

Systemy Uczące się laboratorium

Ćwiczenie 2. Algorytm ILA

opracował: P. Myszkowski

Cel ćwiczenia

Zapoznanie się z metodą indukcją reguł ILA (*Inductive Learning Algorithm*) na przy samodzielnej implementacji (np. Java/C/C++/C#).

Realizacja ćwiczenia

- Zapoznanie się z regułową metodą reprezentacji wiedzy
- Wybór trzech zbiorów danych (mogą być te same co na poprzednich zajęciach + „pogodowy” przykład z wykładu dla sprawdzenia poprawności implementacji; trzeba pamiętać o ew. transformacjach dziedzin atrybutów). Dla przypomnienia 2/3 (glass, wine, diabetes) + dwa wybrane przez studenta (zbiory iris lub weather możemy używać tylko dla sprawdzenia poprawności kodu).
- Przebadanie działania metody ILA na wyżej wymienionych zbiorach
- Przebadanie wpływu zbioru/atributów/wartości na skuteczność/efektywność metody
- Sporządzenie sprawozdania z ćwiczenia

Informacje pomocnicze

Proces pozyskiwania wiedzy z baz danych (KDD, *Knowledge Discovery in Databases*) jest jednym z ważniejszych zastosowań metod sztucznej inteligencji. W tym procesie najbardziej interesuje nas etap *data mining* (drażenie danych), które zawężamy do zadania klasyfikacji. Dodatkowo skupiamy się tylko na jednej konkretnej metodzie, jaką jest indukcja reguł.

Obowiązujące miary: Acc, Prec, Rec i Fsc. Dodatkowo, wybór innych trzeba uzasadnić.

Zadanie polega na implementacji algorytmu regułowego ILA, który na podstawie zbiorów danych z zbuduje reguły je opisujące. Ćwiczenie zakłada pracę z trzema zbiorami.

Krosvalidacja (walidacja krzyżowa, *crossvalidation*) jest to sposób na podział danych na zbiory uczące i testowe w taki sposób, aby zminimalizować ich wpływ na obserwowane wyniki. Zwykle stosuje się krosvalidacje 2, 3 lub 5, rzadziej 10. Tak jak w poprzednim zadaniu.

Przydatne linki

- dane można użyć z poprzedniego ćwiczenia (UCL Repository) - <http://archive.ics.uci.edu/ml/>
Trzeba pamiętać o dostosowaniu danych do algorytmu.
- Artykuł opisujący działanie i użycie ILA
<http://www.dis.uniroma1.it/~sassano/STAGE/LearningRule.pdf>

Ocena ćwiczenia (max 10pkt)

2pkt	Implementacja algorytmu ILA
1pkt	Implementacja i testowanie 2 różnych metod <u>dyskretyzacji</u>
1pkt	Krótki opis działania algorytmu ILA
2pkt	Zbadanie działania algorytmu ILA na 3 wybranych zbiorach
3pkt	Porównanie działania algorytmu przy różnych podziałach danych – tabelki, wnioski
1pkt	Porównanie działania algorytmu – graficzne przedstawienie uzyskanych wyników

Uwaga! Przy testach proszę pamiętać o krosvalidacji

Literatura

- 1.Cichosz P. "Systemy uczące się", WNT Warszawa
- 2.Zasoby Internetu: uczenie maszynowe (*machine learning*), *data mining*, reguły

W razie pytań, uwag czy komentarzy prosimy o kontakt z prowadzącym zajęcia.

P. Myszkowski