perspektywy klasy źródłowej (F) i klasy docelowej (B). Wybór wariantu zależy od rozwiązywanego problemu, gdyż te dwa podejścia dają w wyniku zbiory reguł różniące się cechami jakościowymi takimi jak średnia długość reguły, średnie pokrycie klasy źródłowej i docelowej. W szczególności wariant B może być użyteczny w zadaniach, w których klasa docelowa jest niedoreprezentowana w zbiorze treningowym.

Prześledźmy wzrost jednej reguły akcji na zbiorze Monk1, syntetycznym zbiorze danych pochodzącym z repozytorium UCI [3]. Wszystkie atrybuty warunkowe $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ w tym zbiorze są kategoryczne. Atrybut decyzyjny *class* przyjmuje wartość 1, gdy spełniony jest jeden z dwóch warunków: $a_1 = a_2$ lub $a_5 = 1$. W przeciwnym przypadku, atrybut *class* przyjmuje wartość 0. Przyjmijmy, że klasa 0 jest klasą źródłową, a 1 – docelową.

Początkowo reguła ma pustą przesłankę i ustaloną konkluzję w postaci $(class, 0 \rightarrow 1)$. Warunki elementarne wybierane w następujących po sobie iteracjach, które budują akcje elementarne dodawane do przesłanki reguły akcji przedstawione zostały w poniższej tabeli:

iteracja	w_{best_s}	q_{r_s}	w_{best_t}	q_{r_t}
1	$(a_1 = 1)$	0.69	$(a_1 = 3)$	0.70
2	$(a_2 = 2)$	0.88	$(a_2 = 3)$	1.00
3	$(a_6 = 2)$	0.90	$(a_6 = 2)$	1.00

Tabela 1: Kolejne warunki elementarne wybierane w trakcie wzrostu reguły akcji, gdy użytą miarą oceny jakości jest precyzja reguły.

Po zakończeniu specjalizacji reguły, prezentuje się ona następująco:

przesłanka	q_{r_s}	q_{r_t}
$(a_1 = 1) \rightarrow (a_1 = 3) \wedge (a_2 = 2) \rightarrow (a_2 = 3) \wedge (a_6 = 2) \rightarrow (a_6 = 2)$	0.13	0.27
$(a_1 = 1) \rightarrow (a_1 = 3) \wedge ((a_2 = 2) \rightarrow (a_2 = 3)) \wedge (a_6 = 2) \rightarrow$	0.13	0.27
$(a_1 = 1) \rightarrow (a_1 = 3) \wedge (a_2 = 2) \rightarrow (a_2 = 3)$	0.21	0.27
$(a_1 = 1) \rightarrow (a_1 = 3) \wedge ((a_2 = 2) \rightarrow$	0.21	0.24
$(a_1 = 1) \rightarrow (a_1 = 3) \wedge \rightarrow (a_2 = 3)$	0.26	0.27

Tabela 2: Ilustracja przycinania reguły z zastosowaniem miary RSS.

IF $((a_1 = 1) \rightarrow (a_1 = 3)) \wedge ((a_2 = 2) \rightarrow (a_2 = 3)) \wedge ((a_6 = 2) \rightarrow (a_6 = 2))$ **THEN** $(class = 0) \rightarrow (class = 1)$.

po zakończonej fazie wzrostu rozpoczyna się przycinanie reguły. Kolejne kroki przycinania przedstawione zostały w tabeli 2, która prezentuje kolejne uzyskane przesłanki reguły akcji, oraz wyznaczoną jakość reguł składowych r_s i r_t . W procesie przycinania użyto miary RSS.

Na koniec przycinania, reguła przyjmuje następującą formę:

IF $((a_1 = 1) \rightarrow (a_1 = 3)) \wedge (\rightarrow (a_2 = 3))$ **THEN** $(class, 0) \rightarrow (class, 1)$

Ostateczna reguła poprawnie identyfikuje jeden z warunków przynależności do pożądanej klasy 1 w zbiorze *monk1* poprzez żądanie ustalenia wartości atrybutów a_1, a_2 na taką samą wartość.

Algorytm 1 został w pracy dostosowany do działania także na danych regresyjnych. Przez problem regresji rozumie się przewidywanie wartości pewnego wyróżnionego atrybutu numerycznego (zmiennej zależnej) na podstawie wartości pozostałych atrybutów w danym zbiorze danych (zmiennych niezależnych). W przypadku reguł regresyjnych zbiory przykładów pozytywnych i negatywnych zmieniają się wraz z rozszerzaniem (lub przycinaniem) przesłanki reguły. W przyjętej przez autora implementacji, konkluzja reguły regresji jest determinowana przez jej przesłankę. Ze względu na brak ostrych klas w zbiorach danych, w których zmienna decyzyjna jest ciągła, pojęcie klasy źródłowej i docelowej zanika w przypadku regresyjnych reguł akcji. Konkluzja takich reguł akcji zawiera dwie wartości, wskazujące na oczekiwaną zmianę zmiennej objaśnianej pod warunkiem wykonania wszystkich akcji z przesłanki, jednak użytkownik nie może wskazać, jaki zakres wartości jest preferowany. Jeśli wystąpi taka konieczność, należy dokonać dyskretyzacji atrybutu

decyzyjnego, wyróżnić klasę źródłową i docelową i przeprowadzić indukcję reguł akcji dla danych klasyfikacyjnych. Wobec powyższego, dla danych regresyjnych w pracy przedstawiono tylko jeden wariant algorytmu.

Oprócz algorytmów indukcji reguł akcji, opisano także algorytm odkrywania rekomendacji. Rekomendacja to szczególna reguła akcji, która tworzona jest specjalnie pod zadany przykład (pojedynczą obserwację). Intencją jest wskazanie możliwie najlepszych akcji dla danego przykładu. Motywacją utworzenia takiego algorytmu był problem zbiorów reguł akcji, w których jeden przykład może być pokryty przez więcej niż jedną regułę. Użytkownik metody, chcąc zastosować (wdrożyć w życie) reguły akcji, musi wówczas rozwiązać konflikt, tj. wybrać jedną z potencjalnie wielu reguł. Problem ten jest dwustopniowy - po pierwsze, przykład może być pokryty przez reguły o sprzecznych decyzjach, po drugie, przykład może być pokryty przez wiele reguł o tej samej decyzji. Reguły dodatkowo mogą zawierać zróżnicowaną liczbę akcji w swych przesłankach, zbudowanych w oparciu o różnorodne atrybuty. Algorytm rekomendacji rozwiązuje ten problem, poprzez podawanie kompletnych reguł akcji jako odpowiedź na zadany przykład.

Algorytm rekomendacji zaproponowany w pracy jako dane wejściowe przyjmuje zbiór odkrytych wcześniej reguł akcji. Na ich podstawie tworzy wewnętrzną strukturę danych, zwaną meta-tablicą, która stanowi niejako kompilację wiedzy odkrytej już przez algorytm indukcji reguł akcji. W trakcie tworzenia meta-tablicy wyznaczane są przedziały atrybutów warunków na podstawie akcji elementarnych zawartych w wejściowym zbiorze reguł. Przedziały te są jednak bardziej granularne i nie mające ze sobą nawzajem części wspólnych. Z każdym przedziałem skojarzona jest informacja o liczbie przykładów klasy źródłowej i docelowej pokrywanych przez te przedziały.

W celu wytworzenia rekomendacji, prowadzona jest wspinaczka (ang. *hill climbing*) po meta-tablicy. Proces ten jest nadzorowany miarą oceny jakości reguł. Źródłową część wynikowej rekomendacji stanowią wartości odczytane z analizowanego przykładu. Część docelowa jest wynikiem przeszukiwania meta-tablicy pod kątem możliwie najlepszego warunku, maksymalizującego miarę jakości wynikowej reguły.

2.1 Ocena skuteczności modeli akcji

Zbiory reguł akcji oraz rekomendacje utworzone dla pewnego zbioru przykładów są w pracy zbiorczo nazywane modelami akcji. Modele akcji mogą znaleźć zastosowanie w sugerowaniu działań mających na celu wpływanie na fizyczne obiekty tak, aby w

wyniku uzyskać oczekiwaną zmianę klasy tych obiektów. Działanie takie jednak może być kosztowne, czasochłonne lub - ogólnie mówiąc - skomplikowane we wdrożeniu. Trudno podjąć decyzję, który model akcji okaże się najlepszy w takim zastosowaniu, bez testów na obiektach rzeczywistych.

W pracy przedstawiono propozycję eksperymentu, który ma za zadanie określić skuteczność modelu akcji. Idea takiej oceny skuteczności opiera się o zastosowanie klasyfikatora wysokiej jakości jako punktu odniesienia dla dobroci modelu akcji. Przykładowo, jeśli dysponujemy klasyfikatorem, który z bardzo wysoką dokładnością jest w stanie rozpoznawać obiekty z pewnej populacji, na której chcemy zastosować reguły akcji, to przed wdrożeniem fizycznych działań możemy użyć tego samego klasyfikatora do stwierdzenia, czy przykłady opisujące obiekty z tej populacji poddane zmianom wartości atrybutów zgodnie z sugestiami podawanymi przez reguły akcji są uznawane jako należące do pożądanej klasy przez ten klasyfikator.

Zaproponowany eksperyment zakłada wykorzystanie klasyfikatora XGBoost [2] dla danych klasyfikacyjnych. Ten sam algorytm – po odpowiednim ustawieniu parametrów – znajdzie zastosowanie w ocenie reguł akcji dla danych regresyjnych. W eksperymencie zbiór testowy jest poddawany transformacji. Wartości atrybutów każdego z przykładów w tym zbiorze są modyfikowane zgodnie z sugestiami podawanymi przez testowany model akcji. Następnie na tak zmodyfikowanym zbiorze testowym przeprowadzana jest klasyfikacja. Jeśli klasyfikator rozpozna przetransformowany przykład jako należący do klasy docelowej, uznajemy to za sukces modelu akcji. Na tej podstawie można wyznaczyć kilka metryk opisujących skuteczność modeli akcji dla klasyfikacji i regresji. W pracy zaproponowano po 3 takie metryki dla każdego rodzaju danych.

3 Eksperymenty z danymi

W celu oceny efektywności zaproponowanych w pracy algorytmów przeprowadzono eksperymenty na zestawie 16 zbiorów danych klasyfikacyjnych i 6 zbiorach z danymi regresyjnymi.

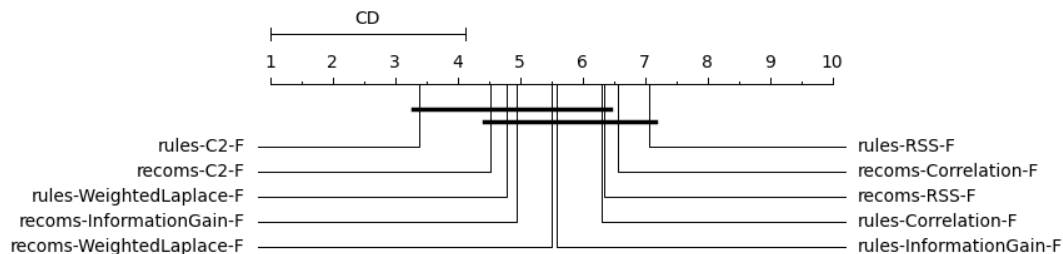
W przypadku klasyfikacji, doświadczenie polegało na przeprowadzeniu eksperymentu omówionego wcześniej na kilkunastu różnych modelach akcji. Algorytmy indukcji reguł były nadzorowane jedną z pięciu miar jakości. Miary jakości zostały wybrane ze względu na ich efektywność pod kątem dokładności i liczby reguł w wynikowych zbiorach reguł decyzyjnych [15]. Oprócz wariantu F i B metody indukcji reguł akcji

i odpowiednio algorytmu rekomendacji F i rekomendacji B, przetestowano także wariant algorytmu regułowego FB, w którym zbiór reguł stanowi sumę zbiorów reguł utworzonych przez warianty F i B osobne. Ostatecznie, porównano ze sobą grupę $2 \times 3 \times 5 = 30$ ($\{rules, recoms\} \times \{F, B, FB\} \times \{C2, Correlation, InformationGain, RSS, WeightedLaplace\}$) algorytmów. Angielskie nazwy *rules* i *recoms* oznaczają odpowiednio algorytm indukcji reguł i rekomendacji.

Jako czynnik oceniający dobroć reguł akcji posłużyła jedna z zaproponowanych w pracy metryk jakości. Oprócz samej analizy skuteczności w pracy przedstawiono kilka sposobów wizualizacji działania modeli akcji, mających na celu wskazanie alternatywnych sposobów doboru najlepszego algorytmu do rozwiązywanego zadania.

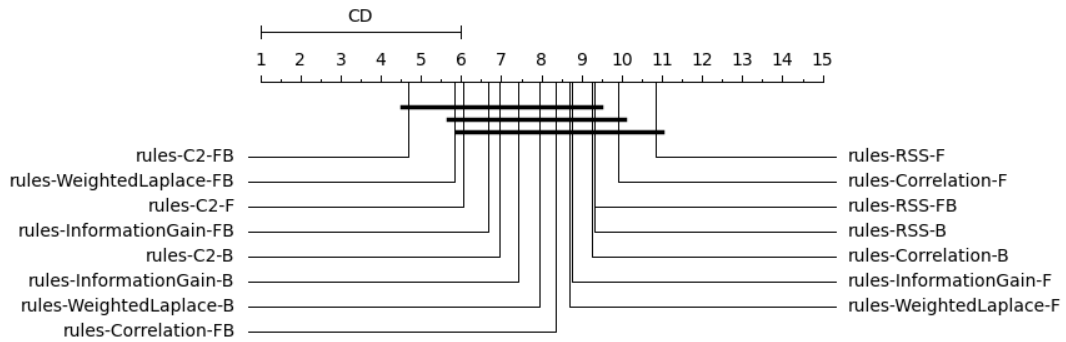
W celu wyłonienia najlepszego algorytmu (grupy algorytmów) przeprowadzono test Friedmana wraz z analizą post-hoc Nemenyi. Porównanie algorytmów przeprowadzono w podgrupach: metody nadzorowane tą samą miarą, metody pogrupowane po kierunku indukcji (F/B/FB), osobno modele akcyjne regułowe i oparte o algorytm rekomendacji.

Na rysunku 1 przedstawiono diagram CD uzyskany dla grupy modeli akcji z wariantu F na danych klasyfikacyjnych. Test Nemenyi na poziomie istotności $\alpha = 0.1$ wykrył dwie grupy algorytmów. Najczęściej najwyższe lokaty w rankingu metod zajmowały algorytmy nadzorowane miarą jakości C2.

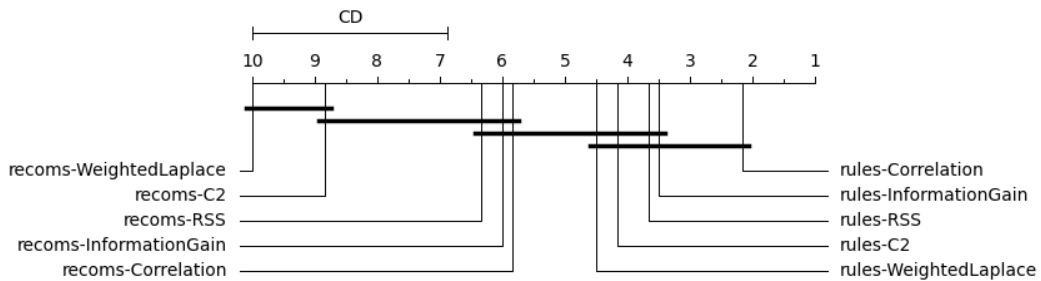


Rysunek 1: Diagram CD wskazujący na różnice w uzyskanej skuteczności transformacji przykładów między grupami modeli akcyjnych utworzonych z algorytmów regułowego i rekomendacji o kierunku indukcji *F*.

Diagram 2 obrazuje porównanie wszystkich wariantów algorytmu regułowego pod kątem uzyskanej skuteczności transformacji przykładów na danych klasyfikacyjnych. Ponownie okazało się, że zastosowanie miary C2 prowadzi do wysokiej skuteczności transformacji. Inną obserwacją jest wysoka skuteczność wariantu FB, którego jednak wadą jest konieczność dwukrotnego wykonania indukcji reguł akcji.



Rysunek 2: Diagram CD wskazujący na różnice w uzyskanej skuteczności transformacji przykładów między grupami modeli akcyjnych utworzonych na podstawie różnych wariantów algorytmu regułowego



Rysunek 3: Diagram CD wartości MAE dla wszystkich wariantów algorytmu rekomendacyjnego w zadaniu regresji.

W przypadku danych regresyjnych testy przeprowadzono na 6 zbiorach danych. Mniejsza jest również liczba testowanych algorytmów: $2 \times 5 = 10$ ($\{rules, recoms\} \times \{C2, Correlation, InformationGain, RSS, WeightedLaplace\}$), ze względu na brak wariantów związanych z kierunkiem indukcji.

Diagram CD przedstawiony na rys. 3 posiada odwróconą oś tak, aby najlepsza metoda znalazła się z lewej strony. Wynika to z tego, że dla danych regresyjnych przyjętą w testach metryką był błąd MAE. Im mniejsza wartość tego błędu, tym lepiej, jednak na diagramach CD tradycyjnie wysokie lokaty (tj. bliższe 1) przypisuje się algorytmom, które uzyskały wysoką wartość analizowanego wskaźnika. Z zamieszczonego diagramu można wysnuć wniosek, że algorytm rekomendacyjny na ogół jest skuteczniejszy w transformacji przykładów niż algorytm regułowy.

W tabeli 3 zaprezentowano uzyskaną skuteczność transformacji przykładów dla wszystkich zbiorów testowych. Przedstawione wyniki dotyczą algorytmu indukcji reguł i rekomendacji w wariancie B i są uśrednione na 10 przebiegach.

Wskazanie najlepszego wariantu algorytmu nie może się jednak odbyć bez głębszej analizy. Oprócz skuteczności transformacji należy zbadać zdolność modelu do pokrywania przykładów testowych, ponieważ model pokrywający wszystkie przykłady, ale czasem popełniający błąd, będzie zazwyczaj bardziej pożądanym od takiego, który jest w pełni skuteczny, ale tylko dla kilku przykładów ze zbioru testowego. Informacje o pokryciu na każdym ze zbiorów testowych, wyrażoną w procentach, zamieszczono w tabeli 4.

Wynika z niej przede wszystkim, że algorytm rekomendacji cechuje się znacznie lepszym pokryciem na każdym z testowanych zbiorów. Wyjątkiem jest zastosowanie miary WeightedLaplace, dla której algorytm, jakkolwiek skuteczny w transformacji okazuje się słaby w generowaniu reguł akcji dla zadanych przykładów. Ostatecznie, analizując łącznie tabelę 4 i diagram 3 dla danych regresyjnych należy zarekomendować zastosowanie algorytmu rekomendacyjnego nadzorowanego miarą C2.

Podsumowując, analiza przeprowadzona w rozdziale 6 referowanej pracy wskazuje, że najwyższą skuteczność transformacji uzyskuje się dla algorytmów nadzorowanych miarami C2 i WeightedLaplace, w przypadku danych klasyfikacyjnych. Prawidłowość ta jest widoczna dla wszystkich rodzin: indukcji „w przód” (wariant F), „wstecz” (B) oraz fuzji tych wariantów (FB). W przypadku danych regresyjnych, po uwzględnieniu wszystkich opisanych szczegółów, najlepiej spisują się algorytmy nadzorowane miarą C2. Pod kątem zdolności do pokrywania przykładów testowych metoda rekomendacji przewyższa metodę regułową zarówno w przypadku danych klasyfikacyjnych jak i regresyjnych.

Jeśli istotna jest ilość reguł akcji wygenerowanych przez algorytmy nadzorowane poszczególnymi miarami, to w przypadku danych regresyjnych średnio najwięcej reguł daje zastosowanie miary RSS, a najmniejszą liczbę reguł pozwala uzyskać miara InformationGain. W przypadku danych klasyfikacyjnych, zarówno przy indukcji w przód (wariant F) i jak wstecz (wariant B) najwięcej reguł daje miara WeightedLaplace, a najmniej RSS.

zbiór	algorytm	C2	Correlation	Information Gain	RSS	Weighted Laplace
car-reduced	rekomendacje	0.994	0.714	0.716	0.455	0.994
car-reduced	najlepsza reguła	1.000	0.782	0.779	0.406	1.000
credit-a	rekomendacje	0.717	0.544	0.549	0.546	0.208
credit-a	najlepsza reguła	0.910	0.454	0.451	0.471	0.838
credit-g	rekomendacje	0.652	0.932	0.930	0.925	0.697
credit-g	najlepsza reguła	0.883	0.903	0.912	0.901	0.910
diabetes-c	rekomendacje	0.689	0.843	0.856	0.854	0.724
diabetes-c	najlepsza reguła	0.981	0.937	0.990	0.951	0.895
echocardiogram	rekomendacje	0.978	0.978	0.978	0.944	1.000
echocardiogram	najlepsza reguła	0.977	1.000	1.000	0.994	1.000
heart-c	rekomendacje	0.430	0.612	0.606	0.385	0.527
heart-c	najlepsza reguła	0.693	0.440	0.454	0.355	0.688
heart-statlog	rekomendacje	0.483	0.592	0.696	0.625	0.446
heart-statlog	najlepsza reguła	0.912	0.852	0.917	0.867	0.733
hepatitis	rekomendacje	0.817	0.767	0.783	0.783	0.667
hepatitis	najlepsza reguła	0.847	0.563	0.765	0.723	0.943
horse-colic	rekomendacje	0.922	0.907	0.900	0.919	0.956
horse-colic	najlepsza reguła	0.924	0.850	0.852	0.891	0.953
hungarian-heart-disease	rekomendacje	0.774	0.547	0.571	0.500	0.397
hungarian-heart-disease	najlepsza reguła	0.836	0.618	0.626	0.586	0.492
iris-reduced	rekomendacje	1.000	1.000	1.000	1.000	1.000
iris-reduced	najlepsza reguła	1.000	1.000	1.000	1.000	1.000
monk1_train	rekomendacje	1.000	1.000	1.000	1.000	1.000
monk1_train	najlepsza reguła	0.763	0.782	0.755	0.809	0.747
mushroom	rekomendacje	0.952	0.944	0.944	0.944	0.491
mushroom	najlepsza reguła	0.698	0.635	0.635	0.635	0.599
tic-tac-toe	rekomendacje	0.947	0.798	0.806	0.762	0.936
tic-tac-toe	najlepsza reguła	0.947	0.921	0.934	0.925	0.953
titanic	rekomendacje	1.000	0.718	0.718	0.718	1.000
titanic	najlepsza reguła	0.567	0.702	0.706	0.711	0.359
vote	rekomendacje	0.988	0.982	0.988	0.988	0.997
vote	najlepsza reguła	0.921	0.988	0.988	0.988	0.980
średnia skuteczność	rekomendacje	0.834	0.805	0.815	0.772	0.752
średnia skuteczność	najlepsza reguła	0.866	0.777	0.798	0.763	0.818

Tabela 3: Skuteczność transformacji przykładów ze zbioru testowego z zastosowaniem wariantu B algorytmu indukcji reguł i rekomendacji. W ramach każdego ze zbiorów pogrubiono wynik lepszy. W przypadku remisu, żadna z wartości nie została wyróżniona. Ostatnie dwa wiersze zawierają wyniki uśrednione dla każdej miary na wszystkich zbiorach.

zbiór	algorytm	C2	Correlation	Information Gain	RSS	Weighted Laplace
auto-mpg	rekomendacje	100.00	100.00	100.00	100.00	20.00
auto-mpg	reguły	59.41	75.15	91.88	80.30	18.71
bodyfat	rekomendacje	100.00	100.00	100.00	100.00	30.20
bodyfat	reguły	92.60	93.80	94.20	92.00	90.20
boston_housing	rekomendacje	100.00	100.00	100.00	100.00	20.00
boston_housing	reguły	59.41	75.15	91.88	80.30	18.71
gas	rekomendacje	100.00	100.00	100.00	100.00	0.00
gas	reguły	93.62	94.83	93.62	94.83	91.90
houses_taiwan	rekomendacje	100.00	100.00	100.00	100.00	0.00
houses_taiwan	reguły	40.72	40.12	39.28	44.34	40.84
methane-train-minimal	rekomendacje	100.00	100.00	100.00	100.00	80.00
methane-train-minimal	reguły	79.00	85.75	86.25	89.00	27.75

Tabela 4: Średnie pokrycie zbioru testowego w trakcie oceny jakości modelu akcyjnego, dane w procentach, uśrednione z 10 przebiegów.

4 Podsumowanie

Wszystkie zaproponowane algorytmy wykorzystują paradygmat sekwencyjnego pokrywania do indukcji reguł akcji. Wpływ na wynikową postać reguł ma zastosowana miara jakości nadzorująca wzrost i przycinanie reguły, a dodatkowo w przypadku danych klasyfikacyjnych – kierunek indukcji.

W pracy zaproponowano także sposób eksperymentalnej weryfikacji skuteczności zbiorów reguł akcji w zadaniu transformacji przykładów. Wykorzystując opisaną metodę porównano ze sobą wyniki działania podanych wcześniej algorytmów. Przeprowadzona analiza wykazała, że stosowanie różnych miar jakości i kierunków indukcji prowadzi do uzyskania zbiorów reguł o dobrych zdolnościach prognostycznych i zróżnicowanych cechach jakościowych związanych ze zdolnościami opisowymi uzyskanych zbiorów reguł akcji.

W szczególności, metoda agregacji wiedzy zawartej w istniejących zbiorach reguł, nazwana algorytmem rekomendacji, pozwala na uzyskanie specjalizowanych zbiorów

reguł charakteryzujących się bardzo wysoką zdolnością do rozpoznawania nieznanych do tej pory przykładów i podawania skutecznych akcji.

Przedstawione wyniki pozwalają uzasadnić sformułowane w pracy tezę i tezy pomocnicze.

Dalsze prace w obszarze odkrywania reguł akcji mogą przebiegać w kilku kierunkach. W algorytmach indukcji reguł decyzyjnych oprócz zastosowanego w metodach zaproponowanych w niniejszej pracy procesu przycinania reguł, stosuje się także krok filtracji [8, 11]. Filtracja polega na usuwaniu ze zbioru tych reguł, które pod względem przyjętego kryterium jakości są uznane za niepotrzebne. Jest więc to forma przycinania zbioru reguł, która nie zmienia postaci poszczególnych reguł. Interesującym byłoby zbadanie wpływu filtracji na skuteczność modelu akcji. Filtracja zbiorów reguł doprowadziłaby wówczas do uzyskania bardziej zwartej reprezentacji wiedzy o koniecznych zmianach wartości atrybutów, przy – potencjalnie – zachowaniu tej samej skuteczności.

Reguły akcji mogą znaleźć zastosowanie także w analizie ciągów czasowych. Nietrudno wyobrazić sobie praktyczne zastosowania w obszarach konserwacji predykcyjnej czy analizie danych ekonomicznych. Algorytm pokryciowy z powodzeniem odkrywa reguły akcji w zadaniach regresji, klasyfikacji i analizy przeżycia, co pozwala żywić nadzieję, że to podejście sprawdzi się także w analizie danych temporalnych.

W przedstawionych w pracy algorytmach ta sama miara nadzoruje proces wzrostu (specjalizacji) i przycinania reguł. W literaturze [15] do indukcji reguł decyzyjnych spotyka się stosowanie kombinacji miar: jedna miara nadzoruje wzrost reguły, a inna jej przycinanie. Interesującym byłoby zbadanie, czy podobne podejście do indukcji reguł akcji pozwala uzyskać wyższą skuteczność.

Pokryciowy algorytm indukcji reguł decyzyjnych pozwala na wykorzystanie w procesie odkrywania reguł znanej już wiedzy, wprowadzanej przez eksperta domenowego. Uzyskiwane zbiory reguł charakteryzują się wyższą skutecznością, niż odkrywane w pełni automatycznie [12]. Podobne podejście mogłoby znaleźć zastosowanie w odkrywaniu reguł akcji.

Literatura

- [1] An, A., Cercone, N. Rule quality measures for rule induction systems: Description and evaluation. *Computational Intelligence*, 17:409–424, 08 2001.

- [2] Chen, T., Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, strona 785–794, New York, NY, USA, 2016. Association for Computing Machinery.
- [3] Dheeru, D., Karra Taniskidou, E. UCI machine learning repository, 2017.
- [4] Fürnkranz, J. Separate-and-conquer rule learning. *Artif. Intell. Rev.*, 13(1):3–54, Luty 1999.
- [5] Fürnkranz, J., Gamberger, D., Lavrač, N. *Foundations of Rule Learning*. Cognitive Technologies. Springer Berlin Heidelberg, 2012.
- [6] He, Z., Xu, X., Deng, S., Ma, R. Mining action rules from scratch. *Expert Systems with Applications*, 29:691–699, 10 2005.
- [7] Powell, L., Gelich, A., Ras, Z. W. How to raise artwork prices using action rules, personalization and artwork visual features. *J. Intell. Inf. Syst.*, 57:583–599, 2021.
- [8] Ågotnes, T., Komorowski, J., Løken, T. Taming large rule models in rough set approaches. Żytkow, J. M., Rauch, J., redaktorzy, *Principles of Data Mining and Knowledge Discovery*, strony 193–203, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.
- [9] Raś, Z., Dardzińska, A. Action rules discovery based on tree classifiers and meta-actions. *ISMIS09*, wolumen 5722, strony 66–75, 09 2009.
- [10] Raś, Z. W., Wieczorkowska, A. Action-rules: How to increase profit of a company. Zighed, D. A., Komorowski, J., Żytkow, J., redaktorzy, *Principles of Data Mining and Knowledge Discovery*, strony 587–592, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.
- [11] Sikora, M. Filtracja zbioru reguł decyzyjnych wykorzystująca funkcje oceny jakości reguł. *Studia Informatica*, number 46, strony 57–72, 2001.
- [12] Sikora, M., Gudyś, A., Wróbel, Ł. GuideR: a guided separate-and-conquer rule learning in classification, regression, and survival settings. *Knowledge-Based Systems*, 173:1–14, 2019.

- [13] Tarnowska, K., Raś, Z., Daniel, L. *Recommender System for Improving Customer Loyalty*. Springer International Publishing, 01 2020.
- [14] Touati, H., Raś, Z. W., Studnicki, J., Wiczorkowska, A. A. Mining surgical meta-actions effects with variable diagnoses' number. Andreasen, T., Christiansen, H., Cubero, J.-C., Raś, Z. W., redaktorzy, *Foundations of Intelligent Systems*, strony 254–263, Cham, 2014. Springer International Publishing.
- [15] Wróbel, Ł., Sikora, M., Michalak, M. Rule quality measures settings in classification, regression and survival rule induction — an empirical approach. *Fundamenta Informaticae*, 149:419–449, 12 2016.
- [16] Yang, Q., Yin, J., Ling, C., Pan, R. Extracting actionable knowledge from decision trees. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):43–56, 2007.