

Wprowadzenie do sztucznej inteligencji - ćwiczenie 4

Kacper Marchlewicz

Zaimplementować klasyfikator ID3 (drzewo decyzyjne). Atrybuty nominalne, testy tożsamościowe. Podać dokładność i macierz pomyłek na zbiorach: Breast cancer i mushroom. Dlaczego na jednym zbiorze jest znacznie lepszy wynik niż na drugim? Do potwierdzenia lub odrzucenia postawionych hipotez konieczne może być przeprowadzenie dodatkowych eksperymentów ze zmodyfikowanymi zbiorami danych. Sformułować i spisać wnioski.

Wyniki działania klasyfikatora

Zgodnie z zaleceniem zbior danych został losowo podzielony na zbiór uczący i testujący w stosunku 3:2. Podana dokładność jest średnią z 25 uruchomień algorytmu.

Dla zbioru Breast Cancer:

Dokładność: 67,82%

Macierz pomyłek:

		Klasa rzeczywista	
		Recurrence-events	No-recurrence-events
Klasa predykowana	Recurrence-events	6	11
	No-recurrence-events	22	75

Dokładność dla zwiększonego (stosunek 6:1) zbioru uczącego: 68,71%

Dokładność dla zmniejszonego (stosunek 2:3) zbioru uczącego: 64,35%

Dla zbioru Mushroom

Dokładność: 99,23%

Macierz pomyłek:

		Klasa rzeczywista	
		edible	poisonous
Klasa predykowana	edible	1705	2
	poisonous	23	1520

Dokładność dla zwiększonego (stosunek 6:1) zbioru uczącego: 99,08%

Dokładność dla zmniejszonego (stosunek 2:3) zbioru uczącego: 99,04%

Wnioski

Dlaczego na jednym zbiorze jest znacznie lepszy wynik niż na drugim?

Zbiór Mushroom pozwolił na stworzenie drzewa decyzyjnego o wskaźniku dokładności o wiele większym od zbioru Breast Cancer. Zestaw danych o grzybach zawiera 2,5 razy więcej cech oraz 28 razy więcej instancji. Dzięki temu algorytm jest w stanie wygenerować dokładniejsze drzewo decyzyjne, co zaowocuje większą dokładnością. Odpowiedni dobór zbiorów uczących jest kluczowy do stworzenia dobrego klasyfikatora.