

Dr hab. Krzysztof Diks, prof. UW
Instytut Informatyki
Uniwersytet Warszawski
ul. Banacha 2
02-097 Warszawa
e-mail: diks@mimuw.edu.pl

Warszawa, 30.03.2006

Recenzja rozprawy doktorskiej
pani mgr inż. Doroty Cendrowskiej
p.t. *Konstrukcja klasyfikatora obiektów
z wykorzystaniem algorytmu badania
rozdzielności liniowej dwóch zbiorów*

Problem klasyfikowania obiektów jest jednym z podstawowych problemów pojawiających się w analizie dużej ilości danych (obiektów). Zazwyczaj celem takiej analizy jest podział zbioru danych na podzbiory o zbliżonych własnościach oraz znalezienie takiej "algorytmicznej" charakterystyki tych podzbiorów, która pozwoli na automatyczne przypisywanie (klasyfikowanie) nowych obiektów do właściwych podzbiorów. Jedno z podejść do rozwiązania tego zagadnienia polega na potraktowaniu badanych obiektów jako punktów (wektorów) w wielowymiarowej przestrzeni euklidesowej, a następnie podzielenie tej przestrzeni na rozłączne obszary zawierające tylko obiekty o podobnych charakterystykach. Najprostszy klasyfikator jest liniowym klasyfikatorem dwu-decyzyjnym, co w podejściu opisanym powyżej odpowiada wyznaczeniu hiperpłaszczyzny dzielącej dwukolorowy zbiór punktów na dwa zbiory punktów o takich samych kolorach. Niestety nie zawsze taka hiperpłaszczyzna istnieje. Wówczas poszukuje się hierarchii hiperpłaszczyzn, które dzielą przestrzeń na obszary zawierające tylko punkty o tych samych kolorach. Wyznaczona hierarchia hiperpłaszczyzn powinna także umożliwiać łatwe i efektywne klasyfikowanie nowych obiektów do odpowiednich obszarów, a tym samym zaliczanie ich do zbioru punktów tego samego koloru, charakteryzujących się zbliżonymi własnościami. Klasyfikatory dwu-decyzyjne można następnie wykorzystywać do konstrukcji klasyfikatorów wielo-decyzyjnych. Omówione powyżej zagadnienia są właśnie przedmiotem rozprawy doktorskiej pan mgr inż. Doroty Cendrowskiej.

Rozprawa składa się z wprowadzenia, pięciu rozdziałów, podsumowania oraz dwóch dodatków.

We wprowadzeniu autorka krótko omawia, co jest przedmiotem rozprawy i opisuje zawartość pozostałej części pracy.

Rozdział pierwszy rozpoczyna się od omówienia istoty samej klasyfikacji oraz ogólnych zasad budowy klasyfikatorów i oceny ich jakości. Autorka także uzasadnia przydatność algorytmu liniowego rozdzielania dwóch zbiorów w konstrukcji hierarchicznego klasyfikatora odcinkowo-liniowego. Podrozdział 1.3 zawiera kluczowe dla pracy definicje liniowej rozdzielności dwóch zbiorów oraz krótki przegląd algorytmów testowania rozdzielności liniowej, jak i znajdowania "świadectwa" tej rozdzielności, czyli hiperpłaszczyzny (lub hiperpłaszczyzn) rozdzielającej (rozdzielających). W sformułowaniu definicji 2 pojawia się pewna niezręczność. Z jednej strony żąda się, żeby $g^*(x) > 0$, dla x ze zbioru $X1$, oraz $g^*(x) < 0$, dla x ze zbioru $X2$, a z drugiej strony dopuszcza się możliwość $g^*(x) = 0$. W definicjach 4 i 5 pojawia się tajemnicze a_{n+2} . Autorka twierdzi, że najbardziej interesujące jest wyznaczenie

K. D.

