

Narzędzie eksploracji danych

Kornel Lewandowski, Małgorzata Napieraj

Opis projektu

Proste w obsłudze narzędzie do eksploracji danych medycznych, które umożliwia tworzenie klasyfikatorów dla zawartych w danych wzorców, wstępne przetwarzanie danych i atrybutów, import i eksport danych do popularnych formatów plików oraz łatwy w obsłudze i przyjazny interfejs.

Koncepcja realizacji

Aplikacja działająca na wielu platformach

Przejrzysty interfejs oparty o przeglądarkę

Nie chcemy odkrywać koła na nowo

Specyfikacja

<i>Lp</i>	<i>Opis funkcjonalności</i>	<i>Typ</i>	<i>Status</i>
1	Normalizacja i dyskretyzacja danych	obowiązkowa	
2	Selekcja atrybutów	obowiązkowa	
3	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>NaiveBayes</i>)	obowiązkowa	
4	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>J48</i>)	obowiązkowa	
5	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>IBk</i>)	obowiązkowa	
6	Klasyfikacja przykładów na zbudowanym klasyfikatorze	obowiązkowa	
7	Manipulacja parametrami algorytmów klasyfikacji	obowiązkowa	
8	Import danych w formacie XLS/CSV/ARFF	obowiązkowa	
9	Eksport danych w formacie XLS/CSV/ARFF	obowiązkowa	
10	Pomoc kontekstowa	obowiązkowa	
11	Import/eksport modelu klasyfikatora	opcjonalna	
12	Import danych w formacie XML	opcjonalna	
13	Eksport danych w formacie XML	opcjonalna	
14	Udostępnienie dodatkowego algorytmu selekcji atrybutów (<i>GainRatio</i>)	opcjonalna	
15	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>DTNB</i>)	opcjonalna	

Stopień realizacji

Lp	Opis funkcjonalności	Typ	Status
1	Normalizacja i dyskretyzacja danych	obowiązkowa	100,00%
2	Selekcja atrybutów	obowiązkowa	100,00%
3	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>NaiveBayes</i>)	obowiązkowa	100,00%
4	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>J48</i>)	obowiązkowa	100,00%
5	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>IBk</i>)	obowiązkowa	100,00%
6	Klasyfikacja przykładów na zbudowanym klasyfikatorze	obowiązkowa	100,00%
7	Manipulacja parametrami algorytmów klasyfikacji	obowiązkowa	100,00%
8	Import danych w formacie XLS/CSV/ARFF	obowiązkowa	100,00%
9	Eksport danych w formacie XLS/CSV/ARFF	obowiązkowa	100,00%
10	Pomoc kontekstowa	obowiązkowa	100,00%
11	Import/eksport modelu klasyfikatora	opcjonalna	0,00%
12	Import danych w formacie XML	opcjonalna	0,00%
13	Eksport danych w formacie XML	opcjonalna	0,00%
14	Udostępnienie dodatkowego algorytmu selekcji atrybutów (<i>GainRatio</i>)	opcjonalna	0,00%
15	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>DTNB</i>)	opcjonalna	100,00%

Interfejs

MediMiner

Simple data-mining tool for medical purposes

Kornel Lewandowski & Malgorzata Napieraj, 2014

Classifier is ready.

Import data

Export to ARFF

Export to CSV

Export to XLS

Close

Select attributes...DiscretizeNormalizeHide unselected

PatientId	Mer1	Mer2	Mer3	Diagn
1	'(27.5-66.5]'	72	95	diarrhoea
2	'(27.5-66.5]'	62	8	flu
3	'(-inf-27.5]'	9	46	flu
4	'(-inf-27.5]'	52	32	flu
5	'(27.5-66.5]'	9	3	flu
6	'(-inf-27.5]'	18	45	flu
7	'(-inf-27.5]'	23	77	flu
8	'(66.5-inf)'	52	77	cold
9	'(27.5-66.5]'	7	51	flu
10	'(27.5-66.5]'	34	20	flu
11	'(27.5-66.5]'	12	14	flu

Naive Bayes Classifier

Estimate kernel?
Yes ▾

Build

J48 Classifier

Confidence level
0.01 ▾

Classify

Rebuild

IBk Classifier

Neighbours
1 ▾

Build

DTNB Classifier

Evaluation measure
RMSE ▾

Build

If you don't know how any option works, move cursor on the button and wait for tooltip (little comic's bubble with description).

MediMiner a WEKA Explorer - różnice

Przyjazny interfejs

Pomoc kontekstowa

Podgląd danych

Import/eksport danych w formacie XLS

Napotkane problemy

UTF8 \longleftrightarrow ANSI: polskie znaki

Normalizacja - wszystkie atrybuty

Atrybut decyzyjny typu liczbowego

Dyskretyzacja atrybutu decyzyjnego

Instancje: reprezentacja danych oraz ich etykiet

Co dalej?

Odczyt/zapis modelu klasyfikatora

Odczyt/zapis wyników klasyfikacji

Dodawanie/usuwanie instancji ze zbioru treningowego

Udostępnienie dodatkowych funkcji WEKI

Technologie

Java

Maven

Weka

HTML + CSS

JavaScript

jQuery

Jetty

Prezentacja aplikacji