

article polski [utf8]inputenc [T1]fontenc graphicx [hscale=0.7,vscale=0.8]geometry
System wspomagania diagnostyki medycznej
Zuzanna Pacholczyk, 230709
document
Zaoenia wstpne

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 wystąpienia 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.

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).

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 wybraliśmy odpowiedniego modelu chcąc 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 dla niedoświadczonych użytkowników 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.

Działanie aplikacji

Użytkownik ma możliwość podania ciekich danych do pliku, 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: itemize

- g grubość warstwy,
- r rozmiar komórki równomierny,
- k kształt komórki równomierny,
- m średnia przyczepność komórki,
- t rozmiar tkanki nabłonkowej,
- j jądra komórkowe,
- agoodna chromatyna,
- n normalne jądra komórki,
- m mitoz.