

System wspomagania diagnostyki medycznej

Systemy wspomagania decyzji – projekt

Zuzanna Pacholczyk, 230709
Dawid Robaczyński, 194860

15 stycznia 2016

1 Założenia wstępne

Celem projektu jest stworzenie aplikacji wspomagającej diagnostykę medyczną w zakresie wykrywania raka piersi. Aplikacja wykorzystuje dane dostarczone przez użytkownika lub wcześniej przygotowane (dane zaczerpnięte z UCI Machine Learning Repository), dostarczane wraz z aplikacją w celu zbudowania klasyfikatora wykorzystywanego do określenia prawdopodobieństwa występowania raka piersi. Program pozwala użytkownikowi na samodzielne wprowadzenie wartości szeregu atrybutów i obliczenie na ich podstawie prawdopodobieństwa choroby. Wynik zostaje zaprezentowany jako wartość liczbową wraz z drzewem decyzyjnym, które doprowadziło do uzyskania wyniku.

2 Technologia

Projekt został zaimplementowany w języku Java w wersji 8 przy wykorzystaniu framework'a JavaFX w celu utworzenia interfejsu graficznego oraz biblioteki Weka pozwalającej na wykorzystanie algorytmu eksploracji danych (ang. data mining).

3 Wykorzystane modele i algorytmy

Dla naszego problemu charakterystyczna jest niewielka liczba atrybutów: 9. Wszystkie z nich to atrybuty numeryczne przyjmujące wartości całkowitoliczbowe z zakresu 1-10, dlatego wybór odpowiedniego modelu chcieliśmy jak najlepiej dopasować do tej charakterystyki.

Pierwszym z analizowanych modeli był Random Forest wykorzystujący kilka drzew decyzyjnych w celu zbudowania klasyfikatora. Opiera się o tzw. bagging (pakowanie do worków) w połączeniu z losowym wyborem ustawień.

Model ten został jednak przez nas odrzucony ze względu na zbyt skomplikowaną wizualizację, która nie byłaby użyteczna na niedoświadczonego użytkownika oraz zbyt skomplikowane ustawienia dla potrzeb naszego programu.

Model, który wybraliśmy, to J48 wykorzystujący algorytm C4.5, który buduje pojedyncze drzewo decyzyjne. Doskonale sprawdza się w zastosowaniu dla atrybutów numerycznych, pozwala na estetyczną wizualizację powstałego drzewa i daje wiarygodne wyniki.

4 Działanie aplikacji

Użytkownik ma możliwość podania ścieżki do pliku z danymi, na podstawie których ma zostać zbudowany klasyfikator. Jeśli informacje te nie zostaną dostarczone, wykorzystane zostaną dane zaczerpnięte z UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+>

Następnie użytkownik ma możliwość wprowadzenia wartości atrybutów:

- grubość warstwy,
- rozmiar komórki równomierny,
- kształt komórki równomierny,
- marginalna przyczepność komórki,
- tozmiar tkanki nabłonkowej,
- jądra komórkowe,
- łagodna chromatyna,
- normalne jądra komórki,
- mitozę.

Atrybuty te przyjmują wartości od 1 do 10 i są one wybierane przy pomocy suwaków.

Po wybraniu przycisku "Oblicz", następuje wyznaczenie prawdopodobieństwa występowania raka piersi, a informacje wynikowe zostają wyświetlone w postaci wartości procentowej i schematu drzewa decyzyjnego, które doprowadziło do tego wyniku.