

Projekt z Zastosowań informatyki w medycynie

Kolejne etapy realizacji projektu

1. Zapoznać się z algorytmami diagnostycznymi (algorytmami klasyfikacji), określonymi w temacie:

- (a) **sieci neuronowe** — sieć jednokierunkowa, z jedną warstwą ukrytą, uczona metodą wstecznej propagacji błędów (z członem momentum i bez członu momentum) [3, 7, 8, 9],
- (b) **algorytmy minimalno-odległościowe** — algorytm najbliższa średnia (NM – nearest mean) i algorytm k-najbliższych sąsiadów [1, 2, 4, 5],
- (c) **empiryczny naiwny algorytm bayesowski** — algorytm Bayesa z prawdopodobieństwami a priori oraz prawdopodobieństwami cech w poszczególnych klasach (dla cech dyskretnych) szacowanymi (estymowanymi) ze zbioru uczącego metodą częstotliwościową i/lub warunkowymi gęstościami cech w klasach (dla cech ciągłych) szacowanymi metodą histogramu, metodą empirycznej dystrybucyjności lub metodami jądrowymi [1, 2, 4, 5, 6].

2. Zapoznać się z materiałem empirycznym – zdefiniować problem rozpoznawania (klasyfikacji) – określić liczbę i znaczenie klas, liczbę i znaczenie cech oraz charakter cech (ciągłe, wielowartościowe, binarne, itp.).

3. Wyznaczyć ranking cech pod względem ich przydatności do klasyfikacji korzystając z dowolnej miary (kryterium) jakości cech stosowanych w selekcji cech (są to metody selekcji cech zwane filtrami)¹ [1, 2, 10].

4. Zaplanować badania eksperymentalne dla następujących założeń

- (a) Trenowanie i testowanie zastosowanych klasyfikatorów z wykorzystaniem 5 razy powtarzanej metody 2-krotnej walidacji krzyżowej (5x2cv)².
- (b) Przeprowadzić badania dla różnej liczby cech (poczynając od jednej – najlepszej wg. wyznaczonego rankingu, a następnie dokładać kolejno po jednej (również według wyznaczonego rankingu) tak długo, aż zostanie znaleziona najlepsza liczba cech. Dodawanie cech powinno poprawiać jakość klasyfikacji, ale do pewnego momentu – dalsze dodawanie cech jakości pogorszy. Trzeba znaleźć optimum.
- (c) Przeprowadzić badania dla następujących algorytmów:
 - Dla sztucznych sieci neuronowych – sieć jednokierunkowa z 1 warstwą ukrytą dla 3 różnych liczb neuronów w warstwie ukrytej oraz dla uczenia metodą propagacji wstecznej z momentum i bez momentum.
 - Dla metod minimalno-odległościowych – algorytmy NM, 1-NN, k-NN (dla 2 różnych wartości k), dla dwóch różnych miar odległości (np. Euklidesowa, Manhattan) oraz z normalizacją i bez normalizacji.

- Dla empirycznego naiwnego algorytmu bayesowskiego – algorytm dla 0-1 funkcji strat z założeniem, że cechy są niezależne.

(d) Dla każdego pojedynczego eksperymentu (pojedynczy eksperyment to doświadczalne wyznaczenie jakości klasyfikacji (patrz punkt 4a) dla danego algorytmu, danych wartości parametrów algorytmu i dla danej liczby cech) należy przedstawić wyniki (jakości klasyfikacji) w formie uśrednionej (względem 5 powtórzeń metody 2-krotnej walidacji krzyżowej) oraz dla jednej (dowolnie wybranej) macierzy konfuzji.

5. Zaimplementować algorytmy diagnostyczne (klasyfikacji) (środowisko implementacji dowolne), tak aby można było przeprowadzić badania eksperymentalne według przedstawionych założeń.

6. Zrealizować badania eksperymentalne według przedstawionych w punkcie 4 założeń.

7. Przeprowadzić dyskusję wyników i przedstawić wnioski.

Zawartość sprawozdania

1. Charakterystyka analizowanego problemu (jako zadania rozpoznawania).
2. Opis stosowanych algorytmów.
3. Informacja o środowisku implementacyjnym.
4. Opis badań eksperymentalnych.
5. Wyniki (w formie graficznej, tabeli, itp.).
6. Dyskusja wyników i wnioski wynikające z badań.

Literatura

1. M. Kurzyński, Rozpoznawanie obiektów – metody statystyczne, Oficyna Wydawnicza Pol. Wr., Wrocław 1997
2. K. Stapor, Metody klasyfikacji obiektów w wizji komputerowej, PWN, Warszawa 2011
3. L. Rutkowski, Metody i techniki sztucznej inteligencji, PWN, Warszawa 2005
4. R. Tadeusiewicz, M. Flasiński, Rozpoznawanie obiektów, PWN, Warszawa 1991
5. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WNT, Warszawa 2005
6. M. Krzyśko, Systemy uczące się, WNT, Warszawa 2008
7. J. Łęski, Systemy neuronowo-rozmyte, PWN, Warszawa 2004
8. S. Osowski, Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Pol. Warszawskiej, Warszawa 2006
9. Sieci Neuronowe, seria: Biocybernetyka i Inżynieria Biomedyczna, tom 6, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000
10. W. Sobczak, W. Malina, Metody redukcji i selekcji informacji, WNT, Warszawa 1987

¹Proponuję kryterium Kolmogorowa, gdyż jest bardzo proste rachunkowo.

²Metoda 2cv polega na tym, że dostępny zbiór danych dzielimy (losowo) na dwie jednakowe części (jedna część stanowi zbiór uczący, a druga zbiór testujący) i wykonujemy badania. Następnie powtarzamy badania dla odwrotnej sytuacji – zbiór, który był uczącym teraz jest testującym, a testujący – uczącym. 5x2cv oznacza, że procedurę powtarzamy 5 razy (każdorazowo na nowo losujemy dwie połowy) i wyniki uśredniamy.

Lista tematów

Komputerowe wspomaganie diagnozowania ...

1. ...białaczek u dzieci z wykorzystaniem sztucznych sieci neuronowych
2. ...białaczek u dzieci z wykorzystaniem empirycznego algorytmu bayesowskiego
3. ...białaczek u dzieci z wykorzystaniem algorytmów minimalno-odległościowych
4. ...choroby niedokrwiennej u dzieci z wykorzystaniem sztucznych sieci neuronowych
5. ...choroby niedokrwiennej u dzieci z wykorzystaniem empirycznego algorytmu bayesowskiego
6. ...choroby niedokrwiennej u dzieci z wykorzystaniem algorytmów minimalno-odległościowych
7. ...stanów ostrego brzucha z wykorzystaniem sztucznych sieci neuronowych
8. ...stanów ostrego brzucha z wykorzystaniem empirycznego algorytmu bayesowskiego
9. ...stanów ostrego brzucha z wykorzystaniem algorytmów minimalno-odległościowych
10. ...zawałów z wykorzystaniem sztucznych sieci neuronowych
11. ...zawałów z wykorzystaniem empirycznego algorytmu bayesowskiego
12. ...zawałów z wykorzystaniem algorytmów minimalno-odległościowych
13. ...nowotworów piersi z wykorzystaniem sztucznych sieci neuronowych
14. ...nowotworów piersi z wykorzystaniem empirycznego algorytmu bayesowskiego
15. ...nowotworów piersi z wykorzystaniem algorytmów minimalno-odległościowych
16. ...chorób tarczycy z wykorzystaniem sztucznych sieci neuronowych
17. ...chorób tarczycy z wykorzystaniem empirycznego algorytmu bayesowskiego
18. ...chorób tarczycy z wykorzystaniem algorytmów minimalno-odległościowych
19. ...ostrego zapalenia dróg moczowych z wykorzystaniem empirycznego algorytmu bayesowskiego
20. ...ostrego zapalenia dróg moczowych z wykorzystaniem algorytmów minimalno-odległościowych
21. ...ostrego zapalenia dróg moczowych z wykorzystaniem sztucznych sieci neuronowych
22. ...zapalenia wątroby z wykorzystaniem empirycznego algorytmu bayesowskiego
23. ...zapalenia wątroby z wykorzystaniem algorytmów minimalno-odległościowych
24. ...zapalenia wątroby z wykorzystaniem sztucznych sieci neuronowych
25. ...chorób wątroby z wykorzystaniem empirycznego algorytmu bayesowskiego
26. ...chorób wątroby z wykorzystaniem algorytmów minimalno-odległościowych
27. ...chorób wątroby z wykorzystaniem sztucznych sieci neuronowych

Zbiory danych

1. **Tematy 1–12** — <https://bit.ly/2HGmfZx>
2. **Tematy 13–15** — <http://archive.ics.uci.edu/ml/datasets.html> (Breast Cancer Wisconsin (Original) Data Set)
3. **Tematy 16–18** — <http://archive.ics.uci.edu/ml/datasets.html> (Thyroid disease data set)
4. **Tematy 19–21** — <http://archive.ics.uci.edu/ml/datasets.html> (Acute Inflammations)
5. **Tematy 22–24** — <http://archive.ics.uci.edu/ml/datasets.html> (Hepatitis)
6. **Tematy 25–27** — <http://archive.ics.uci.edu/ml/datasets.html> (ILPD Indian Liver Patient Dataset)