

Prof. dr hab. inż. Katarzyna Stapor
Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Instytut Informatyki

RECENZJA

pracy doktorskiej mgr inż. **Adama Skowrona**

p.t. „**Indukcja i optymalizacja regułowych modeli danych za pomocą pokryciowych algorytmów indukcji reguł regresyjnych**”

(ang. “**Sequential covering regression rule induction and optimization of regression rule-based data models**”)

1. **Obszar problemowy rozprawy oraz jej pozycja w stosunku do aktualnego stanu wiedzy**

Recenzowana praca doktorska jest rozprawą z obszaru uczenia maszynowego, a konkretnie dotyczy problemu indukcji reguł dla problemów regresyjnych. W dziedzinie klasyfikacji przykładów opracowano wiele algorytmów indukcji, stosujących różnorodne strategie poszukiwania reguł. Można tutaj wymienić metody: bazujące na teorii zbiorów przybliżonych, poszukujące reguł asocjacyjnych, stosujące strategie ewolucyjne oraz przede wszystkim metody polegające na tzw. sekwencyjnym pokrywaniu dostępnego zbioru przykładów. Idea sekwencyjnego pokrywania polega na iteracyjnej indukcji kolejnych reguł w taki sposób, aby kolejna z wyznaczanych reguł pokrywała jak najwięcej niepokrytych dotychczas przykładów reprezentujących pojęcie, które w danej chwili działania algorytmu jest klasą docelową. Efektywność podejść pokryciowych w zastosowaniu do różnego rodzaju danych (syntetycznych oraz opisujących rzeczywiste i praktyczne problemy) potwierdzona jest w dużej liczbie publikacji.

Szerokie zastosowanie algorytmów pokryciowych do rozwiązywania zadań klasyfikacji nie przełożyło się jak dotąd na analizę możliwości zastosowania pokryciowych algorytmów indukcji reguł do rozwiązywania problemów regresyjnych (takich w których atrybut decyzyjnych jest typu numerycznego). W literaturze przedmiotu można znaleźć zaledwie kilka prób adaptacji algorytmów indukcji reguł klasyfikacyjnych do rozwiązania problemów regresyjnych.

Przygotowana przez mgr inż. Adama Skowrona praca doktorska, podejmuje – moim zdaniem z sukcesem – tematykę indukcji i optymalizacji reguł regresyjnych za pomocą podejścia pokryciowego, wypełniając w ten sposób istniejącą lukę.

Tematykę rozprawy uważam za oryginalną, ważną oraz aktualną dla współczesnych prac z zakresu indukcji reguł. Podjęta przez Doktoranta tematyka jest oryginalna i istotna z naukowego punktu widzenia, a opracowane w ramach rozprawy metody i algorytmy z pewnością będą mieć duże znaczenie dla praktyki .

2. Zawartość rozprawy

Rozprawa składa się z siedmiu rozdziałów oraz bibliografii liczącej 137 pozycji.

W rozdziale pierwszym scharakteryzowano strukturę pracy, jej cele oraz wartości dodane. Rozdział drugi jest wprowadzeniem w regułowe modele danych. W rozdziałach - trzecim i czwartym opisane zostały zaproponowane przez Doktoranta sposoby realizacji dwóch podstawowych celów pracy: w rozdziale trzecim - modyfikacje dwóch strategii: bottom-up i top-down dla indukcji reguł regresyjnych, w czwartym natomiast optymalizacja reguł. W rozdziale piątym przedstawiono przeprowadzone badania empiryczne mające na celu ewaluację zaproponowanych algorytmów, w rozdziale szóstym natomiast dokonano tej ewaluacji na zbiorach rzeczywistych. W ostatnim, siódmym rozdziale znajduje się podsumowanie pracy i wskazówki dalszego rozwoju zaproponowanych metod.

3. Uzyskane wyniki

Realizacja postawionych celów doprowadziła do szeregu oryginalnych, szczegółowych i konkretnych osiągnięć. W pracy przedstawiono w sumie cztery algorytmy: dwa działające na zasadzie wstępującego budowania przesłanki reguły; oraz dwa - zstępującego budowania przesłanki. Nową strategią nazwaną przez Doktoranta *fixed*, jest strategia polegająca na budowaniu reguły w taki sposób, aby jej konkluzja miała z góry określoną postać (atrybut decyzyjny ma określoną wartość). Jest to podejście nowe, gdyż w przypadku reguł regresyjnych konkluzja reguły zmienia się wraz ze zmieniającymi się w trakcie indukcji przesłankami, co nie pozwala kontrolować modelu. Doktorant zbadał również wpływ miar oceniających jakości reguł na błąd i liczbę reguł generowanych przez algorytm, efektywność różnych strategii rozstrzygania konfliktów „klasyfikacji” – proponując również własne nazwane w pracy miary *intersection of coverage*. Podejście to podczas rozstrzygania konfliktu uwzględnia część wspólną pokrywaną przez reguł biorące udział w „klasyfikacji”.

W pracy zaproponowano (bądź zaadaptowano) kilka algorytmów filtracji reguł -- Doktorant zaproponował dwa własne algorytmy *Disjoint* oraz *ForwBack*. Praca zawiera także elementy statystycznej analizy wyznaczanych reguł. Efektywność wszystkich algorytmów zbadano eksperymentalnie na wielu zbiorach danych.

Do szczególnie wartościowych elementów pracy o charakterze oryginalnym należy zaliczyć:

- Opracowanie zmodyfikowanej strategii zstępującej indukcji reguł regresyjnych.
- Poprzez wprowadzenie modyfikacji polegającej na wprowadzeniu pewnej stałej do konkluzji każdej generowanej reguły – opracowanie quasi-pokryciowej wersji wstępującej i zstępującej.
- Zaproponowanie metody przeszukiwania tabu do znalezienia optymalnej reguły w fazie jej przycinania.
- Zaproponowanie dwóch metod filtracji reguł regresyjnych - metody *Disjoint* i *ForwBack*.
- Zaproponowanie nowej metody rozwiązywania konfliktów pomiędzy regułami pokrywającymi przykład testowy.
- Zaproponowanie statystycznej korekty na liczbę przykładów pozytywnych i negatywnych pokrywających indukowaną (tworzoną) regułę.

4. Uwagi o charakterze dyskusyjnym

Praca napisana jest zwięzłym językiem oraz cechuje się bardzo staranną stroną edytorską. Podczas czytania pracy nasuwają się pewne uwagi, nie wpływające jednakże na ogólnie wysoką ocenę pracy.

- W pracy brakuje informacji na temat jakości generowanych reguł, rozważanej z punktu widzenia ich konkluzji.
- Brak uzasadnienia wyboru tych, a nie innych miar jakości reguł jakie wykorzystano w pracy.
- Podczas fazy przycinania reguł usuwano całe warunki elementarne (chodzi o atrybuty numeryczne). Dlaczego nie próbowano również zmieniać zakresu warunków elementarnych ?
- Dlaczego podczas przycinania reguł jako alternatywną strategię przeszukiwania zbadano jedynie algorytm tabu, mimo że liczba warunków elementarnych w przesłankach reguł nie jest chyba zbyt duża i można było porównać ze strategią przeszukiwania wyczerpującego

5. Ocena końcowa rozprawy

Cele pracy zostały jasno sformułowane, a ich realizacja wymagała obszernych badań teoretyczno-eksperymentalnych. Cele te zostały osiągnięte.

Praca napisana jest przejrzysto i stanowi przemyślaną całość. Zawiera także wszystkie niezbędne elementy dobrej rozprawy naukowej: krytyczny przegląd literatury, wskazanie zagadnień wymagających rozwiązania, postawienie celów, propozycję nowych rozwiązań oraz badanie ich przydatności. Zawarte w niej rezultaty obejmujące są oryginalne i zostały przedstawione na wysokim, matematycznym poziomie.

Reasumując stwierdzam, że mgr inż. Adam Skowron wykazał się odpowiednią wiedzą z zakresu uczenia maszynowego, indukcji reguł, a także dobrym opanowaniem i posługiwaniem się warsztatem badawczym.

Uważam, że recenzowana praca doktorska mgr inż. Adama Skowrona w pełni spełnia wszystkie wymagania stawiane pracom doktorskim przez odpowiednią Ustawę i wnioskuję o jej przyjęcie i dopuszczenie do publicznej obrony.



K. Kowalski