

prof. dr hab. inż. **Piotr Kulczycki**  
Polska Akademia Nauk, Instytut Badań Systemowych  
Akademia Górniczo-Hutnicza, Wydział Fizyki i Informatyki Stosowanej  
ul. Jaracza 30, 31-216 Kraków  
tel.: 601 314159  
e-mail: kulczycki@ibspan.waw.pl, kulczycki@agh.edu.pl

## RECENZJA

rozprawy doktorskiej

mgr inż. Sylwester Czmił

„Algorytmy nadzorowanego inkrementalnego uczenia maszynowego oraz  
ocena jakości ich klasyfikacji”

Recenzowaną rozprawę wykonano pod kierunkiem promotora Pana prof. dr hab. inż. Jacka Kluski. Przewód doktorski został wszczęty na Politechnice Rzeszowskiej, w dziedzinie nauk inżyniersko-technicznych, dyscyplinie *Informatyka techniczna i telekomunikacja*.

### 1. Charakterystyka rozprawy

Przedmiotem pracy jest synteza nowatorskiej procedury klasyfikacji inkrementalnej. Przymiotnik „inkrementalny” wskazuje tu, że do rozważanego zbioru danych można sukcesywnie dodawać nowe elementy lub podzbiory, bez konieczności przeprowadzania obliczeń od początku. Takie ujęcie pozwala efektywnie zwiększać liczbę zbioru danych, co lokuje przedmiot przedstawionych badań w dziedzinie obecnie wyjątkowo rozwojowej, w dużym stopniu dzięki sukcesywnemu postępowi w automatycznych systemach pozyskiwania i transmisji informacji.

Recenzowana rozprawa liczy 114 stron luźno, starannie sformatowanego tekstu. Układ rozprawy przywodzi na myśl dyktoryjki o dawnych pracach doktorskich, o przełomowej tematyce, liczących 3 strony. Otóż, sedno recenzowanej tu rozprawy sprowadza się do 3-stronicowego podrozdziału 5.1. Wszystko co przedtem, to wprowadzenie do niego, a po – weryfikacja, suplementy i podsumowania.

WPLYNĘŁO

07. MAR. 2025

POLITECHNIKA RZESZOWSKA  
im. Ignacego Łukasiewicza  
WYDZIAŁ ELEKTROINŻYNIERSTWA I INFORMATYKI  
ul. W. Pola 2  
35-059 Rzeszów  
tel. 17 865 1764

Kulczycki: recenzja doktoratu Sylwestra Czmiła

I tak, rozdziały 1 oraz 2 mają charakter wprowadzenia; pierwszy zawiera opisowe, natomiast drugi ściśle sformułowanie celu i zakresu pracy. W rozdziale 3 omówione są pokrewne procedury klasyfikacji danych, zwłaszcza te, których aspekty zostaną wykorzystane w dalszej części pracy. Z kolei rozdział 4 zawiera powszechnie stosowane miary (wskaźniki) jakości klasyfikacji – w istocie, jest ich wiele, a dotychczasowe doświadczenia aplikacyjne nie pozwalają jednoznacznie określić którego wskazania są statystycznie bardziej wiarygodne, czy też w praktyce po prostu lepsze. W metodach probabilistycznych zawsze pozostaje również pytanie czy otrzymany wynik nie jest przypadkowy. Dlatego w rozdziale 4 podano także metody porównawczego wnioskowania o zasadności wyników uzyskanych z użyciem powyższych wskaźników.

Rozdział 5, a zwłaszcza jego pierwszy podrozdział 5.1, stanowi kwintesencję rozprawy – przedłożono w nim nowatorską koncepcję inkrementalnego klasyfikatora SEVQ, opierając jego zamysł na ewolucyjnej kwantyzacji wektorowej (procedury EVQ) i adaptacyjnej teorii rezonansu (algorytm SFAM). W podrozdziałach 5.2-5.3 zilustrowano jego działanie na wyrazistych przykładach, po czym przyrównano go do wyjściowych koncepcji EVQ i SFAM. W rozległym rozdziale 6 dokonano szczegółowej i wielostronnej weryfikacji wyników otrzymywanych z użyciem skonstruowanego klasyfikatora.

W rozdziale 7 przedstawiono implementację sprzętową algorytmu SEVQ w układzie FPGA i programową w języku Python, poruszając tu zagadnienie czasu obliczeń. Kolejny rozdział 8 jest niejako wartością dodaną – opisane zostało oryginalne oprogramowanie służące do oceny jakości algorytmów klasyfikacyjnych. Pracę kończy podsumowanie, dodatki i 71-elementowa bibliografia.

Zamysł klasyfikatora SEVQ jest ciekawy. Otóż, w badanym zbiorze wyróżnia się podzbiory (jakby klastry) przypisane do poszczególnych klas, przy czym do każdej klasy może być przypisanych wiele takich podzbiorów. W trakcie działania procedury możliwa jest kreacja nowych podzbiorów. Każdy podzbiór jest charakteryzowany przez jeden wektor (zwany w rozprawie wektorem wag) i pomocnicze parametry, które w trakcie działania procedury są odpowiednio modyfikowane, w reakcji na informację dostarczaną przez kolejne elementy. Powyższy proces stanowi sedno algorytmu SEVQ. Sama klasyfikacja polega na przypisaniu badanego elementu do najbliższego jej wektora charakteryzującego poszczególne podzbiory i w konsekwencji do odpowiedniej klasy. Klasyfikowany element bierze udział w ewentualnym tworzeniu nowych podzbiorów, a także modyfikacji wspomnianych wektorów charakteryzujących oraz pomocniczych parametrów. Tu właśnie pojawia się inkrementalny charakter algorytmu – informacja wynikła z dotychczas rozważanych elementów jest zapisana syntetycznie w postaci owych wektorów i parametrów, a same elementy mogą już być „zapomniane”.

Skonstruowany niniejszym algorytm został wszechstronnie sprawdzony z wykorzystaniem 36 empirycznych zbiorów dostępnych w repozytorium zbiorów KEEL. Wydaje się wystarczająco



liczny i różnorodny, aby uznać go za reprezentatywny. Porównanie obejmowało 10 metod nieinkrementalnych oraz 12 inkrementalnych, zaimplementowanych głównie przy użyciu powszechnie dostępnych bibliotek, przy standardowych parametrach. Także ten wybór można uznać za prawidłowy. Wykorzystano 5 powszechnie używanych wskaźników jakości: AUC (Area Under Curve, w domyśle krzywej ROC), jak również ACC (Accuracy), PRE (Precision), REC (Recall) oraz F1 (średnia harmoniczna dwóch ostatnich). AUC bywa najskuteczniejszy, ACC najbardziej intuicyjny, zwłaszcza przy zbliżonej liczności klas, natomiast trzy ostatnie dobrze spisują się dla zbiorów niezrównoważonych, ze wskazaniem na ostatni. Żaden jednak nie dominuje nad pozostałymi i w praktyce używane są wszystkie.

Pozwolę sobie w dalszej części skupić uwagę na metodach inkrementalnych, do jakich należy skonstruowana procedura SEVQ.

Najpierw, w sekcji 6.3.3, porównano średnią wartość powyższych wskaźników na badanych zbiorach przy użyciu różnych metod klasyfikacji inkrementalnej. Tu algorytm SEVQ uzyskał najlepszą średnią względem wskaźnika AUC oraz wice-najlepszą dla pozostałych. Wyniki te SEVQ otrzymał przy średniej – względem innych badanych metod – wartości odchylenia standardowego, co dyskretnie uczula na potencjalną niestabilność wyników.

W sekcji 6.4.3 przedstawiono ranking algorytmów, przypominający olimpijską klasyfikację medalową, aczkolwiek o kolejności świadczy suma „medali”. Algorytm SEVQ zajmuje pierwsze miejsce względem każdego z pięciu wskaźników. Ten ranking ma w pracy znaczenie raczej ilustracyjne, niewątpliwie bardzo ciekawe. Szkoda, że Doktorant nie starał się uzyskać tu odpowiedzi do jakich zbiorów algorytm SEVQ jest szczególnie predestynowany – o dużej lub małej liczności, zrównoważonych czy nierównoważonych, dwu- albo więcej-klasowych, itd.

Z kolei, sekcja 6.5.2 prezentuje wykresy pudełkowe wyników poszczególnych metod. Zaznaczona jest mediana, której wartość ustala miejsce w rankingu, ale także pierwszy  $Q_1$  oraz trzeci kwartyl  $Q_3$  jako brzegi pudełka, a także końce tzw. wąsów  $Q_1 - 1,5 \cdot (Q_3 - Q_1)$  oraz  $Q_3 + 1,5 \cdot (Q_3 - Q_1)$ . Liczba 1,5 jest uzasadniona praktyką, gdyż odwołuje się do powszechnie stosowanej reguły trzech sigm – odchylenie standardowe i odchylenie ćwiartkowe  $Q_3 - Q_1$  są zdefiniowane tak, iż pierwsze jest z zasady około dwukrotnie mniejsze, stąd 1,5 zamiast 3. W przypadku wszystkich wskaźników jakości, algorytm SEVQ okazuje się najlepszy. Jednak w przypadku czterech bezpośrednio opartych na macierzy pomyłek – czyli ACC, PRE, REC, F1 – niepokoi bardzo duża liczba elementów odstających i to wyrazistych, natomiast dla AUC duża szerokość „dolnych wąsów”. Ewidentnie wskazuje to na potencjalną niestabilność wyników i istnienie rodziny zbiorów danych wobec których SEVQ działa słabiej. Ponownie żałuję, że nie dokonano próby analizy i wskazania rodzaju zbiorów, dla których algorytm ten działa dobrze, a dla których gorzej.



W sekcji 6.6.2 przedstawiono wyniki testu Scotta-Knotta, polegającego na hierarchicznym grupowaniu. Względem każdego wskaźnika jakości, algorytm SEVQ lokował się w najlepszej grupie.

I wreszcie, przedstawiony w sekcji 6.7.3 test Wilcozona spina powyższe dywagacje wskazując czy wcześniejsze wyniki nie były przypadkowe, co przy ujęciu probabilistycznym zawsze powinno być brane pod uwagę. Testowana jest hipoteza o braku statystycznie istotnych różnic. Tak więc jej odrzucenie stanowi, że wynik nie jest przypadkowy, co w połączeniu z poprzednimi badaniami wskazywałoby na probabilistycznie zasadną przewagę algorytmu SEVQ.

Niestety, dalsza część rozumowania wydaje się błędna. Test Wilcozona jest tzw. testem istotności, gdzie weryfikowaną hipotezę zerową odrzuca się z założonym prawdopodobieństwem błędu pierwszego rodzaju (odrzuć hipotezę prawdziwej) lub nie odrzuca, bynajmniej nie przyjmując hipotezy alternatywnej (nawet można jej nie formułować), gdyż zachodziłoby niebezpieczeństwo błędu drugiego rodzaju, który w testach istotności bywa całkowicie niekontrolowany. Wartość powyższego prawdopodobieństwa nazywana jest poziomem istotności i w praktyce zwykle przyjmuje się 0,01, 0,05 lub 0,1. Mniejsze wartości wskazują na mocniej uzasadnione odrzucenie testowanej hipotezy, większe stanowią o słabszym uwarunkowaniu. Dokonane w rozprawie sformułowanie hipotezy alternatywnej jest bezcelowe, a jej przyjmowanie absolutnie bezzasadne, merytorycznie błędne. Szczęśliwie doktorant zamieścił tzw. wartości  $p$  ( $p$ -value), co pozwala uratować sytuację. Na przekopiowanej poniżej z rozprawy tabeli 6.5 zaznaczyłem na zielono te wyniki, które pozwalają odrzucić testowaną hipotezę z dowolnym stosowanym w praktyce poziomem istotności, na żółto te które pozwalają to zrobić dla mniej „ostrych” poziomów, a czerwonym te które nie pozwalają odrzucić testowanej hipotezy. (To jednak nie wskazuje na gorsze wyniki algorytmu SEVQ, a jedynie nie pozwala wnioskować o istotnej różnicy między wynikami porównywanych procedur!)

Tabela 6.5: Wyniki testu Wilcozona użytego do porównania algorytmu SEVQ oraz klasyfikatorów inkrementalnych

	Algorytm	Wartość $p$ dla ACC	Wartość $p$ AUC	Wartość $p$ PRE	Wartość $p$ dla SEN	Wartość $p$ F1
1	AEE	0,0001	0,0013	0,0001	0,0001	0,0000
2	ARF	0,0374	0,0001	0,0000	0,0374	0,0000
3	DWM	0,0000	0,0000	0,0000	0,0000	0,0000
4	EFDI	0,0199	0,0000	0,0000	0,0199	0,0001
5	EVQ	0,3275	0,0209	0,4649	0,3275	0,8284
6	HAT	0,1737	0,0587	0,0023	0,1737	0,0061
7	HT	0,2214	0,0120	0,0012	0,2214	0,0050
8	KNNI	0,9419	0,0003	0,0786	0,9419	0,1796
9	LVQ	0,0000	0,0000	0,0000	0,0000	0,0000
10	LVQ2	0,0001	0,0000	0,0000	0,0001	0,0000
11	LVQ2.1	0,0000	0,0000	0,0000	0,0000	0,0000
12	LVQ3	0,0000	0,0000	0,0000	0,0000	0,0000
13	NB	0,0189	0,5311	0,1056	0,0189	0,0079
14	OB	0,3693	0,0025	0,4098	0,3693	0,7303
15	SFAM	0,0000	0,0000	0,0003	0,0000	0,0000

Jak widać, w przypadku każdego algorytmu porównywanego z SEVQ, chociaż jeden ze wskaźników daje wynik upoważniający do wnioskowania o probabilistycznej istotności różnic wobec SEVQ (kolor zielony i nieco ostrożniej żółty), a poprzednie badania wskazywały na jego większą skuteczność. A zatem uwzględnienie algorytmu SEVQ „w koszyku” rozważanych do użycia procedur jest w pełni uzasadnione. W przypadku klasyfikacji istnieje bowiem możliwość (m.in. walidacja krzyżowa, metoda *leave-one-out*) sprawdzenia która z metod wydaje się najlepsza dla konkretnego przypadku. Przedłożenie w ramach rozprawy doktorskiej algorytmu, który jak najbardziej zasadnie może być uwzględniany w „koszyku” najlepszych znanych obecnie procedur, potencjalnych „zwycięzców”, jest godnym podkreślenia wynikiem. Warto zwrócić uwagę, że jest on dodatkowo relatywnie prosty i dogodnie interpretowalny oraz nie wymaga estymacji żadnych parametrów, co w praktyce jest często kłopotliwe.

## 2. Uwagi krytyczne i dyskusyjne

Pracę czyta się trudno. Zawiera ona dużą ilość nieścisłości, pomyłek, opuszczeń, czy też niewłaściwie użytych terminów.

a.

Strona 15<sub>12</sub>: „minimalną liczbę parametrów”. Można od razu wskazać, że minimalną jest zero. Chodziło raczej o możliwie najmniejszą.

b.

Strony 18-19: wzory (3.1) i (3.50) oraz (3.2) i (3.51) są takie same. Różnią się tylko symbolem indeksu, który można przecież „podmianiać” dowolnie.

c.

Strony 23<sub>9</sub> oraz 24<sub>1</sub>: cytowany wzór (5.1) zostanie wprowadzony dopiero w następnym rozdziale.

d.

Strona 26: wielkość *AUC* zdefiniowana jest jako uśrednienie elementów jakiejś dwuwymiarowej macierzy, która nie została jednak podana.

e.

Strona 29<sup>13</sup>: czy nie jest konieczne założenie, że każda klasa jest reprezentowana przez co najmniej 1 element? Nota bene: wtedy warunek  $r \leq z$  byłby automatycznie spełniony.

f.

Strona 29<sup>14</sup>: atrybutami są współrzędne wektorów, a nie same „*j*-wymiarowe wektory rzeczywiste”.

g.

Strona 29<sub>7</sub>: skąd nazwa „wektor wag”? Wagami są najczęściej wielkości, które pomnożone przez inne elementy wskazują istotność (ważność, wagę) poszczególnych elementów.



h.

Strona 30, listing 1: jakie są wartości początkowe wektorów  $w_i$ ? Jak będzie działać linia 4 gdy  $k = 0$  i zbiór argumentów operacji minimum będzie pusty? Podobnie gdy  $k = i = 1$ .

i.

Strona 32: czy przewidywana jest procedura łączenia lub usuwania kategorii? Przy dużych zbiorach ich liczebność może nadmiernie „puchnąć”.

j.

Strony 32-34: dlaczego elementy analizowanego zbioru nazywane są tu „przykładami”? (Takie nazewnictwo występuje jedynie w tej sekcji.)

k.

Strona 41<sub>6</sub>: po co w opisie algorytmu Naiwnego Bayesa zapisano „Dla klasyfikatora *NaiveBayes* atrybuty były numeryczne.”? Przecież to dotyczy każdego rozważanego algorytmu, co nawet jest podane na stronie 37<sup>8</sup> i powtórzone w 37<sup>12</sup>, aczkolwiek tu występuje określenie „liczbowe”.

l.

Strony 43-44, 69, 72: tabele 6.2 i 6.3 oraz 6.4 i 6.5 dostarczają ciekawego testu na inteligencję czytającego, gdyż kolumny dotyczące wskaźników AUC i ACC zostają zniekształcone. Test ten został dodatkowo utrudniony, gdyż początkowa logiczna kolejność AUC, ACC, PRE, SEN, F1 została zamieniona na mniej uzasadnioną ACC, AUC, PRE, SEN, F1. Wskaźnik ACC pochodzi bowiem z tej samej rodziny co PRE, SEN, F1 i raczej powinien być lokowany przy nich. Podobnie kolejność AUC ↔ ACC zamieniana jest w tekście.

m.

Strona 46<sub>4-5</sub>: nie przewyższa pozostałych i to tym bardziej znacząco, gdyż AB też ma 4 pierwsze miejsca.

n.

Strona 48<sub>5-6</sub>: absolutnie nie ma największej liczby trzecich miejsc.

o.

Strona 51<sup>5-6</sup>: duża liczba elementów odstających jest naturalną konsekwencją krótkich wąsów.

p.

Strona 51<sub>1-4</sub>: algorytm QDA ma bardzo szerokie pudełko i wąsy, ale to trudno nazwać wynikiem ciekawym. To jest raczej niebezpieczna własność. Stąd też wynika naturalny w tej sytuacji brak elementów odstających. Wyrażenie „prezentują pewien stopień wartości odstających” nie powinno zaistnieć w dysertacji doktorskiej.

r.

Strona 68<sub>8-10</sub>: testy nieparametryczne nie wymagają nie tyle założenia o normalności rozkładu, ale jakiegokolwiek założenia o konkretnej postaci rozkładu (normalnego, jednostajnego, wykładniczego, itd.).

s.

Strony 75-78: w rozdziale tym przedstawione jest zagadnienie czasu obliczeń, ale rozważaniom podlegają tu dwie różne implementacje algorytmu SEVQ, a nie porównanie czasu jego obliczeń z czasami innych procedur, co byłoby ważniejsze, a nawet wysoce pożądane w przedstawionych badaniach.

t.

Strony 67-89: bardzo brakuje mi wskazań do jakich zbiorów i uwarunkowań szczególnie predestynowany jest skonstruowany algorytm SEVQ. Przynajmniej próby uzyskania wskazań w tym zakresie.

u.

Strona 89<sup>4</sup>: gorąco rekomenduję Doktorantowi badania nad niestacjonarnością danych (por. także powyższa uwaga „i”), co umożliwiłoby zastosowania przedłożonego, bardzo wartościowego algorytmu SEVQ do strumieni danych, ważnej i obecnie znacząco rozwijającej się dyscypliny.

Ogólnie: przytoczone powyżej uwagi mają charakter drugorzędny oraz dyskusyjny i nie negują generalnie pozytywnej mojej oceny zawartości przedstawionej dysertacji. Jej sednem było stworzenie inkrementalnego algorytmu klasyfikacji o dużej wartości aplikacyjnej, co zostało z powodzeniem wykonane i przedstawione.

### 3. Konkluzja

**Podsumowując uważam, że recenzowana praca doktorska Pana mgr inż. Sylwestra Czmiła spełnia wymagania stawiane w odpowiednich przepisach rozprawom doktorskim i wobec tego wnioskuję o dopuszczenie Go do dalszych, przewidzianych ustawą, etapów przewodu doktorskiego.**



(prof. dr hab. inż. Piotr Kulczycki)

Kraków/Warszawa, 28 lutego 2025.