



UNIWERSYTET ŚLĄSKI  
W KATOWICACH

IMI. RPITT  
25.08.2022  
M. Skon

dr hab. Beata Zielosko, prof. UŚ  
Instytut Informatyki  
Wydział Nauk Ścisłych i Technicznych  
Uniwersytet Śląski w Katowicach

Sosnowiec, 22 sierpnia 2022r.

## Recenzja rozprawy doktorskiej

Tytuł rozprawy: **INDUKCJA REGUŁ AKCJI NA PODSTAWIE METODY  
SEKWENCYJNEGO POKRYWANIA**

Autor rozprawy: **mgr inż. Paweł Matyszok**

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

Promotor pomocniczy: **dr inż. Łukasz Wróbel**

Dziedzina: nauki techniczne

Dyscyplina: Informatyka Techniczna i Telekomunikacja

Recenzja wykonana na podstawie zlecenia Przewodniczącego Rady  
Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej,  
prof. dr hab. inż. Andrzeja Polańskiego, z dnia 27 czerwca 2022 r., zgodnie  
z uchwałą Rady Dyscypliny.

### 1. Obszar problemowy pracy

Rozwój technologiczny jak i dynamika procesów występujących  
w otaczającym świecie wiąże się z koniecznością przetwarzania różnego  
rodzaju danych. Problemem jest zastosowanie odpowiednich metod, które





pozwołą wydobyć wiedzę z tych danych. Celem procesu odkrywania wiedzy jest jej wykorzystanie do podejmowania decyzji przynoszących korzyść użytkownikowi w różnego rodzaju zastosowaniach np. decyzji biznesowych pozwalających na zwiększenie zysku, zdobycie nowych i zatrzymanie istniejących klientów, czy decyzji związanych z zastosowaniami medycznymi. Podczas procesu odkrywania wiedzy z danych KDD (ang. *knowledge data discovery*) jednym z istotnych elementów jest etap eksploracji danych (ang. *data mining*), w tym tworzenie modelu danych i jego ocena. Najczęściej oceniane są zdolności predykcyjne wyindukowanego modelu, jednak równie często, zwłaszcza w przypadku podejmowania decyzji przez użytkownika dotyczących np. finansów związanych z działalnością firmy, czy zdrowia pacjenta, zdolności opisowe modelu i sposób reprezentacji odkrytej wiedzy odgrywa istotną rolę. W tym kontekście reguły decyzyjne stanowią jedną z popularnych form reprezentacji wiedzy i mają wiele różnorodnych zastosowań, m.in. ze względu na prostotę i intuicyjność formy zapisu, co z kolei przekłada się na interpretację wyników przez eksperta dziedzinowego. Istnieją obszerne studia literaturowe dotyczące podejść i algorytmów indukcji reguł oraz stosowanych miar oceny ich jakości, w zależności od celów prowadzonej analizy.

Jednym ze stosunkowo nowych obszarów wykorzystania reguł decyzyjnych jest tworzenie tzw. reguł akcji pozwalających na badanie jak zmiany wartości atrybutów występujących w przesłankach reguły wpływają na zmianę wartości atrybutu decyzyjnego. Reguły te sugerują podjęcie działań prowadzących do zmiany stanu obiektu i mogą też pełnić rolę rekomendacji dla użytkownika. Wykrywanie jak i sugerowanie działań prowadzących do zmiany stanu obiektu, które jednocześnie mogą przynieść „korzyść” dla użytkownika odgrywa istotną rolę w procesach podejmowania decyzji, co potwierdzają przytoczone w rozprawie przykłady zastosowań reguł akcji.

Recenzowana praca dotyczy indukcji reguł akcji z wykorzystaniem paradygmatu sekwencyjnego pokrywania. Główne wyniki rozprawy

obejmują nowe metody indukcji reguł akcji zarówno w kontekście zadania klasyfikacji danych jak i regresji, uwzględniając różne kryteria jakości reguł. Analiza opracowanych rozwiązań pozwala na ich ocenę, zarówno z punktu widzenia reprezentacji wiedzy tj. uwzględniając zdolności opisowe modelu danych, jak i z punktu widzenia odkrywania wiedzy tj. uwzględniając zdolności prognostyczne modelu. Metody te zapewniają możliwość działania algorytmów także w przypadku brakujących wartości atrybutów. Ponadto Autor zaproponował metodę generowania rekomendacji dla określonego obiektu.

Podsumowując uważam, że podjęta w badaniach problematyka, sformułowany cel i zadania badawcze są interesujące i istotne dla rozwoju dyscypliny. Autor przeprowadził analizę istniejących podejść w zakresie realizowanego tematu badawczego, zaproponował autorskie rozwiązania oraz zweryfikował ich skuteczność, co pozwoliło na przeprowadzenie analizy porównawczej zaproponowanych algorytmów oraz ocenę opracowanych rozwiązań.

## **2. Kompozycja i zawartość pracy**

Rozprawa została napisana w języku polskim i składa się z sześciu rozdziałów, oraz bibliografii.

Rozdział pierwszy stanowi wstęp do tematyki rozprawy. Autor przedstawił motywację do podjęcia badań dotyczących odkrywania reguł akcji, tezę i cele pracy oraz strukturę rozprawy.

Rozdział drugi przedstawia podstawowe definicje i miary oceny jakości reguł decyzyjnych, ich zastosowanie w procesie klasyfikacji danych oraz wybrane algorytmy indukcji tych reguł, w tym opis metod dotyczących paradygmatu sekwencyjnego pokrywania. Dalszą część rozdziału stanowi opis reguł akcji, ich definicje, znane z literatury podejścia do indukcji reguł akcji oraz przykłady ich zastosowań.



Rozdział trzeci przedstawia zaproponowane przez Autora metody indukcji reguł akcji z zastosowaniem paradygmatu sekwencyjnego pokrywania, które można sklasyfikować, w zależności od kierunku indukcji na metody indukcji „w przód” (ang. *forward*) oraz „wstecz” (ang. *backward*). W rozdziale tym Autor odniósł się także do problemu indukcji reguł akcji w zadaniu regresji oraz zaproponował algorytm ich konstruowania.

W rozdziale czwartym Autor przedstawił podejście do tworzenia rekomendacji na podstawie zbiorów reguł akcji dotyczących pojedynczych przykładów. Ważną częścią tego rozdziału jest weryfikacja reguł akcji i rekomendacji w kontekście ich wykorzystania w zadaniach klasyfikacji i regresji oraz zaproponowane przez Autora metryki jakości modelu.

Rozdział piąty dotyczy eksperymentów przeprowadzonych m.in. na zbiorach danych dostępnych w repozytorium UCI Machine Learning Repository, z zastosowaniem zaproponowanych i opisanych w pracy metod. Eksperymenty te dotyczą zarówno zadań klasyfikacji jak i regresji a przedstawione wyniki określają dwa aspekty indukcji reguł akcji tj. ocena ich właściwości objaśniających jak i prognostycznych. Istotnym elementem przeprowadzonych badań jest analiza skuteczności reguł akcji i rekomendacji związana z rzeczywistym działaniem dotyczącym zmiany klasy źródłowej obiektu na docelową. Na uwagę zasługuje także przeprowadzona analiza statystyczna uzyskanych wyników.

Rozdział szósty stanowi podsumowanie rozprawy. Autor wskazał główne osiągnięcia i możliwe kierunki kontynuacji prac badawczych.

Rozprawę kończy bibliografia zawierająca 111 pozycji. Cytowane prace to przede wszystkim publikacje z czasopism naukowych oraz prace konferencyjne.

Podsumowując tę część recenzji uważam, iż praca jest obszerna ale też stanowi odzwierciedlenie złożoności podjętej problematyki badawczej. Rozprawa

przedstawia dużą zawartość merytoryczną, w tym naukową i eksperymentalną związaną z przeprowadzonymi badaniami.

### **3. Opinia o rozprawie i oryginalność osiągnięć**

Ważnym elementem pracy jest podjęta tematyka badawcza, z jednej strony odnosząca się do popularnej formy reprezentacji wiedzy jaką są reguły decyzyjne, a z drugiej strony dotycząca opierających się na nich reguł akcji, których metody indukcji stanowią złożony i aktualny problem badawczy.

Algorytmy indukcji reguł decyzyjnych są od wielu lat rozwijane i stanowią przedmiot różnorodnych badań. W tym zakresie istnieją ugruntowane metody indukcji reguł decyzyjnych z wykorzystaniem idei sekwencyjnego pokrywania. Zatem zastosowanie tego paradygmatu do indukcji reguł akcji jest interesującym, nowym rozwiązaniem, pozwalającym na wykorzystanie i weryfikację jego zalet w odniesieniu do reguł akcji. Warty podkreślenia jest fakt, iż Autor podszedł do zadania w sposób kompleksowy tj. uwzględnił dwa podstawowe aspekty indukcji reguł, czyli ich zdolności opisowe jak i prognostyczne, co zostało odzwierciedlone w prezentowanych wynikach eksperymentów.

W ramach kompleksowości proponowanego przez Autora podejścia do indukcji reguł akcji należy uwzględnić fakt, iż badania dotyczyły nie tylko zastosowanej w procesie indukcji miary jakości reguł ale także kierunku indukcji reguł akcji. Jak wykazano w części eksperymentalnej uzyskane modele reguł akcji indukowane „w przód” oraz „wstecz” różnią się między sobą z punktu widzenia zdolności opisowych, jednak z punktu widzenia oceny ich skuteczności połączenie tych dwóch wariantów przyczynia się do poprawy uzyskiwanych wyników. W tej sytuacji Autor zwrócił uwagę, iż metoda analizy powinna być dostosowana do danych, gdyż jest zależna od konkretnych zastosowań praktycznych a sposób badania skuteczności modeli akcyjnych powinien uwzględniać specyficzne wymagania użytkownika.

Kolejnym ważnym elementem proponowanego przez Autora rozwiązania jest adaptacja algorytmu indukcji reguł akcji do pracy z atrybutami o wartościach ciągłych, co eliminuje etap dyskretyzacji danych i analizę jego wpływu na wyindukowany zbiór reguł. W opisaney przez Autora implementacji w przypadku wyznaczania klasy docelowej reguły regresji wykorzystano wartość odchylenia standardowego wartości atrybutu decyzyjnego, odczytanych z przykładów pokrywanych przez przesłankę reguły. Ponadto, Autor dostosował zaproponowane przez siebie rozwiązanie do pracy z danymi o wartościach brakujących wykorzystując podejście oparte na pomijaniu atrybutów z takimi wartościami w procesie generowania warunków elementarnych reguł akcji.

Kolejna propozycja Autora dotyczyła algorytmu rekomendacji mającego na celu wskazanie najlepszych akcji prowadzących do zmiany przyporządkowania do klasy docelowej, dla określonego obiektu. W tym kontekście Autor także uwzględnił zarówno ciągłe jak i nominalne wartości atrybutów. Rozwiązanie to opiera się na budowie tzw. meta-tablicy stanowiącej iloczyn kartezyjski wartości tzw. meta-atributów czyli wszystkich zakresów wartości atrybutów jakie wystąpiły w regułach będących danymi wejściowymi algorytmu. Celem tworzenia tej struktury jest wyznaczenie w ramach każdego atrybutu warunkowego takich przedziałów wartości, które zapewnią utworzenie jak najbardziej precyzyjnych akcji. Rekomendacja przyjmuje formę reguły akcji, której część źródłowa tworzona jest na podstawie wpisów w meta-tablicy, które odpowiadają wartościom atrybutów w analizowanym przykładzie. Docelowa część rekomendacji jest wyznaczana w procesie wyszukiwania i tworzenia koniunkcji, na podstawie wpisów w meta-tablicy, najlepszych warunków elementarnych, gdzie kryterium stanowi wybrana przez użytkownika miara oceny jakości reguł.

Ważnym i interesującym aspektem pracy są badania dotyczące weryfikacji reguł akcji i rekomendacji mające na celu uzyskanie informacji czy wykonanie akcji sugerowanych przez model znajduje pokrycie w danych. W tym celu,



zarówno w zadaniu klasyfikacji jak i regresji został wprowadzony tzw. klasyfikator weryfikujący dla określenia poziomu skuteczności rozpoznawania przykładów poszczególnych klas oraz zaproponowane zostały metryki dla oszacowania skuteczności modelu akcji. Takie podejście zapewnia bardziej wiarygodną weryfikację procesu odkrywania wiedzy z wykorzystaniem reguł akcji, a w zastosowaniach praktycznych pozwoli na symulację modelu reguł akcji przed ich wdrożeniem.

Podsumowując uważam, że zaproponowane, opisane i zrealizowane w ramach przedstawionej rozprawy badania stanowią ważny wkład w rozwój metod dotyczących indukcji reguł akcji, zwłaszcza w oparciu o paradygmat sekwencyjnego pokrywania. Rozprawa prezentuje dobry poziom naukowy i zasługuje na ocenę pozytywną.

Praca napisana jest starannie, bez poważnych pomyłek językowych aczkolwiek występują błędy redakcyjne. Zawarte w rozprawie przykłady pozwalają lepiej zobrazować i wyjaśnić działanie przedstawionych algorytmów.

#### 4. Uwagi i problemy do dyskusji

Podczas opisywania proponowanych przez Autora metod zabrakło analizy lub przynajmniej komentarza dotyczącego złożoności obliczeniowej proponowanych i stosowanych algorytmów, np. w przypadku algorytmu rekomendacji wykorzystującego meta-tablicę.

Pewnym niedociągnięciem rozprawy jest brak nawiązania do zasady minimalnego opisu MDL (ang. *minimum description length*) principle wprowadzonej przez Rissanena, która może być wykorzystywana w podejmowaniu decyzji o wyborze najlepszych warunków elementarnych podczas indukcji reguł (cytowana w rozprawie praca 102 oraz np. Cichosz P.: "Systemy uczące się"). Zasada ta jest związana z kodowaniem danych oraz reprezentacją hipotez. W łącznym kodzie wyróżnia się część związaną



z dopasowaniem hipotezy do przykładów uczących (interpretowaną jako błąd na zbiorze uczącym) oraz część związaną ze złożonością hipotezy (rozumianą jak długość jej optymalnego kodu). Dla danego zbioru przykładów i różnych hipotez należy wybrać tę, która minimalizuje łączną długość kodu. Ta optymalizacja powinna uwzględniać specyfikę danych.

Zastosowane w pracy podejście do brakujących wartości polegające na ignorowaniu atrybutów z takimi wartościami może doprowadzić do utraty pewnych informacji. Warto byłoby zbadać jak jakość klasyfikatora zależy od innych strategii imputacji brakujących wartości np. wykorzystanie najczęstszej wartości atrybutu (ang. *common value strategy*).

W rozprawie brakuje spisu tabel, rysunków i algorytmów, co ułatwiłoby nawigację po treści pracy i zapewniłoby jej kompletność.

Podczas opisu wybranych algorytmów indukcji reguł decyzyjnych (rozdział 2.3) Autor mógł określić kryterium ich opisu np. indukcja pośrednia za pomocą drzew decyzyjnych, należało tutaj wspomnieć także o reduktach oraz indukcja bezpośrednia.

We Wprowadzeniu podczas opisu uczenia nienadzorowanego Autor powinien wymienić grupowanie jako przykład. Aktualny opis mógłby sugerować, iż grupowanie to główna metoda uczenia nienadzorowanego.

Przykład 4.2.3 wydaje się być niekompletny, gdyż dotyczy tylko wyznaczenia źródłowej części rekomendacji dla obiektu.

#### **Uwagi redakcyjne:**

Praca napisana jest starannie jednak występują błędy edycyjne i stylistyczne. W treści pracy na końcu linii bardzo często występują tzw. „sieroty” (przyimki „i”, „w”, „z”), które powinny zostać przeniesione do nowej linii.

Kropka na końcu tytułu rozdziału 4 powinna zostać usunięta.

Wprowadzenie str. 7 linia 9 – powinno być reguł akcji.

Rozdział 2.1, str. 13 linia 2 – powinno być atrybut decyzyjny.



Odwołania w treści tekstu do tabel, rysunków, algorytmów czasem zaczynają się z dużej, czasem z małej litery np. „algorytm 2” lub są pominięte np. „przedstawiony w 2.1” – powinno być przedstawiony w Tabeli 2.1.

W Przykładzie 2.2.1 (str. 20) brakuje odwołania oraz opisu (tytułu) tabelki przedstawionej na początku strony 21.

Podpisy pod rysunkami powinny zostać ujednolicone, w niektórych występują kropki na końcu, w innych nie, np. Rysunki 5.9 i 5.10.

Nagłówek Algorytmu 2 powinien być bardziej uszczegółowiony w stosunku do tytułu Algorytmu 1.

W Rozdziale 3.3.1, str. 72, Autor odwołuje się do linii 13 i 16 w Algorytmie 13, numery linii nie zostały umieszczone w opisie tego algorytmu na str.74.

Przedstawione powyżej uwagi, kwestie dyskusyjne oraz dostrzeżone uchybienia nie umniejszają wartości całej pracy i nie zmieniają mojej pozytywnej opinii o rozprawie. Raczej wynikają z chęci podjęcia dyskusji i mogą być traktowane jako kierunki dalszych badań.

## 5. Wnioski

Pan Paweł Matyszok przedstawił rozprawę doktorską stanowiącą oryginalne rozwiązanie problemu naukowego z zakresu indukcji reguł akcji w oparciu o paradygmat sekwencyjnego pokrywania. Opracowane metody zostały zweryfikowane w sposób eksperymentalny.

Autor wykazał się dużą wiedzą w zakresie tematyki rozprawy, umiejętnością pracy naukowej oraz znajomością metod badawczych. Osiągnięte wyniki świadczą o dobrym przygotowaniu Autora do pracy naukowej.

Ponadto, należy uwzględnić, iż Pan Paweł Matyszok jest współautorem pięciu publikacji, w tym czterech prac bezpośrednio dotyczących realizowanej tematyki rozprawy, a jedna z nich to praca opublikowana w renomowanym czasopiśmie Information Sciences.



UNIWERSYTET ŚLĄSKI  
W KATOWICACH

Recenzowana praca spełnia wymagania stawiane rozprawom doktorskim przez ustawę o stopniach naukowych i tytule naukowym w dyscyplinie Informatyka Techniczna i Telekomunikacja. Zatem wnoszę, by mgr Paweł Matyszok został dopuszczony do publicznej obrony.

Beata Ziobło

