

Warszawa, 31.01.2006 r.

Recenzja rozprawy doktorskiej pani mgr inż. Doroty Cendrowskiej, napisanej pod kierunkiem prof. dr. hab. Witolda Kosińskiego, zatytułowanej:

„Konstrukcja klasyfikatora obiektów z wykorzystaniem algorytmu badania rozdzielnosci liniowej dwóch zbiorów.”

1. Wstęp.

Po przeczytaniu rozprawy doktorskiej pani Doroty Cendrowskiej przyszła mi na myśl następująca uwaga. Wśród prac naukowych można wyróżnić dwie kategorie. Pierwsza polega na formułowaniu nowych zadań badawczych, stawianiu pytań ich dotyczących i udzielaniu na nie odpowiedzi. Druga to włączenie się w nurt badań w toku i podjęcie problemów sformułowanych, lecz nie rozwiązanych, przez innych badaczy (oczywiście z twórczą ich modyfikacją) i próba uzyskania rozwiązania tych problemów. W tym drugim przypadku istnieje ryzyko, że problem może zostać rozwiązany wcześniej przez kogoś innego, lub też jest na tyle trudny (nie bez przyczyny nie miał jeszcze rozwiązania), że nie uda się go rozwiązać. Pani Dorota Cendrowska zdecydowała się zająć zagadnieniem z tej drugiej kategorii i, trzeba przyznać, odniosła tu znaczący sukces.

2. Charakterystyka pracy

Rozprawa doktorska pani mgr Doroty Cendrowskiej poświęcona jest konstrukcji hierarchicznego klasyfikatora obiektów wykorzystującego algorytm rozdzielania liniowego zbiorów ze skończonym prześwitem. Rozprawa ma 97 stron, składa się z e wstępu, pięciu rozdziałów, podsumowania, dwóch dodatków i bibliografii zawierającej 45 pozycji, wśród których jest 6 prac własnych autorki (1 praca w międzynarodowym czasopiśmie recenzowanym, 2 konferencyjne w języku polskim i 3 konferencyjne w języku angielskim).

W rozdziale **pierwszym** zaprezentowane są podstawowe metody wykorzystywane w zagadnieniach klasyfikacji, przede wszystkim klasyfikatory liniowe. Przedstawione są ich definicje, podstawowe własności oraz przedyskutowane są ich cechy pod względem przydatności do budowy hierarchicznego klasyfikatora odcinkowo-liniowego, którego budowa jest celem recenzowanej rozprawy.

W rozdziale **drugim** przedstawiony został szczegółowo algorytm LS2S (zaczepnięty z literatury), którego własności i stopień przebadania predestynują, zdaniem autorki, do roli algorytmu bazowego w konstrukcji klasyfikatora hierarchicznego.

Rozdział **trzeci** jest poświęcony algorytmom rozdzielającym zbiory punktów z zadaniem skończonym prześwitem. W szczególności, przedstawiono w nim definicję algorytmu SLS2S rozdzielania dwóch zbiorów z maksymalnym prześwitem, czyli, jak to zdefiniowano, w optymalny sposób. Rozdział ten zawiera też sformułowanie twierdzenie o istnieniu rozwiązania procedury algorytmu SLS2S oraz jednoznaczności tego rozwiązania (dowód twierdzenia jest w dodatku A).

Rozdział **czwarty** dotyczy optymalizacji wspomnianego algorytmu pod względem minimalizacji liczby wykonań jednego z jego kroków (kroku, w którym znajdowany jest wektor prostopadły do hiperpłaszczyzny rozpiętej na zadanych punktach). Poprawna

wykonalność tego kroku (tj. jednoznaczna i wymagająca skończonej liczby operacji) jest przedmiotem wspomnianego twierdzenia 1.

Rozdział **piąty** zawiera wynik, który jest końcowym celem pracy, czyli konstrukcję algorytmu klasyfikatora hierarchicznego odcinkowo-liniowego, wykorzystującego wspomniany algorytm SLS2S optymalnego rozdzielania z prześwitem.

Prace kończy podsumowanie oraz załączone dwa dodatki. W pierwszym z nich przedstawiono dowód poprawności algorytmu SLS2S (obejmujący 25 stron), a w drugim przykład kodu programu realizującego algorytm LC2S iteracyjnego budowania drzewa hierarchii hiperpłaszczyzn rozdzielających w klasyfikatorze.

3. Ocena pracy

Praca zawiera jasny wywód matematyczny, w którym, poczynając od historycznych uwarunkowań, przedstawiono zagadnienie liniowej rozdzielności zbiorów i sformułowano problem budowy klasyfikatora hierarchicznego. W całej pracy autorka uzasadnia podejmowane decyzje dotyczące analizowanych algorytmów, w szczególności tych, które zostały przez nią wybrane bądź zaprojektowane w końcowej konstrukcji klasyfikatora. Opisuje też wszelkie błędne tropy, na które natrafiła w czasie swych badań. Takie przedstawienie problemu sprawia, że pracę czyta się z zainteresowaniem. Także piękny język pracy zachęca do jej czytania. Ponadto, główna część pracy zawiera jedynie główny wywód prowadzący od sformułowania problemu do końcowego wyniku, z pominięciem dowodów, co także ułatwia odbiór rozprawy. Z kolei zaprezentowany dowód głównego wyniku matematycznego pracy, czyli twierdzenia 1 o istnieniu i jednoznaczności uzyskania wyniku przez algorytm SLS2S imponuje swą precyzją wywodu i przejrzystością prezentacji (choć jego długość jest trochę przerażająca).

Podstawowym zastrzeżeniem do pracy jest brak obliczeń numerycznych. W całej pracy autorka analizuje przykłady obliczeń, które są znakomitą ilustracją problemów decyzyjnych występujących we wszystkich rozważanych algorytmach, jednak w gruncie rzeczy mogą być przeprowadzone na kartce papieru. W pracy jest wprawdzie zamieszczony kod programu realizującego algorytm LC2S stanowiący podstawę budowy klasyfikatora hierarchicznego, jednak nie przedstawiono obliczeń wykorzystujących ten kod. Brak jest zarówno typowych obliczeń testowych (na wektorach próbnych, z weryfikacją poprawności wyniku, uwarunkowania warunków początkowych i wydajnością algorytmu), jak i rozwiązania jakiegoś zagadnienia praktycznego większej skali. Brak jest kodu podstawowego algorytmu SLS2S realizującego rozdzielanie liniowe dwóch zbiorów z maksymalnym prześwitem. Uwaga ta nie umniejsza wyników matematycznych uzyskanych w pracy przez autorkę, jednak brak obliczeń numerycznych sprawia, że temat pracy, mający ogromne znaczenie praktyczne, zaowocował rozprawą bardzo teoretyczną.

Jak to już wspomniałem, sposób prezentacji pracy, przejrzyste rozumowania i znakomicie dobrane, pod względem dydaktycznym, przykłady ilustracyjne sprawiają, że praca mogłaby być dobrym podręcznikiem z zakresu algorytmów rozdzielności liniowej. Dlatego też sugeruje autorce druk tej pracy, np. w pracach IPPT, w celu udostępnienia jej szerszej grupie odbiorców. (Można wówczas pracę uzupełnić o pełne specyfikacje algorytmów rozdzielności liniowej, o których autorka wspomina w pracy, lecz szerzej o nich nie pisze uznając za nieprzydatne do głównego celu rozprawy, jakim jest budowa klasyfikatora hierarchicznego.)

Reasumując stwierdzam, że rozprawa doktorska pani mgr Doroty Cendrowskiej, zatytułowana „Konstrukcja klasyfikatora obiektów z wykorzystaniem algorytmu badania rozdzielności liniowej dwóch zbiorów”, spełnia wymogi stawiane rozprawom doktorskim

przez USTAWĘ z dnia 14 marca 2003 r o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki, wraz z późniejszymi zmianami, Dz.U. z 2003 r. Nr 65, poz. 595; z 2005 r. Nr 164, poz. 1365 i może być dopuszczona do publicznej obrony.

4. Uwagi szczegółowe i redakcyjne

Ogólna uwaga dotycząca redakcji pracy wiąże się z założeniem, poczynionym przez autorkę o dobrej pamięci czytelnika. Brak jest zatem powtórzenie pewnych oznaczeń, założeń, itp. bezpośrednio w miejscu, gdzie są używane, jeśli były przyjęte we wcześniejszym miejscu w pracy. Przed ewentualnym drukiem rozprawy warto te założenia uzupełnić. Np., w rozdziale 2.2, str. 26, rozważany jest przykład liczbowy, w którym dane zostały zdefiniowane na rys. 2.1, str. 24, i tam też przyjęto ogólne założenie, że przykłady będą dotyczyć tych danych liczbowych. Później (do str. 30), brak do nich odwołania, co utrudnia lekturę pracy.

Na str. 11, punkt 1.3.1, zdanie zaczynające się od „Jeśli n cech opisujących...” powinno być połączone ze zdaniem następnym, aby całość miała sens.

Na str. 12, Definicja 1 – zamiast ”taką” powinno być „taka”.
Zdanie zaczynające się od „Chcąc uniknąć...” niepoprawnie sformułowane.

Na str. 14. przed pkt. 1.3.3, „naprzeciw” powinno być pisane łącznie.

Zamiast „dwoma nierównościami” powinno być „dwoma nierównościami”
„określoną” ... „opisaną”

Na str. 14, wzór (1.8). Brakuje

$t=f(x)=...$,

co sprawia, że następny wzór (1.9) nie jest w pełni zdefiniowany.

W całej pracy (począwszy od wzoru (1.8)) autorka zakłada, że $x \in X = X_1 \cup X_2$ (czasem jest to oznaczenie dziedziny operatora i wówczas taki zapis jest niezbędny. To założenie nie jest konieczne dla poprawności sformułowania problemu. Brak jest w tych wzorach założenia $X_1 \cap X_2 = \emptyset$, co jest niezbędne do poprawności sformułowania problemu klasyfikacji.

W całej pracy, zwrot „przy pomocy” należałoby zastąpić zwrotem „za pomocą”.

Pozostałe zauważone nieliczne drobne usterki (np. źle umieszczone przecinki, literówki, poprawki językowe, itp.) zaznaczyłem w swoim egzemplarzu pracy i mogę je udostępnić autorce w razie potrzeby.

Kotulski