



**POLITECHNIKA
RZESZOWSKA**
im. IGNACEGO ŁUKASIEWICZA

al. Powstańców Warszawy 12, 35-959 Rzeszów
tel./fax: +48 17 854 12 60, tel.: +48 17 865 11 00
www.prz.edu.pl

dr hab. inż. Dominik Strzałka, prof. PRz

Rzeszów, dn. 31.01.2025 r.

Politechnika Rzeszowska

Wydział Elektrotechniki i Informatyki

Zakład Systemów Złożonych

Tytuł rozprawy: Identyfikacja anomalii w dziedzinowych zbiorach danych złożonych

Autorka rozprawy: mgr inż. Czesław Horyń

Promotor rozprawy: dr hab. Agnieszka Nowak-Brzezińska, prof. UŚ

Dziedzina: nauki ścisłe i przyrodnicze

Dyscyplina: informatyka



ZAKŁAD SYSTEMÓW ZŁOŻONYCH

Wydział Elektrotechniki i Informatyki
ul. Wincentego Pola 2, 35-959 Rzeszów, tel. 17 865 1340
zsz.prz.edu.pl



Syntetyczne przedstawienie rozprawy

Cel pracy

Zasadniczym celem pracy jest opracowanie zestawu metod i narzędzi, które zwiększą czułość i efektywność wykrywania anomalii w złożonych zbiorach danych, z szczególnym naciskiem na zastosowania praktyczne i optymalizację algorytmu Local Outlier Factor (LOF).

W szczególności autor pracy założył następujące cele szczegółowe:

- zdefiniowanie pojęcia anomalii oraz identyfikację wyzwań związanych z ich wykrywaniem w złożonych, kategoriowych danych wielowymiarowych,
- zastosowanie i ocenę nowoczesnych technologii, takich jak uczenie maszynowe i głębokie uczenie, w kontekście identyfikacji anomalii, z naciskiem na zwiększenie czułości,
- rozwój i optymalizację zaawansowanych technik zespołowych, w celu zwiększenia skuteczności i wydajności procesu wykrywania anomalii,
- implementację i testowanie proponowanych metod na rzeczywistych zbiorach danych, aby ocenić ich efektywność, czułość oraz użyteczność w praktycznych zastosowaniach.

Zakres pracy

Rozprawa doktorska jest bardzo obszerna. Cały tekst składa się z niemal 400-u stron opracowania: tekst rozdziałów, dodatki, literatura, różnego rodzaju spisy. W pracy można wyodrębnić dwie zasadnicze części objętościowo zbliżone do siebie: pierwsza (nieco dłuższa) obejmująca głównie treść rozdziałów 1-6, prezentuje teoretyczną analizę najważniejszych aspektów problemu wykrywania anomalii (lub wartości odstających – określenia te są stosowane zamiennie) w danych złożonych, istotnych dla dalszych badań naukowych. Nie jest to jednak tylko i wyłącznie przywołanie definicji znanych z literatury ponieważ autor proponuje także i własną (str. 12) jako podstawę do rozważań w odniesieniu do analizy i metody przetwarzania danych kategoriowych i wielowymiarowych. Natomiast w drugiej części pracy (nieco krótszej, ale moim zdaniem będącej najważniejszym elementem rozprawy) położono nacisk na wykrywanie wartości odstających (anomalii) w różnych danych, z uwzględnieniem





uczenia maszynowego i uczenia głębokiego. Obejmuje ona nie tylko opis projektu systemu Trinity SALT i przykłady praktyczne dotyczące zastosowania nowoczesnych technik uczenia maszynowego dla autorskiego rozwiązania, ale także zaawansowane techniki zespołowe, które zwiększają efektywność identyfikacji anomalii. Czytelnik może zapoznać się z przeglądem i porównaniem miar oceny skuteczności metod wykrywania anomalii zarówno dla proponowanego systemu Trinity SALT, który integruje w sobie algorytmy Sieci Samoorganizujących (SOM), Autoenkoderów (AE) i Local Outlier Factor (LOF) jak i w odniesieniu do samodzielnej pracy powyższych algorytmów. Skuteczna okazuje się także autorska technika Maksymalnej Znormalizowanej Agregacji (MNA). Ta część prac jest bardzo intensywnie zobrazowana wieloma przykładami, uzyskiwanymi wynikami oraz ich pogłębioną interpretacją.

Charakter rozprawy

Praca ma charakter teoretyczno-eksperymentalny z mocno wyeksponowanym elementem praktycznym w postaci projektu systemu Trinity SALT, który jest szczegółowo opisany w rozdziale 7. Jest to rozwiązanie oparte o aplikację webową z odpowiednimi interaktywnymi panelami nawigacji do wizualizacji i analizy danych w czasie rzeczywistym. Narzędzie służy do detekcji anomalii. Aplikacja jest także w wersji umożliwiającej jej uruchomienie na własnym komputerze. Zaimplementowano w niej algorytmy SOM, Autoenkoder oraz LOF a także autorską technikę maksymalnej znormalizowanej agregacji MNA (ang. *maximum normalized aggregation*, MNA). System Trinity SALT oblicza wynik końcowy dla każdego przetwarzanego obiektu jako suma maksymalnej wartości znormalizowanych wyników z algorytmów SOM, AE i LOF oraz bonusu wg zależności (7.1). W pracy są także wyniki licznych symulacji komputerowych dla zaproponowanych eksperymentów również na bazie autorskiego oprogramowania.





Opinia:

- o poprawności o oryginalności postawionej tezy i stopniu w jakim została ona wykazana.

W pracy postawiono następującą tezę:

Zastosowanie zaawansowanych technik, takich jak SOM, LOF z podziałem na bloki oraz autoenkodery w ramach zespołu algorytmów, umożliwi zwiększenie czułości i wydajności wykrywania anomalii w rzeczywistych, złożonych zbiorach danych, przy jednoczesnym zachowaniu wydajności procesu analizy, szczególnie dzięki optymalizacji LOF.

Postawiona teza jest poprawna i ma oryginalny charakter w odniesieniu do:

(i) eksperymentalnego charakteru wielu fragmentów doktoratu, w których autor, w bardzo przekonujący sposób z punktu widzenia uzyskiwanych wyników dostarcza odpowiednich rozwiązań i narzędzi (oprogramowanie) dla skutecznego radzenia sobie z rzeczywistymi problemami,

(ii) bardzo szerokiego spektrum potencjalnych praktycznych zastosowań zaproponowanego podejścia w odniesieniu np. do systemów wykrywania i zapobiegania włamaniom, wykrywania oszustw z wykorzystaniem kart kredytowych, diagnoz medycznych, nietypowe zdarzenia czujników,

(iii) analizy i interpretacji otrzymanych wyników dla przetwarzanych konkretnych różnorodnych zbiorów danych m.in. takich jak: pomiary fizykochemiczne win, roszczenia dotyczące pojazdów, końcowe pozycje w szachach, dane medyczne przypadków Covid-19, klasyfikacja grzybów, transakcje kartami kredytowymi, oraz badania mutacji białka p53,

(iv) wykonanej analizy porównawczej wydajności zaproponowanego rozwiązania.

Treść pracy oraz wykonane eksperymenty i analizy jednoznacznie przekonują, że teza została udowodniona, a opracowany system Trinity SALT (SOM-AE-LOF-TriDetect) zapewnia przekonujące wyniki w różnych zbiorach danych.





- o analizie źródeł:

Auto pracy wykonał bardzo duży wysiłek, aby swoje rozważania bardzo mocno osadzić w kontekście dostępnych źródeł literatury. Zgodnie ze spisem jest ich 436 i odnoszą się one do bardzo szerokiego kontekstu omawianych w pracy zagadnień, zarówno tych teoretycznych, jak i praktycznych oraz w odniesieniu do interpretacji wyników. W tym zakresie praca prezentuje się wzorowo przekonując czytelnika o dużej dbałości w zakresie studium literatury, pokazuje bardzo dobrą znajomość literatury tematu i przekonuje, że analiza źródeł literaturowych jest zrobiona bardzo sumiennie. Jest to bardzo dobrze widoczne na przykładzie rozdziałów 1-6, które można także potraktować jako przeglądowe kompendium wiedzy w zakresie identyfikacji anomalii, uczenia maszynowego, technik zespołowych oraz skuteczności identyfikacji anomalii.

W recenzowanej rozprawie autor powołał się także na 9 swoich (głównie współautorskich z Panią promotor) prac, są to poz. [82, 84, 85, 101, 102, 264, 286, 304, 305].

- o pozycji rozprawy w stosunku do stanu wiedzy i/lub stanu techniki reprezentowanych przez literaturę światową

Recenzowana rozprawa doktorska ze względu na bardzo szeroki zakres poruszanych zagadnień oraz prace eksperymentalne z danymi i przetwarzającymi je algorytmami stanowi ważny wkład do stanu techniki i wiedzy reprezentowanego przez literaturę światową. Biorąc pod uwagę stale aktualny kontekst detekcji anomalii w różnego rodzaju danych, skuteczności ich wykrywania, powszechne zastosowania uczenia maszynowego oraz potencjalne możliwości zastosowań autor rozprawy przedstawił nie tylko szeroki konteksty swoich rozważań, ale także dostarczył narzędzie dostępne na stronie: <http://trinitysalt.pl:8000/>. Zaproponowany system Trinity SALT łączy wyniki pracy trzech algorytmów i znacząco podnosi skuteczność identyfikacji anomalii. Zaś autorska technika maksymalnej znormalizowanej agregacji MNA jest rozwiązaniem, które wyróżnia koncepcję autora dodatkowo wzmacniając dokładność wykrywania anomalii umożliwiając generowanie list odstających obiektów oraz wskazując te, które są uznawane za anomalie przez algorytmy składowe. Zazwyczaj w literaturze albo spotyka się przeglądowe opracowania pokazujące np. jak można wykrywać anomalie, albo opracowania dotyczące





skuteczności poszczególnych podejść – tu widać skuteczną próbę integrowania różnych podejść oraz praktyczne opracowanie, które pomysł realizuje.

Ponadto autor rozprawy wskazał, że w trakcie jej realizacji przygotował kilka propozycji własnych artykułów naukowych. Praca pt. *Automatic block size optimization in the LOF algorithm for efficient anomaly detection* została opublikowana w czasopiśmie Applied Soft Computing, Volume 170, February 2025, 112675. Czasopismo posiada IF>7.

- o znaczeniu uzyskanych wyników dla danej dyscypliny naukowej

Bardzo znaczącym elementem rozprawy, oprócz przedstawionych wyników eksperymentów, jest dostarczony system Trinity SATL dostępny na stronie <http://trinitysalt.pl:8000/>. Rozprawa zawiera przykłady pokazujące, efekty użycia tego narzędzia (Rozdziały 7 i 8). Z punktu widzenia dyscypliny naukowej informatyka jest to wkład znaczący, bowiem zaprojektowany i wykonany system działa w oczekiwany sposób zwracając wyniki analiz, które są prezentowane w rozprawie. Ponadto autor zaproponował kilka istotnych (np. z punktu widzenia ograniczenia złożoności obliczeniowej) usprawnień dla istniejących algorytmów co wpłynęło pozytywnie na ogólną wydajność przetwarzania: w przypadku algorytmu LOF wykorzystano zaproponowany podział na bloki, wprowadzono dynamiczną optymalizację rozmiaru bloku, poprawiono skuteczność wykrywania anomalii.

- o umiejętności autora do poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników.

Od strony technicznej i sposobu prezentacji uzyskanych wyników praca jest przygotowana bardzo starannie i w zasadzie nie ma w jej treści znaczących usterek technicznych czy też językowych. Mając na uwadze, że powstało około 400 stron opracowania, to w zasadzie jest to dokument wykonany bardzo wzorowo, z dużą dbałością o szczegóły, aspekty techniczne, formatowanie, wyróżnienie określonych, istotnych fragmentów. Rozprawę od tej strony czyta się bardzo dobrze i bez cienia wątpliwości autor wykazał się tutaj bardzo dużym wysiłkiem. Drobnym mankamentem jest to, że w pracy pewna grupa rysunków jest dość niskiej rozdzielczości lub przyjęta kolorystyka oznaczeń nieco utrudnia lepsze zrozumienie



uzyskanych wyników i można było nieco skupić się na ich lepszym wyeksponowaniu. To nie jest znaczący mankament, ale trochę obniża bardzo pozytywny odbiór.

Generalnie autor posiada umiejętność poprawnego i przekonującego prezentowania otrzymanych wyników bowiem są one nie tylko widoczne w postaci danych tabelarycznych ale i rysunków. Jest to także uzupełnione o odpowiednie komentarze, analizy, pogłębione wnioski.

Główne wady rozprawy i jej słabe strony.

Jak już wspomniano praca jest zdecydowanie za długa prezentując zbyt wiele szczegółów, które można znaleźć w literaturze i to wydaje się być jej najważniejszą słabą stroną. Przejawia się to m.in. w tym, że część pierwsza zdecydowanie bardziej odnosząca się do aktualnego stanu wiedzy, zajmuje ponad 200 stron, natomiast druga, pokazująca wkład autora, jest trochę skromniejsza i w kilku fragmentach (o czym nadmienię w uwagach szczegółowych) brakuje lepszych uzasadnień i wyjaśnień.

W pracy nie ma bardzo poważnych usterek technicznych, autor rozprawy używa poprawnej polszczyzny, zadbał o niezbędne odwołania, starannie sformatowano tekst.

W rozdziale 7, a najlepiej w dodatku, trochę brakuje lepszej instrukcji postępowania z danymi w Trinity SALT. Taka instrukcja mogłaby też być na stronie tej aplikacji np. zakładka pomoc lub w postaci tutoriala video.

Nie jest też wiadome, jak od strony cybernetycznej zabezpieczono webową wersję systemu Trinity SALT.

Na części prezentowanych wykresów, zwłaszcza w rozdziale 8 można było zastosować nieco większą czcionkę opisów i lepszą kolorystyką, co lepiej uwidoczniliby uzyskiwane wyniki.

Uwagi szczegółowe

Uwagi szczegółowe zostaną wyrażone w postaci konkretnych pytań do autora rozprawy doktorskiej. Nie mają one charakteru krytycznego, ale są związane z chęcią uszczegółowienia pewnych kwestii oraz wyjaśnienia drobnych nieścisłości.





1. Jakie jest teoretyczne uzasadnienie podejścia zaprezentowanego w rozdziale 7.1, w szczególności wzoru (7.1) – podejście MNA, gdzie pojawia się bonus B przyznawany w zależności od liczby algorytmów, które zidentyfikowały obiekt? Co było przyczyną wyboru tej metody w kontekście opisów z rozdziału 5.5?
2. Czy możliwe byłoby zastosowanie innego podejścia zamiast maksymalnej normalizowanej agregacji (MNA) w zakresie premiowania konsensusu modeli SOM, AE oraz LOF i jakie potencjalnie mogłyby być tego wyniki? Aktualnie jest wybierana wartość max znormalizowanych wyników z algorytmów SOM, AE oraz LOF dla badanego obiektu, co wydaje się być uzasadnione np. z kontekstu rozdziału 5, ale czy można sobie wyobrazić w podejściu autora inne techniki.
3. Wartości bonusów b_i są wskazane bez uzasadnienia i zakładają odniesienie się do liczby metod wykrywających anomalie, ale nie do tego, które metody tego dokonały. Jest tutaj w zasadzie ciche założenie, że każda z tych metod jest taka sama pod wieloma względami (np. swojej skuteczności), ale z opisów w pracy jest wiadome, że tak nie jest (rozdział 4.5). Skąd biorą się takie a nie inne wartości bonusów $b_1 - b_3$? Czy te wartości mogą się zmieniać, np. w zależności od analizowanego problemu, rodzaju danych wejściowych?
4. Czy jest rozważane skomercjalizowanie zaproponowanego rozwiązania?

Wniosek końcowy

Po analizie przedłożonej rozprawy doktorskiej Pana mgra Czesława Horynia stwierdzam, że wnosi ona ważny wkład w rozwój dyscypliny informatyka. Jej wybrane fragmenty, a w szczególności autorska propozycja przygotowanego systemu Trinity SALT w połączeniu z metodą MNA są unikalne w zakresie różnego rodzaju podejść do wykrywania anomalii. Praca ma bardzo duży potencjał aplikacyjny i mogłaby być podstawą do przygotowania rozwiązania komercyjnego. Treść pracy, przygotowane eksperymenty, zrealizowane badania, sposób prezentacji wyników i nade wszystko bardzo szeroka prezentacja zagadnień teoretycznych jednoznacznie potwierdzają wiedzę teoretyczną, bardzo dobre przygotowanie merytoryczne, umiejętności praktyczne (w tym realizacji eksperymentów) i umiejętność samodzielnego





prowadzenia badań naukowych przez Doktoranta. Podjęta tematyka jest ważna, ciągle aktualna i posiada liczne zastosowania praktyczne. Recenzowana rozprawa doktorska spełnia ustawowe wymagania stawiane kandydatom do stopnia doktora, określone ustawą o stopniach i tytułach naukowych – uwzględnione w art. 13 ust. 1 Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki w związku z art. 179 Ustawy z dnia 3 lipca 2018 r. Przepisy wprowadzające ustawę - Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r., poz. 1669). Na tej podstawie wnioskuję do Rady Dyscypliny Informatyka Instytutu Informatyki Uniwersytetu Śląskiego o dopuszczenie jej do publicznej obrony.

