

Gliwice, 08.05.2024

Prof. dr hab. inż. Marcin Woźniak
Wydział Matematyki Stosowanej
Politechnika Śląska

Sz. P.
dr hab. inż. Adam Kulawik, Prof. PCz.
Kierownik
Rady Dyscypliny Naukowej
Informatyka Techniczna i
Telekomunikacja
Wydział Informatyki i Sztucznej Inteligencji
Politechnika Częstochowska

Recenzja rozprawy doktorskiej mgr inż. Krzysztofa Ropiaka pt.: „Adaptation of mereological granulation techniques in selected data analysis issues” opracowanej pod kierunkiem dr hab. Piotra Artiemjewa, Prof. UWM.

W odpowiedzi na pismo R-WIMil-BOD-510-7/21 z dnia 24.02.2025 przedkładam niniejszą recenzję.

Przedstawiona do recenzji praca została napisana w języku angielskim i składa się z 5 głównych rozdziałów obejmujących w sumie 172 strony. Praca przedstawia temat modeli granulacji danych oraz osiągnięcia Doktoranta w tym zakresie. Poszczególne części pracy doktorskiej pokazują streszczenie, merytoryczną część opracowania, spis rysunków, spis tabel oraz bibliografię, która składa się 51 pozycji. Załączone na końcu pracy spisy cytują 49 rysunków oraz 64 tabele i 2 algorytmy.

W pierwszym rozdziale Doktorant opisuje teoretyczne podstawy granulacji danych w zależności od konfiguracji modelu analizy. W rozdziale zostały opisane podstawowe aksjomaty mereologii Leśniewskiego. Wprowadzone zostaje również pojęcie „rough set” oraz aksjomatyka działań na zbiorach przybliżonych.

W rozdziale drugim Doktorant przedstawia motywację do podjęcia tematu komputerowych metod granulacji danych. Przedstawione zostały pojęcia różnych

podejść do granulacji danych typu: standardowego, koncepcyjnego oraz warstwowego. Autor przedstawił działanie wymienionych metod na 11 podstawowych zbiorach danych. Wyniki obliczeń zostały opisane w sposób graficzny i tabelaryczny. Modele granulacji zostały porównane w testach wydajnościowych ze standardowymi wersjami innych technik analitycznych typu knn, logistic regression, bayesian classifier, support vector machine, decision tree lub random forest.

Rozdział trzeci przedstawia metodę granulacji jednorodnej. W zaproponowanym podejściu Autor uzależnia dopasowanie danych od ich zbieżności względem informacji jaką zawierają. Dla zaproponowanej metody przedstawione zostały testy wydajnościowe na opisanych wcześniej zbiorach danych. Doktorant opisał modyfikację metody jednorodnej uzależniając model granulacji od zastosowanego podejścia normalizacji danych wejściowych. W modelu oceny została zastosowana metryka Hamminga. Dla zaproponowanej modyfikacji metody jednorodnej przedstawione zostały testy wydajnościowe na zbiorze UCI Repository. Doktorant przedstawił również zbiorczy model granulacji danych, w którym porównuje wyniki metod: random granules, arcing, ada-boost and bagging. Wyniki testów wydajnościowych zostały przedstawione dla zbioru Australian credit data set.

W rozdziale czwartym Doktorat podjął temat tzw. oversamplingu oraz undersamplingu, które w znaczący sposób mogą wpływać na finalną wydajność klasyfikatora. Rozdział przedstawia zasady analizy danych niekompletnych poprzez zastosowanie modelu granulacji. Do przedstawienia testów i analizy potencjalnych zależności zastosowano modelowanie metodą SMOTE.

W kolejnych częściach pracy Doktorant porównuje przedstawione metody i przedstawia przykłady ich zastosowania na wymienionych wcześniej zbiorach danych, aby w rezultacie postawić wnioski z wykonanych eksperymentów obliczeniowych.

Ocena merytoryczna przedstawionej rozprawy doktorskiej:

Przedłożona rozprawa doktorska mgr inż. Krzysztofa Ropiaka pokazuje temat analizy danych, który jest aktualny zwłaszcza w dzisiejszym świecie modeli numerycznych, które analizują ogromne zestawy danych w różnych zastosowaniach decyzyjnych.

Doktorant modyfikuje znane podejścia do wyznaczania istotnych fragmentów danych i stara się pokazać praktyczne aspekty ich zastosowań.

Praca napisana jest poprawnie i zwięźle, choć w kilku miejscach brakuje należytej staranności w naukowym opisie badań i analizy aspektów teoretycznych stosowanych modeli, zwłaszcza pod względem złożoności obliczeniowej i wydajności metod na dużych zbiorach danych. Praca została napisana w języku

angielskim. Dzięki temu uniknięto konieczności tłumaczenia nazw i terminów.

Po przeczytaniu przedstawionej rozprawy doktorskiej nasuwają się następujące pytania i sugestie:

- i. Jaka jest efektywność zaproponowanych aproksymacji systemów decyzyjnych vs inne techniki z tej samej rodziny modeli granulacji? Nie przedstawiono analizy porównawczej ani statystycznej wyników badań.
- ii. Czy efektywność badanych modeli da się zwiększyć poprzez zastosowanie parametryzacji lub zdefiniowanie współczynnika skalującego w modelach granulacji? Czy jest możliwość znalezienia optymalnych parametrów granulacji? W jaki sposób efektywność proponowanej metody granulacji zależy od schematu budowy danych wejściowych?
- iii. Jakie jest podejście najefektywniejsze do granulacji danych różnego typu? Czy dane numeryczne i symboliczne granulujemy w ten sam sposób? Czy były prowadzone badania aby takie metody zidentyfikować, a dopiero na ich podstawie opracować ulepszony model?
- iv. Czy Doktorant badał proponowane metody na tzw. dużych zbiorach danych? Jak zaproponowane rozwiązania działałyby na zbiorach danych, gdzie klasy abstrakcji nie są całkowicie rozłączne?
- v. Jak Doktorant rozumie hybrydowość zaproponowanego rozwiązania? Czy model te można stosować w kontekście klasyfikatorów złożonych?

Ocena formy, taksonomii, języka i edycji pracy:

Przedstawiona rozprawa doktorska jest napisana poprawnym językiem. Edycja i forma dokumentu są poprawne. Rozprawę napisano w języku angielskim. Natomiast moim zdaniem na jej końcu powinno zostać dołączone streszczenie w języku polskim. Przedstawione rysunki, tabele i wzory zostały ponumerowane stosownie do przynależności do rozdziałów opracowania.

Ocena przedstawionego dorobku naukowego:

Doktorant jest współautorem kilku artykułów opublikowanych zarówno w czasopiśmie oraz wolumenach międzynarodowych konferencji. Większość jego prac skupia się na tematyce związanej z przedstawioną rozprawą doktorską.

Przedstawiony dorobek naukowy Doktoranta pokazuje zaangażowanie w prace naukowe, które zostały uznane w międzynarodowym środowisku naukowym. W przedstawionym dorobku dostrzegam kilka prac, które powinny zostać zacytowane we właściwy sposób pokazując należną kolejność autorów. Natomiast obecne cytowanie, choć niewłaściwe, traktuję jako omyłkę pisarską. Moim zdaniem zaprezentowany dorobek pozwala stwierdzić obeznanie Doktoranta z dyscypliną prowadzenia badań naukowych, analizą wyników i podejmowaniem wniosków z prowadzonych badań naukowych.

Podsumowanie:

Pismem R-WIMil-BOD-510-7/21 z dnia 24.02.2025 otrzymałem z Politechniki Częstochowskiej informację o wpłynięciu poprawionej wersji pracy doktorskiej, którą otrzymałem do recenzji. Przedstawiona do recenzji rozprawa doktorska porusza temat modelowania algorytmów augmentacji i analizy danych. Jest to aktualny temat z punktu widzenia rozwoju nowoczesnej informatyki. Doktorant przedstawił ulepszoną wersję dla algorytmu granulacji danych numerycznych wraz z wynikami swoich badań.

Na podstawie przedstawionej powyżej oceny stwierdzam, iż rozprawa doktorska mgr inż. Krzysztofa Ropiaka pt.: „Adaptation of mereological granulation techniques in selected data analysis issues” spełnia warunki określone w art. 183 oraz art. 190 pkt 2 i 3 Ustawy z dnia 20 lipca 2018r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2023 roku poz 742, ze zm.) stosowane w postępowaniach o nadanie stopnia doktora. Wnoszę o przyjęcie w/w rozprawy doktorskiej i dopuszczenie jej do publicznej obrony w dyscyplinie informatyka.