



Prof. dr hab. Marcin Szpyrka
Akademia Górniczo-Hutnicza
Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej
Katedra Informatyki i Sztucznej Inteligencji
Al. Mickiewicza 30, 30-059 Kraków
tel.: 012 617 51 94
e-mail: mszpyrka@agh.edu.pl

Kraków, 11 sierpnia 2026 r.

Recenzja rozprawy doktorskiej mgr. inż. Adama Szczura pt. *Wybrane metody konstrukcji klasyfikatorów dla danych temporalnych*

Rozprawa doktorska została opracowana na Uniwersytecie Rzeszowskim w dyscyplinie Informatyka techniczna i telekomunikacja. Recenzja została opracowana na wniosek Rady Dyscyplin: Nauki fizyczne, Matematyka, Informatyka Techniczna i Telekomunikacja, Inżynieria materiałowa na Wydziale Nauk Ścisłych i Technicznych Uniwersytetu Rzeszowskiego.

1. Cel i zakres rozprawy

Rozprawa doktorska mgr. inż. Adama Szczura dotyczy metod klasyfikacji dla danych temporalnych, które reprezentowane są za pomocą temporalnych tablic decyzyjnych. W podejściu rozważanym w rozprawie obiekt klasyfikowany jest na podstawie sekwencji obserwacji zgromadzonych w określonym przedziale czasu. Obserwacje te grupowane są w sekwencje określone jako okna czasowe, w taki sposób, że całemu oknu przypisywana jest jedna klasa decyzyjna.

Klasyfikacja danych temporalnych nie jest nowym zagadnieniem w obszarze uczenia maszynowego. To co stanowi istotny nowy element, to podejście doktoranta do sposobu użycia danych temporalnych do uczenia modeli. Podstawowy cel jaki sobie postawił, to opracowanie metod umożliwiających aproksymację pojęć czasowych bez wcześniejszego przekształcania danych, np. wyliczania agregatów, co jest typowe dla wielu istniejących podejść. Drugi warunek jaki postawił sobie doktorant to zachowanie interpretowalności uzyskanych modeli. Podejście takie jest szczególnie cenne współcześnie, gdzie co raz częściej obserwujemy, że modele, w przypadku których możemy zrozumieć mechanizm podejmowania decyzji przez model, są wyżej cenione niż modele działające na zasadzie czarnej skrzynki. Biorąc pod uwagę, że część rzeczywistych danych, z którymi pracuje doktorant to dane medyczne, interpretowalność modeli ma szczególne znaczenie.

Rozwój istniejących metod uczenia maszynowego, czy też proponowanie nowych, które mogą pozwolić na lepszą lub porównywalną skuteczność, ale z zachowaniem wyższej interpretowalności modelu jest jak najbardziej pożądany. Tematykę poruszaną w pracy należy więc uznać za ważną i aktualną.

Tezy pracy zostały formalnie zdefiniowane jako „*Możliwe jest skonstruowanie interpretowalnych klasyfikatorów dla danych temporalnych, które działają bezpośrednio na danych źródłowych bez konieczności ich wcześniejszej agregacji, przy jednoczesnym zachowaniu skuteczności porównywalnej z klasycznymi metodami klasyfikacji oraz wybranymi metodami głębokiego uczenia.*”.

2. Struktura i zawartość rozprawy

Opiniowana rozprawa doktorska została przygotowana w języku polskim. Praca składa się z siedmiu rozdziałów wliczając w to wstęp i podsumowanie. Zawiera ponadto wykaz literatury liczący 122 numerowane pozycje, w tym 3, których współautorem jest doktorant. Praca liczy łącznie 128 stron.

- Rozdział 1 zawiera wprowadzenie do problematyki pracy, w szczególności zarysowano w nim problemy klasyfikacji danych temporalnych, motywację podjęcia badań, cel rozprawy, główne osiągnięcia i zakres pracy. W rozdziale tym jawnie zdefiniowano również tezy pracy.
- Rozdział 2 zawiera przegląd podstawowych pojęć, które są kluczowe dla zrozumienia dalszej części pracy. Rozdział rozpoczyna kilka definicji typowych dla ujęcia wiedzy z perspektywy zbiorów przybliżonych, w szczególności bazowe dla rozprawy pojęcia okna czasowego i temporalnej tablicy decyzyjnej. W rozdziale tym zawarto również podstawowe informacje nt. konstruowania drzew decyzyjnych oraz metod i miar oceny jakości klasyfikatorów.
- Rozdział 3 zawiera opis typowych współcześnie podejść do konstruowania klasyfikatorów dla pojęć czasowych. Zawarto w nim m.in. przegląd metod reprezentowania czasu w danych, różne sposoby definiowania okien czasowych, podejścia używane do opisywania wzorców czasowych, sposoby definiowania pojęć czasowych oraz metody aproksymacji pojęć czasowych.
- Rozdział 4 jest pierwszym z rozdziałów, które przedstawiają autorskie rozwiązania doktoranta. Opisano w nim metodę konstruowania drzewa decyzyjnego na podstawie danych temporalnych określanego jako „*drzewa decyzyjne z cięciami temporalnymi*” – TC-Tree. Rozdział rozpoczyna się od przedstawienia motywacji dla opracowania nowej metody. Następnie formalnie zdefiniowano cięcia temporalne i przedstawiono miarę oceny jakości takich cięć – warto podkreślić, że koncepcje te zilustrowano szczegółowo opisanymi przykładami, co ułatwia zrozumienie omawianej tematyki. W kolejnej części przedstawiono algorytmy wyznaczania cięć temporalnych i konstruowania drzewa TC. Rozdział 4 zawiera również wyniki eksperymentów obliczeniowych przeprowadzonych na rzeczywistych zbiorach danych, których celem było porównanie skuteczności klasyfikacji drzew z cięciami temporalnymi z wybranymi popularnymi metodami konstruowania klasyfikatorów (porównano 12 różnych metod z 4 wariantami drzewa TV – zależnie od sposobu dobierania parametru α przy konstruowaniu modelu).
- W rozdziale 5 przedstawiono zmodyfikowaną wersję modelu TC-Tree (VTC-Tree), która wykorzystuje dodatkowo tzw. weryfikujące cięcia temporalne do zwiększenia wiarygodności procesu klasyfikacji poprzez wykorzystanie dodatkowej wiedzy z alternatywnych cięć opartych na innych atrybutach warunkowych. Rozdział ma podobną strukturę jak poprzedni. Po uzasadnieniu podjęcia takich badań, opisano koncepcję cięć weryfikujących, miary oceny ich jakości (wszystkie pojęcia zilustrowane przykładami), algorytmy generowania cięć weryfikujących i konstruowania modelu VTC-Tree oraz eksperymentalne porównanie jakości klasyfikacji między modelem TC-Tree i VTC-Tree.
- Rozdział 6 stanowi dodatkowy wątek w ocenianej rozprawie doktorskiej. Dotyczy on analizy danych medycznych związanych z niewydolnością oddechową wcześniaków, które pozyskano z Oddziału Patologii i Intensywnej Terapii Noworodków Uniwersyteckiego Szpitala Dziecięcego w Krakowie. Problem badawczy dotyczący tych danych związany był z przewidywaniem zmian wzorców zachowania u wcześniaków, w szczególności pogorszenia się stanu zdrowia. W rozdziale opisano metodę przygotowania danych do uczenia modeli, a następnie wyniki eksperymentów obliczeniowych, w których porównano 9 różnych podejść (metody klasyczne) do poszukiwania najbardziej skutecznego klasyfikatora.

- Ostatni rozdział zawiera krótkie podsumowanie rozprawy wraz z zestawieniem wyników badań oraz pewne sugestie prac badawczych mających być jej kontynuacją.

Prezentacja materiału przedstawionego w pracy dokonana jest w sposób relatywnie czytelny i zrozumiały. Stosowana terminologia i symbole nie budzą w zasadzie zastrzeżeń. Drobne uwagi dotyczące składu i języka prezentacji zawarto w dalszej części recenzji.

3. Ocena rozprawy

Główne wyniki naukowe opisane w rozprawie doktorskiej dotyczą metod konstruowania klasyfikatorów (drzew decyzyjnych) na podstawie danych temporalnych. Doktorant przedstawił dwa podejścia do konstruowania takich drzew decyzyjnych, które określono jako TC-Tree i VTC-Tree. Punktem wyjścia dla doktoranta były metody reprezentowania wiedzy typowe dla teorii zbiorów przybliżonych. We wszystkich rozdziałach związanych z oryginalnymi wynikami badań (4–6) punktem wyjścia jest temporalna tablica decyzyjna. Istotną nowością w podejściu do tak przygotowanych danych jest brak „spłaszczenia”, tj. doktorant nie dokonuje transformacji okna czasowego na nowy wiersz w nowej płaskiej tablicy decyzyjnej (np. przez wykorzystanie metod statystycznych do obliczenia agregatów), ale konstruuje swoje klasyfikatory bezpośrednio na pierwotnej tablicy decyzyjnej. Pewnym uproszczeniem, które pozwoliło na taki zabieg, jest przyjęcie założenia, że decyzja jest taka sama dla wszystkich obserwacji z danego okna czasowego.

Dla tak przygotowanych danych zdefiniowano przede wszystkim pojęcie cięcia temporalnego, jako rozszerzenie koncepcji cięcia w konstruowaniu drzew decyzyjnych, oraz przedstawiono metodę oceny jakości takich cięć. Następnie przez analogię do metod konstruowania drzew decyzyjnych, przedstawiono algorytmy budowania modelu w postaci TC-Tree. Drugie z podejść (VTC-Tree) wydaje się być próbą odpowiedzi na naturalne pytanie: *Czy wybierając cięcie na danym etapie konstruowania drzewa można wybrać lepiej?* Model VTC-Tree to z mojej oceny próba pokazania potencjalnemu użytkownikowi tej metody, że zrobiono wszystko, aby zbudować optymalny model.

Ostatni podejmowany w rozprawie wątek dotyczy analizy danych medycznych związanych z niewydolnością oddechową wcześniaków. Wątek jest ciekawy zarówno w kwestii przygotowania danych, jak i porównania wyników eksperymentów obliczeniowych. Ciekawy wątek w tym temacie stanowi metoda określania wzorców zachowań. Pewien niedosyt budzi dość luźne powiązanie rozdziału 6 z dwoma poprzednimi.

Podsumowując ocenę merytoryczną pracy należy podkreślić, że uzyskane wyniki są istotne zarówno z teoretycznego, jak i praktycznego punktu widzenia, a także, że postawiona we wstępie rozprawy teza została udowodniona. Do najważniejszych osiągnięć doktoranta przedstawionych w ocenianej rozprawie doktorskiej należy zaliczyć:

1. Opracowanie koncepcji cięć temporalnych, metody oceny ich jakości oraz bazującej na nich metody konstruowania drzew decyzyjnych z cięciami temporalnymi – model TC-Tree.
2. Przedstawienie koncepcji cięć weryfikujących, metody oceny ich jakości oraz opracowanie bazującej na nich metody konstruowania modeli VTC-Tree.
3. Opracowanie metody identyfikacji i przewidywania zmian wzorców zachowania w danych medycznych związanych z niewydolnością oddechową wcześniaków.
4. Realizacja eksperymentów obliczeniowych na kilku zestawach rzeczywistych danych dotyczących porównania metod klasyfikacji i szczegółowa interpretacja uzyskanych wyników.

3.1. Uwagi

1. W definicji tablicy decyzyjnej (str. 15) atrybut decyzyjny $d \notin A$, tymczasem przy definicji temporalnej tablicy decyzyjnej (str. 19) atrybut decyzyjny $\tilde{d} \in \tilde{A}$. Skąd taka rozbieżność? Czy to ma jakiś sens merytoryczny?
2. Na str. 47 w definicjach wzorców temporalnych w licznikach używany jest symbol $\sqsubset$, a w mianownikach $\in$. Jaka jest semantyczna różnica między tymi dwoma oznaczeniami?
3. Na str. 64 pojawia się stwierdzenie: „Podsumowując, zaproponowana metoda TC-Tree osiąga wyniki porównywalne z klasycznymi metodami klasyfikacji, przy czym w wybranych przypadkach pozwala uzyskać jedne z lepszych rezultatów.” W tej części pracy porównywano jest z metodami, które często oferują podobną do TC-Tree interpretowalność modelu. Jakie cechy TCT w takim razie mogłyby przekonać użytkownika, że proponowany w pracy model TC-Tree jest lepszy niż modele o ustalonej renomie?
4. Na str. 70 pojawia się stwierdzenie: „Przy czym należy wyraźnie zaznaczyć, że dane okno czasowe pasuje jednocześnie do obu wzorców czasowych wtedy, gdy dokładnie te same punkty czasowe okna spełniają warunki dotyczące wartości atrybutów określone przez te wzorce.” Skąd takie ograniczenie? Dopasowanie do wzorca jest zdefiniowane jako spełnienie warunku nierówności (dla ustalenia uwagi powiedzmy, że chodzi o $\geq$). Bez trudu można wyobrazić sobie, że dla okien czasowych T_t^1 i T_t^2 i parametrów α_1 i α_2 spełnione są warunki dla liczby punktów $\geq \alpha_1$ i $\geq \alpha_2$ i nie są to same punkty.
Czy tym przypadkiem musi być $\alpha_1 = \alpha_2$
5. Na str. 101 wprowadzone zostały oznaczenia, które nie są później w ogóle wykorzystywane (atrybuty sensoryczne, atrybuty eksperckie). Jaki jest cel wyróżniania tych pozbiorów atrybutów? Wydaje się, że można było zachować spójność w oznaczeniach z wcześniejszymi rozdziałami pracy.
6. Dane przedstawione w rozdziale 6 mają charakter temporalny. Wydaje się, że można było pokusić się o zastosowanie któregoś z modeli TC-Tree lub VTC-Tree do tych danych. Takie pokazanie użyteczności wprowadzonej metody byłoby świetnym podsumowaniem całej pracy doktorskiej. Dlaczego nie spróbowano użycia nowych metod do tych danych?

3.2. Uwagi techniczno-redakcyjne

Pod względem edytorskim oceniana rozprawa doktorska lokuje się zdecydowanie powyżej przeciętnej. Została dość starannie złożona z użyciem systemu składu L^AT_EX. Tym niemniej autor nie ustrzegł się kilku „typowych” błędów, np.: brak polecenia `\mathit{}` dla wieloliterowych oznaczeń w trybie matematycznym, co skutkowało zbędnymi odstępami między literami, używanie dywizu zamiast półpauzy, a także pozostawienie licznych sierot typograficznych.

4. Wniosek końcowy

Przedstawiona w pracy problematyka dotyczy aktualnych i interesujących zagadnień naukowych związanych z problemem konstruowania klasyfikatorów dla danych temporalnych. Rozprawa doktorska zawiera szereg interesujących oryginalnych wyników pracy badawczej autora. Na uwagę zasługują zarówno aspekty teoretyczne (definicje, algorytmy), jak i eksperymenty obliczeniowe opracowane i zrealizowane przez doktoranta. Mimo przedstawionych wcześniej drobnych uwag krytycznych, rozprawę doktorską jako całość oceniam zdecydowanie pozytywnie.

Stwierdzam, że autor wykazał się ogólną wiedzą teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja, trafnie sformułował problem naukowy i następnie przedstawił jego oryginalne rozwiązanie. Podsumowując, stwierdzam, że recenzowana rozprawa doktorska pt. *Wybrane metody konstrukcji klasyfikatorów dla danych temporalnych* spełnia wymagania stawiane rozprawom doktorskim przez ustawę *Prawo o Szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 r. (z późniejszymi zmianami) i wnoszę o dopuszczenie mgr. inż. Adama Szczura do publicznej obrony.