

Recenzja

rozprawy doktorskiej mgr. Tomasza Kloneckiego

pt. „Cost-constrained feature selection using information theory”

Podstawa prawna

Niniejszą recenzję przygotowałem w związku z uchwałą nr 27/2025 Rady Naukowej Instytutu Podstaw Informatyki PAN. Postępowanie doktorskie mgr. Tomasza Kloneckiego wszczęto 18.09.2025 i prowadzone jest ono na podstawie przepisów ustawy Prawo o szkolnictwie wyższym i nauce (Dz. U. 2024 poz. 1571 z późn. zm.).

Zakres i forma rozprawy

Rozprawa doktorska dotyczy zagadnienia selekcji cech predykcyjnych w przypadku, gdy ich pozyskiwanie skutkuje różnej wysokości kosztami. To istotny temat stanowiący naturalne rozwinięcie i pogłębienie klasycznych metod doboru zmiennych. Wydaje się być jednak niezbyt często podejmowany w praktyce, co objawia się małą liczbą referencyjnych zbiorów danych.

Rozprawa przygotowana została w języku angielskim. Wywody prowadzone są spójnie i jasno. Widoczna jest sprawność językowa oraz wysoka dbałość redakcyjna. Na 116 stron rozprawy dostrzegłem jedynie cztery literówki¹.

Ocena merytoryczna

Tematykę opracowano bardzo starannie. Wstęp klarownie opisuje zagadnienie selekcji zmiennych i główne wyniki literaturowe. Pomocny i dydaktyczny jest też rozdział poświęcony teorii informacji. Rozprawa prezentuje wysoki poziom matematyczny. Wyróżnia ją przejrzystość wnioskowań oraz staranność redakcji.

Zaletą pracy jest kompleksowość podejścia badawczego. Obejmuje ono zarówno przypadek klasyfikacji wieloetykietowej, jak i przypisywania kosztów całym grupom predyktorów zamiast pojedynczym atrybutom. Nowością jest wykorzystanie ograniczeń dolnych na miary informacji wzajemnej. Podejście takie pozwala na identyfikację ogólnych zależności, w tym wyższych rzędów. Co więcej, oprócz stosowania selekcji jako filtru – co pozwala na integrację z dowolnym algorytmem predykcyjnym – zaproponowano uwzględniające koszty penalizowane metody minimalizacji ryzyka empirycznego.

Wyniki teoretyczne poparto przemyślanymi eksperymentami z użyciem danych referencyjnych oraz symulacyjnych. Cechują się one solidną, rzetelną metodologią. Zwraca uwagę pomysłowe

¹ *an uniform* zamiast *a uniform*; *higher the vaule*, *stronger the emphasis* zamiast *the higher the value*, *the stronger the emphasis*; nierówność ostra zamiast nieostrej w dowodzie lematu 1; J_{sfts} zamiast J_{stfs}

wykorzystanie dodatkowych, sztucznie generowanych zmiennych „zacierających”, aby oceniać i ograniczać poziom przeuczenia. Wyniki przedstawiono w czytelnych tabelach i na odpowiednio dostosowanych wykresach.

Ograniczeniem proponowanych metod jest konieczność dyskretyzacji zmiennych ciągłych, co wymaga podjęcia przez użytkownika dodatkowych decyzji oraz rodzi pytania o odporność wyników na wybór liczby i wartości punktów podziału.

Zastanawia mnie też, czy algorytmy literaturowe wybrane do porównania w podrozdziale 5.4.2 adekwatnie odzwierciedlają współczesny front badań. Dla przypadku pojedynczej zmiennej objaśnianej mają one aż 33, 20, 19 i 26 lat. Czyżby przez ostatnie dwie dekady nie poczyniono istotnego postępu w zakresie doboru cech przy ograniczeniu kosztowym? Sytuacja wygląda lepiej dla klasyfikacji wieloetykietowej, gdyż proponowane w rozprawie algorytmy porównuje się z metodami publikowanymi przed 13, 10, 12, 8, 2, i 4 laty.

Moje pozostałe uwagi dyskusyjne dotyczą jedynie drobnych kwestii. W rozdziale 5 zmienne proksy oznaczone są gwiazdką, a w 7 apostrofem. Czytelniej byłoby uspoźnić notację. Opis omawiający rysunek 6.1 nie odpowiada w całości wynikami dla prognozy 10%. W podrozdziale 7.2 parametr regularyzacji λ_k wyznaczany jest za pomocą przeszukiwania siatki stu potencjalnych wartości. To poprawne rozwiązanie, efektywniej jednak byłoby chyba wykorzystać do jego dokładnej lokalizacji metodę bisekcji. Uwagi te nie zaburzają zdecydowanie pozytywnego odbioru całej rozprawy, a ich szczegółowy charakter świadczy o jej wysokim stopniu jej dopracowania.

Konkluzja

Głębia merytoryczna nowozaproponowanych metod, jak również staranność opracowania wyników i ich prezentacji powodują, że bardzo wysoko oceniam recenzowaną rozprawę doktorską. Spełnia ona wszelkie wymagania stawiane tego typu dysertacjom. Wnoszę do Rady Naukowej Instytutu Podstaw Informatyki PAN o dopuszczenie mgr. Tomasza Kloneckiego do dalszych etapów postępowania doktorskiego.

Karol Opara