



Warszawa, dnia 30 marca 2019 roku

Prof. dr hab. Leon Bobrowski
Wydział Informatyki Politechniki Białostockiej
ul. Wiejska 45A, 15-351 Białystok

Recenzja

pracy doktorskiej mgr inż. Marcina Pietrzykowskiego
na temat „*Lokalne uczenie algorytmów regresyjnych metodą mini – modeli*”

Recenzja została opracowana na podstawie pisma Dziekana Wydziału Informatyki Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie dr hab. inż. Jerzego Pejaśia z dnia 24 stycznia 2019 roku. Zgodnie z tym pismem recenzja została przygotowana w trybie warunków określonych w Art. 1 Ustawy z dnia 14 marca 2003 r. przepisów ustawy o stopniach naukowych i tytule naukowym oraz o stopniach tytule w zakresie sztuki (z późniejszymi zmianami).

Praca doktorska mgr inż. Marcina Pietrzykowskiego została złożona w 2019 roku na Wydziału Informatyki Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie. Promotorem pracy jest prof. dr inż. Andrzej Piegat, a promotorem pomocniczym jest dr inż. Marcin Pluciński. Rozprawa doktorska ma objętość 190 stron i została podzielona na 5 rozdziałów. *Rozdział 1* został zatytułowany *Przegląd metod i zagadnień wykorzystanych w pracy*. *Rozdział drugi* ma tytuł *Bryły geometryczne w przestrzeni wielowymiarowej*. Centralną pozycję w pracy zajmuje *Rozdział 3* o objętości 34 stron, zatytułowany *Metoda mini-modeli*. W rozdziale tym opisana została koncepcja *mini-modeli*, która stanowi podstawową oś merytoryczną przedstawionej rozprawy. W *Rozdziale 4* opisane zostały wyniki numerycznych eksperymentów porównawczych z różnego typu mini-modelami. *Rozdział 5* został zatytułowany *Podsumowanie pracy*.

Cel rozprawy został sformułowany jako (s. viii):

„udoskonalenie metody uczenia mini-modeli na bazie próbek lokalnych w przestrzeni wielowymiarowej”

Hipoteza badawcza sformułowana została w poniższej postaci (s. ix):

„udoskonalenie metody uczenia mini-modeli umożliwi wykazanie co najmniej części przewidywanych zalet tej metody względem geometrycznych metod modelowania bazujących na próbkach”

Doktorant przyjął poniższe założenie na początku pracy:

„Hipoteza badawcza zostanie uzasadniona empirycznie poprzez przedstawienie wyników eksperymentów oraz ich analizę”

Rozdział 1 zawiera opisy wybranych metod regresyjnych i klasyfikacyjnych stosowanych w analizie danych wielowymiarowych. Doktorant zamieścił tu opisy podstawowych właściwości powszechnie używanych metod i algorytmów takich jak modele *regresji*

liniowej, metoda wektorów wspierających (ang. *support vector machines* - *SVM*), algorytm *K* najbliższych sąsiadów (*K* - *NN*), algorytm grupowania *K* - średnich (*K* - *means*). Doktorant zamieścił tu też krótkie opisy wybranych modeli sieci neuropodobnych w tym model *Perceptronu* i model *sieci radialnych* (*RBF*).

W *Rozdziale 2* znalazły się opisy brył geometrycznych w przestrzeni wielowymiarowej wybranych za względu na używanie ich przez doktoranta w mini-modelach. Umieszczone zostały tu opisy trzech rodzajów brył geometrycznych: *simpleks*, *hipersześcian* i *hiperośmiościan*. W rozdziale tym zostało sformułowane i udowodnione twierdzenie na temat *simpleksu dualnego*.

Koncepcja *mini-modeli* (*MM*) jest definiowana i analizowana w *Rozdziale 3* rozprawy. Na początku tego rozdziału znajduje się stwierdzenie: „*Podstawowa koncepcja metody mini modeli została opracowana przez profesora Andrzeja Piegata [58, 59]. Motywacją do opracowania metody była chęć ulepszenia algorytmu najbliższych sąsiadów*” (s. 54). Zgodnie z *Bibliografią* rozprawy, prace cytowane w powyższym stwierdzeniu zostały opublikowane w 2011 roku w *Proceedings of 9th Conference on Information Systems in Management* oraz w *Studia Informatica*.

Zgodnie z określeniem doktoranta (s. 55):

„*Metoda MM jest metodą modelowania lokalnego i operuje tylko i wyłącznie na danych znajdujących się w lokalnym otoczeniu punktu zapytania, które jest ściśle określone i nie obejmuje całej modelowanej dziedziny tak jak w przypadku modeli globalnych. Lokalne otoczenie punktu zapytania nazywamy obszarem mini-modelu lub domeną mini-modelu.*”

Używając terminologii rozpoznawania obrazów (ang. *pattern recognition*), zamiast wyrażenia *punkt zapytania* użylibyśmy raczej określenia *aktualny wektor cech*, lub zbiór aktualnych wektorów cech. W przedstawionej rozprawie domeny mini modeli regresyjnych definiowane są głównie za pomocą wypukłych wielościanów w wielowymiarowych przestrzeniach cech. Generalnie, użycie wypukłych wielościanów w definiowaniu mini domen modeli regresyjnych skutkuje określonym sposobem określania sąsiedztwa aktualnego wektora cech innym niż w klasycznych metodach minimalno odległościowych *K-NN*.

Wyniki eksperymentów porównawczych zostały opisane w *Rozdziale 4*, który jest największym rozdziałem rozprawy o objętości 53 stron. Eksperymenty obliczeniowe opisane w tym rozdziale były na ogół wcześniej prezentowane na międzynarodowych konferencjach specjalistycznych i publikowane m.in. w książkach wydawnictwa *Springer*.

Algorytmy regresyjne i klasyfikacyjne wynikające z mini-modeli były porównywane w eksperymentach obliczeniowych m.in. z algorytmami opartymi na modelach radialnych sieci neuropodobnych oraz metodach *K-NN* i *SVM*. Procedury obliczeniowe metody mini-modeli zostały opracowane i implementowane przez doktoranta. W eksperymentach porównawczych metody mini-modeli z wymienionymi powyżej klasycznymi metodami modelowania regresyjnego i klasyfikacyjnego użyte zostały procedury obliczeniowe implementowane w środowisku *MATLAB*. Zgodnie z deklaracją autora parametry tych procedur były optymalizowane, choć nie zostały podane szczegóły takich optymalizacji. Podana została też informacja, że do oceny modeli użyto średnich błędów absolutnych (s. 94).

Oceny porównawcze metody mini-modeli zostały przeprowadzone przez doktoranta głównie na zbiorach danych w dwuwymiarowej przestrzeni cech (2D). Działanie mini-modeli 2D zostało porównane z poniższymi metodami (s. 69):

1. algorytm k -najbliższych sąsiadów,
2. aproksymacja wielomianową stopnia n ,
3. jednowarstwowa, jednokierunkowa sieć neuronowa (MLP)
4. sieć neuronowa typu GRNN (*General Regression Neuron Network*)

Obliczenia w ramach oceny metody mini-modeli przeprowadzane były często przy użyciu zbiorów danych reprezentowanych we współrzędnych sferycznych. Jakość modeli oceniana była przy zastosowaniu techniki ocen krzyżowych (ang. *cross-validation*) co jest zalecane ze względu na redukcję dodatniego obciążenia ocen.

Na początku podrozdziału 4.5 doktorant deklaruje (s. 137): „*Pomimo, że celem niniejszej rozprawy jest zastosowanie mini-modeli jako metody uczenia algorytmów regresyjnych, warto zaprezentować możliwość zastosowania opisywanej metody w zadaniu klasyfikacji*”. W podrozdziale tym zawarte zostały opisy porównawczych eksperymentów klasyfikacyjnych z trzema zbiorami danych z repozytorium *UCI Machine Learning: Iris flower, Wisconsin Breast Cancer, oraz Pima Indians Diabetes*. Doktorant wykonał obliczenia na tych zbiorach jedynie z mini-modelami. Dla celów porównawczych wyniki pozostałych metod zostały wzięte z publikacji innych autorów.

Eksperymenty obliczeniowe z mini-modelami w wielowymiarowych przestrzeniach cech zostały podzielone przez doktoranta w poniższy sposób (s, 122):

- mini-modele 2D,
- mini-modele bazujące na bryłach wielowymiarowych,
- mini-modele bazujące na bryłach wielowymiarowych z redukcją wymiarowości,
- mini-modele bazujące na algorytmach klasteryzacji.

Na zakończenie *Rozdziału 4* doktorant umieścił swoje podsumowanie eksperymentów w którym znalazły się m.in. stwierdzenia: „*Wyniki przeprowadzonych eksperymentów numerycznych potwierdzają skuteczność działania metody mini-modeli*”, oraz „*Szczególnie interesujące są wyniki eksperymentów w przestrzeni wielowymiarowej, w której to mini-modele sprawdziły się jako skuteczna metoda modelowania danych. Dla niektórych zbiorów takich jak Concrete czy Yacht metoda osiągała najlepsze rezultaty spośród porównywanych.*” (s. 145)

Wyniki przeprowadzonych eksperymentów numerycznych pokazały, że dokładność metody mini modeli jest na poziomie metod z którymi była porównywana. Według doktoranta: „*Przewaga mini-modeli nad metodą K-NN polega na wykorzystywaniu nie tylko odległości do najbliższego sąsiada, ale także informacji o nachyleniu powierzchni funkcyjnej.*” (s. 149)

Rozdział 5 zatytułowany *Podsumowanie pracy* rozpoczyna się od poniższego stwierdzenia doktoranta: „*Niniejsza praca podejmuje temat uczenia algorytmów regresyjnych metodą mini-modeli. Główna zasada działania prezentowanej metody opiera się na założeniu, że w procesie modelowania danych bardzo często jesteśmy zainteresowani odpowiedzią na jedno, konkretne zapytanie. Wiedza na temat wartości zmiennej zależnej w całej modelowanej dziedzinie nie zawsze jest wymagana.*”

Do oryginalnych osiągnięć przedstawionych w rozprawie doktorant zaliczył opracowanie poniższych zagadnień w konstrukcji mini-modeli:

- działanie mini-modeli w przestrzeniach wielowymiarowych ($> 3D$),
- oparcie mini- modeli na bryłach wielowymiarowych z redukcją wymiarowości i bez takiej redukcji,
- mini-modele z użyciem algorytmów grupowania.

W zakończeniu rozprawy doktorant wymienia poniższe ograniczenia proponowanej przez siebie metody konstrukcji mini-modeli:

- złożoność obliczeniowa, która wzrasta z licznnością uczących zbiorów danych
- trudności w określeniu optymalnego modelu lokalnego

Moja ocena pracy doktorskiej mgr inż. Marcina Pietrzykowskiego jest generalnie pozytywna pomimo zastrzeżeń co do użytej metodologii jak też wielu usterek w rozprawie.

Zagadnienie budowy lokalnych modeli klasyfikacyjnych i prognostycznych (regresyjnych) jest aktualnym problemem badawczym rozwijanym od wielu lat i stosowanym z sukcesami w ramach bioinformatyki w oparciu o metodę grupowania dualnego (ang. *biclustering*). Metoda grupowania dualnego posługuje się algorytmami grupowania (ang. *clustering*) zarówno w zbiorze obiektów (*pacjentów*) jak i jednocześnie w zbiorze cech (*genów*). Metoda grupowania dualnego (podwójnego) jest stosowana m.in. do analizy wysoko wymiarowych zbiorów danych genetycznych i rozwiana zwłaszcza w kontekście bardzo aktualnych zagadnień medycyny *personalizowanej* (*prezyzyjnej*).

Na podstawie przedstawionej rozprawy doktorskiej odniosłem wrażenie, że doktorant nie zna metod grupowania dualnego a w tym grupowania kolinearnego. Zastosowane tam metody obliczeniowe pozwalają m. in. efektywnie wydzielać lokalne modele prognostyczne nawet na bazie bardzo dużych zbiorów danych. Wydaje mi się, że mgr inż. Marcin Pietrzykowski nadmiernie i niepotrzebnie ograniczył się metodologicznie w swojej pracy doktorskiej.

Moje wątpliwości budzi m. in. próba użycia przez mgr inż. Marcina Pietrzykowskiego współrzędnych sferycznych do reprezentowania i obliczania płaszczyzn (s. 66). W rozpoznawaniu obrazów znane i stosowane są techniki które wydają się być znacznie bardziej efektywne w tego rodzaju zagadnieniach. Niska efektywność zastosowanych przez doktoranta procedur obliczeniowych była zapewne główną przyczyną ograniczenia się w eksperymentach do zbiorów uczących o relatywnie niskich wymiarach.

Niestety w rozprawie doktorskiej znaleźć można błędy językowe (np. *Brest Cancer* s. 137) oraz matematyczne (wzór 2.1).

Pomimo zauważonych błędów i usterek w rozprawie doktorskiej mgr inż. Marcina Pietrzykowskiego moja finalna opinia jest pozytywna. Opinia pozytywna wynika głównie z tego, że uważam zagadnienie budowy lokalnych modeli prognostycznych na podstawie zbiorów danych za aktualne i ważne. Doktorant określił w swojej pracy możliwości zastosowania metody mini-modeli w projektowaniu lokalnych modeli regresyjnych. Możliwości zastosowania metody mini-modeli zostały określone na podstawie obszernego zestawu eksperymentów obliczeniowych. Czynnikiem, który wpłynął na moją pozytywną

ocenę są też wskaźniki bibliometryczne doktoranta. Mgr inż. Marcin Pietrzykowski ma 37 cytowań, a indeks Hirsha (h-indeks) ma wartość 4.

Stwierdzam, że recenzowana rozprawa doktorska mgr inż. Marcina Pietrzykowskiego pt. „*Lokalne uczenie algorytmów regresyjnych metodą mini – modeli*” spełnia wymogi stawiane rozprawom doktorskim w wyżej wymienionej ustawie i na tej podstawie wnioskuję do Rady Wydziału Informatyki Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie o jej dopuszczenie do publicznej obrony.

A handwritten signature in blue ink, reading "Leszko Bobowski". The signature is written in a cursive, flowing style.