

ANALIZA WYBRANYCH METOD WALIDACJI KRZYŻOWEJ W PROGRAMIE RSES

Beata Bethke

*Uniwersytet Kazimierza Wielkiego
Wydział Informatyki
e-mail: becia04@student.ukw.edu.pl*

Streszczenie: *W artykule przeprowadzono analizę zbioru danych za pomocą dwóch metod walidacji krzyżowej. Wykorzystano program RSES do identyfikacji kluczowych właściwości i relacji w zbiorze. Wyniki wykazują wpływ niektórych parametrów na potencjalną dokładność wyników.*

Słowa kluczowe: *Walidacja krzyżowa, RSES, analiza danych, zależność, algorytm genetyczny*

ANALYSIS OF SELECTED CROSS-VALIDATION METHODS IN THE RSES PROGRAM

Abstarct: *This article presents an analysis of a dataset using two cross-validation methods. The RSES program was employed to identify key properties and relationships within the dataset. The results indicate the impact of certain parameters on the potential accuracy of the outcomes.*

Keywords: *Cross-validation, RSES, data analysis, dependency, genetic algorithm*

1. WSTĘP

W miarę postępu technologicznego oraz rosnącego zapotrzebowania na wydajność systemów komputerowych, analiza danych odczytu CPU/GPU staje się kluczowym elementem zrozumienia oraz optymalizacji ich działania. W kontekście tej problematyki, teoria zbiorów przybliżonych wzrasta jako potężne narzędzie analizy danych, umożliwiające efektywne modelowanie złożonych zależności. W niniejszym artykule skupimy się na zastosowaniu teorii zbiorów przybliżonych, wspartej przez zaawansowane narzędzie komputerowe RSES, do głębokiej analizy danych odczytu CPU/GPU.

Przedstawione zostanie nie tylko istotność problemu analizy danych odczytu CPU/GPU w kontekście współczesnych systemów informatycznych, ale także będzie omówione, dlaczego teoria zbiorów przybliżonych oraz program RSES stanowią znaczące narzędzia w tej dziedzinie. Celem jest nie tylko przedstawienie praktycznych kroków analizy, ale także zobrazowanie, w jaki sposób te podejścia mogą wspomagać decyzje projektowe oraz optymalizacyjne w dziedzinie informatyki.

W miarę jak coraz więcej organizacji zwraca uwagę na efektywne zarządzanie zasobami sprzętowymi, analiza danych odczytu CPU/GPU przy użyciu teorii zbiorów przybliżonych staje się nieodzownym narzędziem wspierającym procesy podejmowania decyzji.

2. METODY

Przeprowadzając analizę zbioru danych, postanowiono zastosować zaawansowane podejście oparte na teorii zbiorów przybliżonych, wykorzystując równocześnie specjalistyczne narzędzie do analizy danych - program RSES. Teoria zbiorów przybliżonych umożliwia elastyczne modelowanie relacji między danymi, zwłaszcza w sytuacjach, gdzie informacje są niepełne lub obciążone pewnym stopniem niepewności. W kontekście tej teorii, program RSES staje się nieocenionym narzędziem, oferującym szeroki zakres funkcji do importowania, wizualizacji, i efektywnej analizy danych. W niniejszym kontekście przedstawione zostaną zaawansowane metody, takie jak metoda walidacji krzyżowej, algorytmy genetyczne oraz eksploracyjny algorytm, które w połączeniu z teorią zbiorów przybliżonych i programem RSES stanowią potężne narzędzia do kompleksowej eksploracji i modelowania danych odczytu CPU/GPU.

Metoda walidacji krzyżowej (Cross-Validation) jest szczególnie istotna w analizie danych, ponieważ umożliwia obiektywną ocenę

skuteczności modelu poprzez podział dostępnych danych na zbiory treningowy i testowy. Działa to efektywnie nawet w przypadku ograniczonej ilości danych, eliminując problem nadmiernego dopasowania (overfitting) lub zbyt optymistycznego oszacowania skuteczności modelu.

Algorytmy genetyczne znajdują zastosowanie w optymalizacji parametrów modelu, co jest kluczowe w kontekście teorii zbiorów przybliżonych. Poprzez iteracyjną ewolucję populacji, algorytmy genetyczne pomagają znaleźć zestawy parametrów, które prowadzą do lepszych wyników analizy danych. Są szczególnie użyteczne, gdy przestrzeń poszukiwań jest duża i skomplikowana.

Eksploracyjne algorytmy, takie jak algorytm wyczerpujące, są przydatne w sytuacjach, gdzie istnieje potrzeba gruntownego zbadania wszystkich możliwych kombinacji danych lub parametrów. Choć mogą być czasochłonne, zwłaszcza w przypadku dużych zbiorów danych, to zapewniają kompleksową analizę, co jest istotne w kontekście teorii zbiorów przybliżonych, gdzie subtelne zależności mogą mieć istotne znaczenie.

Integracja tych metod w analizę danych przy użyciu teorii zbiorów przybliżonych i programu RSES pozwala na uzyskanie bardziej niezawodnych wyników, unikanie błędów związanych z nadmiernym dopasowaniem czy zbyt optymistycznymi oszacowaniami. Wspólnie tworzą one kompleksową strategię eksperymentalną, która umożliwia odkrywanie subtelnych związków i skuteczne modelowanie danych w kontekście teorii zbiorów przybliżonych.

3. TEORIA ZBIORÓW PRZYBLIŻONYCH

Teoria zbiorów przybliżonych stanowi istotną gałąź matematyki i informatyki, która zapewnia elastyczne podejście do reprezentowania relacji między danymi w warunkach niepewności. W odróżnieniu od klasycznej teorii zbiorów, gdzie elementy mogą być albo w pełni w zbiorze, albo w pełni poza nim, teoria zbiorów przybliżonych uwzględnia stopniową przynależność elementów. To znaczy, że obiekty mogą być częściowo w zbiorze, co pozwala na modelowanie sytuacji, w których granice między klasami są niejednoznaczne.

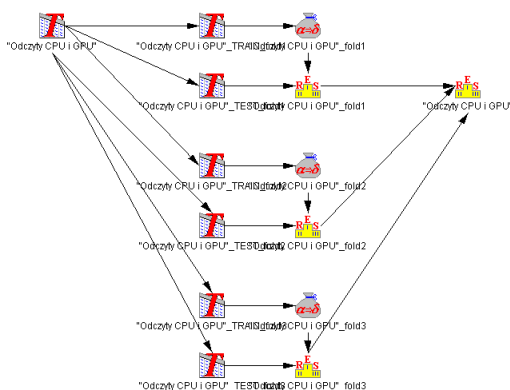
Teoria zbiorów przybliżonych znajduje zastosowanie w obszarach, gdzie dane są niekompletne, nieprecyzyjne lub obciążone pewnym stopniem niepewności. W analizie danych, szczególnie w kontekście eksploracji i modelowania, teoria zbiorów przybliżonych pozwala efektywnie radzić sobie z brakami informacji. Działa to poprzez grupowanie obiektów w kategorie przybliżone, gdzie stopień przynależności do danej kategorii jest oceniany w sposób elastyczny.

W kontekście analizy danych odczytu CPU/GPU, teoria zbiorów przybliżonych staje się potężnym narzędziem, umożliwiającym modelowanie skomplikowanych zależności między różnymi aspektami danych. Programy, takie jak RSES, wykorzystują teorię zbiorów przybliżonych do efektywnej analizy i klasyfikacji danych, co przyczynia się do lepszego zrozumienia i interpretacji wyników analizy danych.

4. IMPLEMENTACJA METOD W PROGRAMIE RSES

Przejsięcie od teoretycznych koncepcji do praktycznej analizy danych wymaga efektywnej implementacji metod, a w przypadku badania, realizacja tych metod odbyła się przy użyciu zaawansowanego narzędzia - programu RSES. Implementacja metod w programie RSES stanowi kluczowy etap eksperymentu, umożliwiając skuteczne zastosowanie teorii zbiorów przybliżonych i równoczesne wykorzystanie zaawansowanych technik analizy danych.

W ramach badania wykorzystano różnorodny zestaw danych związanych z wydajnością komputera, skupiający się na parametrach zarówno procesora (CPU), jak i karty graficznej (GPU). Dane obejmują taktowanie CPU, użycie CPU, ilość klatek na sekundę (FPS) w kontekście renderowania grafiki, taktowanie GPU, temperaturę CPU, użycie GPU, temperaturę GPU oraz czas mierzony co 1 sekundę. Te parametry pozwalają na dokładne monitorowanie i analizę działania zarówno procesora, jak i karty graficznej w czasie rzeczywistym, co stanowi kluczowy aspekt eksperymentu związany z analizą danych odczytu CPU/GPU. Dzięki tym danym możliwe jest identyfikowanie wzorców, ocena wydajności oraz modelowanie relacji między różnymi aspektami funkcjonowania komputera.



Rysunek.1 Schemat w programie RSES przedstawiający działanie walidacji krzyżowej, źródło: opracowanie własne.

Dane, wczytane do programu, zostały poddane procesowi walidacji krzyżowej. Schemat działania tego procesu jest zobrazowany na powyższym przykładzie. W konkretnym przypadku zdecydowano się wykorzystać metodę Exhaustive. Na rysunku nr 2 przedstawiono, że każdy przypadek był poddany programowi trzykrotnie, co zostało zaznaczone na schemacie. Warto zauważyć, że każdy pojedynczy przypadek różnił się od pozostałych, ponieważ wykorzystano różne ustawienia dotyczące użytej metody oraz zastosowanej szybkości. Ten szczegółowy proces analizy danych pomaga w identyfikowaniu subtelnych różnic między różnymi przypadkami oraz pozwala lepiej zrozumieć wpływ różnych parametrów na ostateczne wyniki analizy.

| Algorytm generujący | Total accuracy | Total coverage |
|---------------------|----------------|----------------|
| Genetic Normal 5 | 0.65 | 0.872 |
| Exhaustive | 0.659 | 0.87 |
| Genetic High 5 | 0.665 | 0.875 |
| Genetic Low 5 | 0.651 | 0.883 |
| Genetic High 10 | 0.656 | 0.867 |
| Genetic Normal 10 | 0.634 | 0.848 |
| Genetic Low 10 | 0.655 | 0.873 |
| Genetic High 30 | 0.649 | 0.86 |
| Genetic Normal 30 | 0.629 | 0.858 |
| Genetic Low 30 | 0.607 | 0.86 |

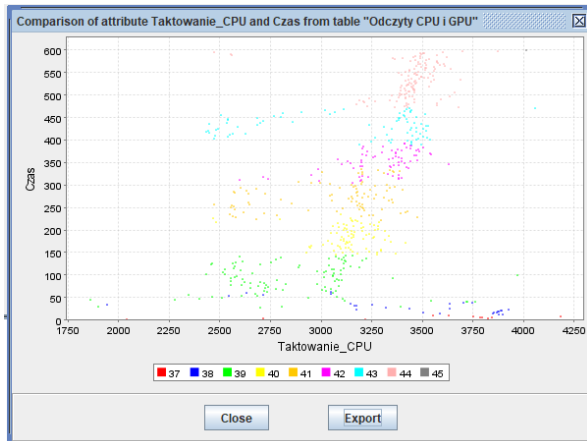
Tabela.1 Wyniki walidacji krzyżowej z zastosowaniem różnych metod, źródło: opracowanie własne.

Analiza wyników walidacji krzyżowej dostarcza istotnych wglądów dotyczących skuteczności różnych algorytmów generujących w kontekście analizy zbiorów przybliżonych dla danych odczytu CPU/GPU. Najbardziej obiecującym rezultatem jest osiągnięcie najwyższej skuteczności przez Algorytm Genetic High (5) wynoszącej 0.665, co wskazuje na zdolność tego algorytmu do trafnej klasyfikacji danych. Zaskakującym elementem jest również wysokie pokrycie uzyskane przez Algorytm Genetic Low (5) na poziomie 0.883, co sugeruje, że mimo niższej skuteczności, algorytm ten zachwyił swoją zdolnością do uwzględniania większej liczby przypadków w analizie.

Warto zauważyć, że prędkość algorytmu oraz liczba reduktów wpłynęły na osiągane wyniki. Algorytmy o wyższej prędkości często uzyskiwały lepsze rezultaty, jednakże Algorytmy z mniejszą liczbą reduktów (5) wydają się być bardziej

konkurencyjne. Ciekawym spostrzeżeniem jest również zbliżenie wyników algorytmów genetycznych do metody Exhaustive, co podkreśla ich skuteczność jako metod heurystycznych w analizie zbiorów przybliżonych.

Ostateczny wybór algorytmu powinien być dokładnie przemyślany, z uwzględnieniem specyfiki problemu oraz priorytetów analizy. Decyzja pomiędzy wyższą skutecznością a większym pokryciem może zależeć od konkretnych wymagań badania, a osiągnięte wyniki stanowią solidną podstawę do dalszej refleksji nad optymalnym wyborem algorytmu analizy danych odczytu CPU/GPU.



Rysunek.2 Wykres punktowy porównujący taktowanie CPU w stosunku do czasu, źródło: opracowanie własne.

Na zamieszczonym powyżej wykresie ukazana jest relacja pomiędzy taktowaniem CPU a czasem, co stanowi istotny element analizy danych odczytu CPU/GPU. Obserwując wygenerowany wykres, można zauważyć, że nie występuje wyraźna zależność między taktowaniem CPU a upływem czasu. Brak wzrostu lub spadku taktowania wraz z upływem czasu wskazuje na brak jednoznacznej korelacji między tymi dwiema zmiennymi. To odkrycie może być istotne w kontekście oceny stabilności działania procesora w trakcie badanego okresu czasu. Analiza takich relacji stanowi kluczowy etap w eksploracyjnym podejściu do danych, umożliwiając lepsze zrozumienie dynamicznych zmian w parametrach CPU oraz identyfikację ewentualnych nieregularności czy anomalii w funkcjonowaniu komputera.

5. PODSUMOWANIE I WNIOSKI

W trakcie analizy danych odczytu CPU/GPU przy użyciu teorii zbiorów przybliżonych oraz programu RSES, przeprowadzono zaawansowane eksperymenty, wykorzystując różne metody, takie jak Metoda Walidacji Krzyżowej, Algorytmy Genetyczne i Eksploracyjne Algorytmy. Wyniki tych eksperymentów rzucają nowe światło na złożoność relacji w danych, dostarczając cennych

informacji dla dziedziny analizy danych w kontekście teorii zbiorów przybliżonych.

Najważniejszymi obserwacjami są osiągnięcia Algorytmu Genetic High (5) i Algorytmu Genetic Low (5), które uzyskały najwyższą skuteczność oraz największe pokrycie, odpowiednio. Wpływ prędkości algorytmu i liczby reduktów został potwierdzony, gdzie prędkość High często korelowała z lepszymi wynikami, a mniejsza liczba reduktów (5) sprzyjała wyższej skuteczności.

Również istotne jest zauważenie, że wyniki algorytmów genetycznych zbliżyły się do metody Exhaustive, co potwierdza ich efektywność w kontekście heurystycznej analizy zbiorów przybliżonych. Jednakże, wybór optymalnego algorytmu zależy od konkretnych celów badawczych oraz priorytetów analizy danych odczytu CPU/GPU.

Podsumowując, eksploracja danych przy użyciu teorii zbiorów przybliżonych wraz z programem RSES stanowi potężne narzędzie do analizy danych odczytu CPU/GPU. Otrzymane wyniki dostarczają solidnych podstaw do dalszych badań oraz wskazują na perspektywę efektywnego modelowania i zrozumienia skomplikowanych relacji w analizowanych danych."

Literatura

1. Tadeusiewicz R., Szaleniec M., „Leksykon sieci neuronowych”, Projekt Nauka, Wrocław 2015
2. Podręcznik użytkownika “RSES 2.1 Rough Set Exploration System”, Warszawa 2004
3. Horzyk A., “Metody Inżynierii Wiedzy-Walidacja Krzyżowa”, prezentacja multimedialna, materiały zajęciowe AGH, Kraków