

Elementy Inteligencji Obliczeniowej - sprawozdanie

Projekt: Drzewa Decyzyjne nr 1

Marek Subocz 136806, Patryk Jedlikowski 136723
grupa wtorek, 15:10

20.10.2020r.

1 Opis Projektu

Wykonaliśmy program projektujący drzewo decyzyjne dla danych zapisanych w pliku csv. Pliki wczytywane są za pomocą biblioteki Pandas, drzewo decyzyjne jest wyświetlane za pomocą biblioteki ete3.

Aby uruchomić program należy zainstalować biblioteki pandas, numpy i ete3:

- pip install pandas
- pip install numpy
- pip install ete3

A następnie wykonać polecenie:

```
python drzewo_decyzyjne.py
```

2 Opis algorytmu

Algorytm składa się z 7 funkcji:

1. entropy (liczenie entropii pojedynczego atrybutu)
2. conditional_entropy
3. information_gain
4. intrinsic_info
5. gain_ratio
6. set_treshold
7. construct_tree

Pierwsze pięć funkcji działa według wzorów opisanych na zajęciach. Do obliczenia gain_ratio atrybutu w postaci zmiennej ciągłej używamy funkcji set_treshold. Do zbudowania drzewa decyzyjnego używamy funkcji construct_tree.

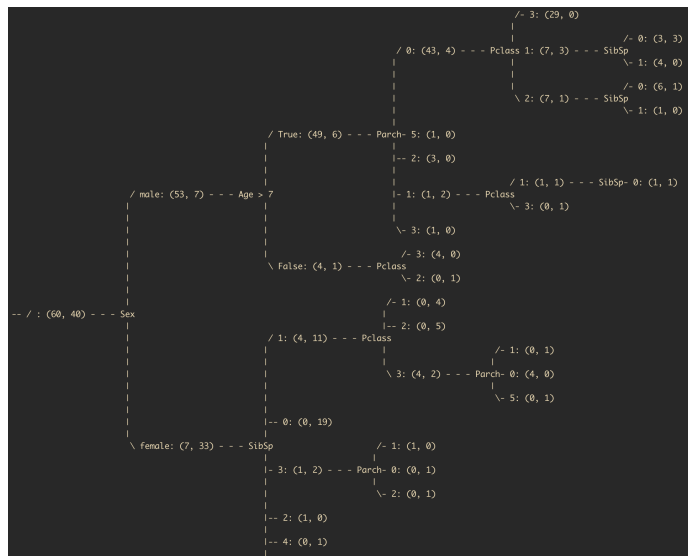
2.1 Set treshold

Funkcja set_treshold dzieli zbiór "Age" na dwa podzbiory. Aby wyznaczyć granicę między nimi, algorytm wylicza wartość gain_ratio dla każdej możliwej granicy (dla szybszego działania ograniczamy się do co trzeciej wartości) i wybiera granicę, dla której ta wartość jest najlepsza.

2.2 Construct tree

Funkcja construct_tree konstruuje drzewo decyzyjne w sposób rekurencyjny. Dla danego zbioru danych wybiera atrybut, dla którego wartość gain_ratio jest największa i dzieli drzewo na poddrzewa zawierające tę samą wartość wybranego atrybutu. Następnie cała funkcja jest uruchamiana dla każdego z tych poddrzew, z danych usuwana jest kolumna dotycząca wybranego atrybutu. Warunkiem zatrzymania rekurencji jest brak atrybutów do rozpatrzenia lub jednoznaczne zakwalifikowanie każdego przykładu do klasy decyzyjnej.

3 Wyniki



Algorytm określił płeć jako najważniejszy atrybut decydujący o przeżyciu danej osoby, co zgadza się z naszymi obserwacjami. Nie udało się przyporządkować każdej osoby do danej klasy decyzyjnej, ale niepewnych elementów pozostało bardzo mało.