

Биомедицинские приборы и системы

УДК 004.9

Г.С. Порєва, Д. Гончарова

Національний технічний університет України «Київський політехнічний інститут»,
вул. Політехнічна, 16, учб. корпус 12, ауд. 412а, м.Київ, 03056, Україна.

Дослідження роботи класифікаторів для оптимізації постановки діагнозів бронхолегеневих захворювань

У статті розглянуто можливість застосування класифікаторів, що є основою машинного навчання, для оптимізації постановки діагнозів бронхолегеневих захворювань. Розглянуто роботу декількох класифікаторів, і в результаті дослідження для поставленого завдання обраний класифікатор на основі методу найближчих сусідів. В якості параметрів методу обрані розраховані на основі поліспектрального аналізу чисельні характеристики сигналів звуків дихання. Встановлено, що даний класифікатор є простим для реалізації і роботи з базою звуків дихання. Отримана точність роботи класифікатора є досить високою. Розроблений алгоритм покликаний істотно спростити роботу лікаря-пульмонолога для постановки своєчасного діагнозу. Бібл. 8, рис. 3.

Ключові слова: звуки дихання; класифікатор; метод k-найближчих сусідів; поліспектральний аналіз.

Вступ

Звуки легень є цінними індикаторами стану респіраторної системи людини. В даний час в світі спостерігається швидке зростання захворювань дихальної системи. Це обумовлює актуальність своєчасної діагностики респіраторних захворювань. Аускультация шумів (звуків) дихання при дослідженні хворого залишається основним неінвазивним методом діагностики респіраторних захворювань. У той же час розвиток цифрових технологій, а також сучасних методів і алгоритмів обробки інформації призводить до постійної появи нових методів аналізу дихальних шумів. Завдяки цьому зросли можливості об'єктивізації аускультативних ознак, тобто кількісної оцінки відповідності зареєстрованих звуків з функціональним станом систем організму. Не дивлячись на різноманітність існуючих методів, не існує єдиного об'єктивного підходу до аналізу звуків легень, а тому дослідження в даній області тривають.

Важливим завданням є класифікація звуків дихання за певними категоріями, що в подальшому дозволить лікарю більш точно встановлювати діагноз. Для вирішення даного завдання широко використовується апарат машинного навчання - великий підрозділ штучного інтелекту, який вивчає методи побудови алгоритмів, здатних навчатися. Машинне навчання знаходиться на стику математичної статистики, методів оптимізації та класичних математичних дисциплін, але має також і власну специфіку, пов'язану з проблемами обчислювальної ефективності і перенавчання [1,2].

Метод k найближчих сусідів або k nearest neighbours - один з найпростіших для реалізації методів машинного навчання. Він дозволяє класифікувати нові дані, ґрунтуючись на наявних [3]. Метод опорних векторів є одним з найвідоміших алгоритмів машинного навчання, що застосовуються в основному для завдання класифікації [4]. Дерево прийняття рішень - засіб підтримки прийняття рішень, що використовується в статистиці та аналізі даних для прогнозних моделей. Класифікація є одним із завдань, де використовуються дерева прийняття рішень. Ці методи і досліджені в даній роботі для класифікації звуків дихання. Результати класифікації можуть цілком реально допомогти лікарю при постановці діагнозу пацієнтам.

Метод k-найближчих сусідів

У попередніх роботах авторів [5-7] було показано, що для виявлення характерних діагностичних ознак дихання ефективним є застосування апарату статистики вищих порядків. Досліджувалися сигнали звуків дихання, що синхронно реєструвалися чотирьохканальним комплексом КОРА з відповідних точок грудної клітини людини. В результаті розраховувалися і аналізувалися такі параметри, як коефіцієнт асиметрії (skewness), коефіцієнт ексцесу (kurtosis), максимальне значення функції бікоге-

рентності (max: bic), середнє за чотирма каналами значення максимальної бїчастоти (f_mean), середнє за чотирма каналами значення бїспектра (v_mean).

Для спрощення аналізу та, вїдповїдно, постановки дїагнозу було запропоновано застосування апарату машинного навчання, що дозволить при бажаннї аналізувати не кожен з параметрїв окремо, а вїдразу видавати дїагностичний результат на основї аналізу всїх досліджувальних параметрїв одночасно.

Класифїкатори використовували базу з 120 пацїєнтїв, 51 з яких лїкарем був поставлений дїагноз хронїчне обструктивне захворювання легень (ХОЗЛ), 35 - практично здоровї пацїєнти і 34 - контрольна група пацїєнтїв, якї працюють шахтарями, але без видимих патологїй бронхолегеневої системи.

Як було вказано, досліджувалася робота трьох класифїкаторїв. В результатї найкращий результат показав метод k-найближчих сусїдїв. Тому програма для попереднього встановлення дїагнозу була розроблена на основї даного методу.

Метод k-найближчих сусїдїв (knn) - метод, що заснований на використаннї пам'ятї. Робота методу заснована на інтуїтивному припущеннї про те, що близько розташованї об'єкти, швидше за все, належать однїй категорїї. Метод k-найближчих сусїдїв дозволяє класифїкувати новї данї, ґрунтуючись на наявних.

Алгоритм може бути застосований до вїбїрок з великою кїлькїстю атрибутїв (багатовимїрних). Для цього перед застосуванням потрїбно визначити функцїю дистанцїї. Класичний

варїант визначення дистанцїї - дистанцїя в евклїдовому просторї. Спочатку встановлюється навчальна вїбїрка пар «об'єкт-клас».

$$X^m = \{(x_1, y_1), \dots, (x_m, y_m)\} \quad (1)$$

Це можуть бути об'єкти, якї ранїше вже були класифїкованї, або об'єкти, якї класифїкують експерти. На безлїчї об'єктїв задається метрика. Для довїльного об'єкта u розташуємо об'єкти навчальної вїбїрки Xї в порядку зростання вїдстанї до u

$$\rho(u, x_{1,u}) \leq \rho(u, x_{2,u}) \leq \dots \leq \rho(u, x_{m,u}), \quad (2)$$

де через xї;u позначається той об'єкт навчальної вїбїрки, який є i-м сусїдом об'єкта u. Так само пронумеровуються класи, до яких вїдносяться xї;u – позначимо їх yї;u. Алгоритм найближчих сусїдїв в загальному виглядї можна виразити формулою

$$a(u) = \operatorname{argmax}_{y \in Y} \sum_{i=1}^m [y(x_{i,u}) = y] \omega(i, u) \quad (3)$$

де ω - вагова функцїя, що показує, чи важливий для об'єкта u елемент пїд номером i. Для методу k найближчих ω(i, u) = [1 <= k]. Таким чином, алгоритм класифїкацїї методом найближчих k сусїдїв дуже простий. Є вїбїрка з даних - навчальна множина, для кожного елемента якої вїдомо, до якого класу вїн належить.

Для кожного наступного елемента невідомого класу береться вїбїрка з k найближчих елементїв, клас яких вже вїдомий. Новому елементу привласнюється той клас, елементїв котрого виявилось найбільше в вїбїрцї [8].

Приклад роботи програми на основї knn показаний на рис. 1.

```
Would you like to remove and check this patients from database to classify,
or just predict new patients? Tips: 0 - remove and check 1 - check new 2 - if you want to skip this process
Write answer: 0

Write a number how many patients you want to delete from database and predict him:2
Write a number of patient in list from database: 2
Write a number of patient in list from database: 31

Work with patients have been ended

Results of Prediction:

1. healthy
2. khozl>>
```

Рис. 1. Приклад роботи програми класифїкатора на основї методу knn

При роботі з методом k-найближчих сусідів дуже важливо правильно вибрати параметр k, так як саме від цього параметра в основному залежить якість роботи класифікатора. Тому були оцінені точності прийняття правильного рішення класифікатора при різних значеннях k, в діапазоні від 3 до 23. При k=5, точність класифікатора виявилася найвищою (рис.2).

Класифікатор при розрахунках може використовувати різні види відстаней. Тому залежність-ність точності від k була розрахована при евклідовій відстані, відстані Махаланобіса, відстані по Манхеттену, косинусній відстані і відстані Хеммінга. Для даного випадку, найбільша точність отримана при використанні класифікатором евклідової відстані.

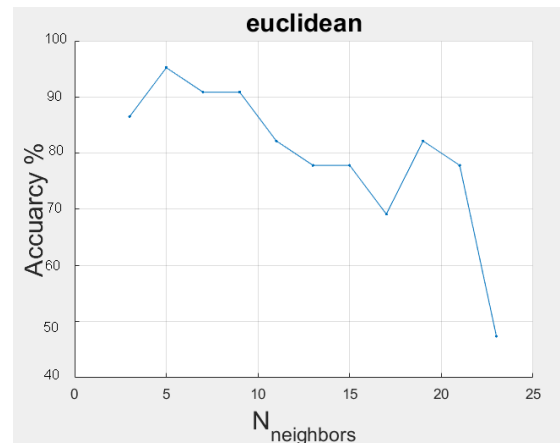


Рис. 2. Залежність точності від кількості найближчих сусідів

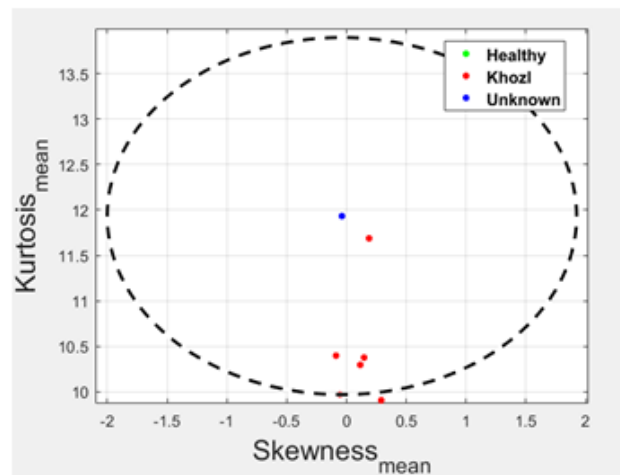
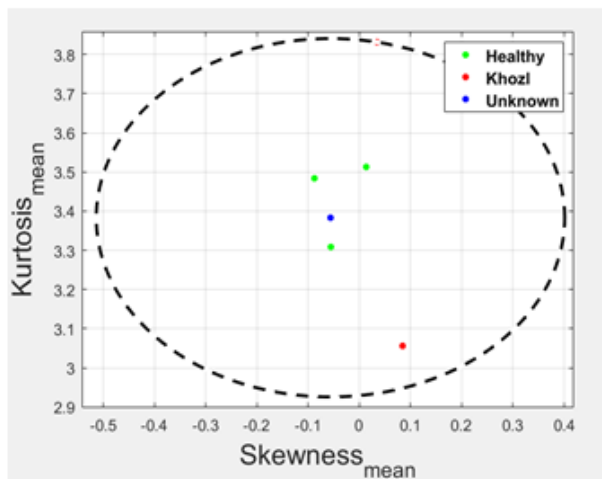


Рис. 3. Ілюстрація результату роботи методу knn для звуків легень

На рис. 3 показаний приклад роботи класифікатора для двох параметрів - коефіцієнта асиметрії (skewness) і коефіцієнта ексцесу (kurtosis) при k= 5. Як було сказано вище, робота розробленого класифікатора була побудована на п'яти параметрах сигналів звуків дихання, що підвищує об'єктивність отриманого результату. Точність роботи даного алгоритму становить 96%, що є досить прийнятним результатом.

Висновки

В даному дослідженні розроблений алгоритм на основі машинного навчання, що дозволяє пульмонологу більш якісно проводити класифікацію пацієнтів по категоріям здоровий - хворий ХОЗЛ. Для початку розраховується ряд параметрів на основі поліспектрального аналізу, які самі по собі вже несуть діагностичну цінність. Однак для спрощення роботи з базою параметрів, а також для об'єктивізації постановки діагнозу пропонується простий і доступний метод

програмної класифікації. Досліджено три різних види класифікатора, і в результаті встановлено, що найкращий результат показав метод найближчих сусідів. Отримана в результаті точність є досить високою. У подальших дослідженнях планується дослідження роботи класифікатора для ще більш великої кількості параметрів звуків дихання, а також провести навчання даного класифікатора для роботи з іншими базами даних пацієнтів.

Список використаних джерел

1. Золотых Н.Ю. Machine Learning and Data Mining // URL: <http://www.uic.unn.ru/~zny/ml/>
2. Воронцов К. В. Математические методы обучения по прецедентам (теория обучения машин) // <http://www.ccas.ru/voron>
3. Mitchell T. Machine Learning. — McGraw-Hill Science/Engineering/Math, 1997. ISBN 0-07-042807-7.

4. Классификация данных методом опорных векторов URL: <https://habrahabr.ru/post/105220/>
5. *Порева Г.С., Карплюк Є.С., Макаренкова А.А., Макаренко А.П.* Виявлення характерних акустичних ознак пацієнтів з ХОЗЛ на основі біспектрального аналізу звуків дихання // Electronics and Communications. – 2014. - Volume 19, №6(83) – сс. 82-86
6. *Порева А.С., Макаренкова А.А., Карплюк Є.С., Гончаренко А.А.* Применение полиспектрального анализа для определения диагностических признаков в звуках дыхания больных ХОБЛ - Вісник Національного Технічного Університету "ХПІ" - збірник наукових праць. Серія : Нові рішення в сучасних технологіях. - Х.: НТУ "ХПІ" - 2014р. - , №36(1079)2014 - 200с., с. 49-55.
7. *Anna Poreva Yevgeniy Karplyuk, Anastasiia Makarenkova, Anatoliy Makarenkov* Detection of COPD's Diagnostic Signs Based on Polyspectral Lung Sounds Analysis of Respiratory Phases 2015 IEEE 35th International Scientific Conference Electronics and Nanotechnology (ELNANO), pp.351-355
8. Метод к ближайших соседей на Python, URL: <http://glebmikhaylov.com>

Поступила в редакцию 19 июля 2016 г.

УДК 004.9

А.С. Порева, Д. Гончарова

Национальный технический университет Украины «Киевский политехнический институт», ул. Политехническая, 16, уч. корпус 12, ауд. 412а, г. Киев, 03056, Украина.

Исследование работы классификаторов для оптимизации постановки диагнозов бронхолегочных заболеваний

В данной статье рассмотрена возможность применения классификаторов, являющихся основой машинного обучения, для оптимизации постановки диагнозов бронхолегочных заболеваний. Рассмотрено работу нескольких классификаторов, и в результате исследования для поставленной задачи был выбран классификатор на основе метода ближайших соседей. В качестве параметров данного метода были выбраны рассчитанные на основе полиспектрального анализа численные характеристики сигналов звуков дыхания. Было установлено, что данный классификатор является простым для реализации и для работы с базой звуков дыхания. Полученная точность работы классификатора является достаточно высокой. Разработанный алгоритм призван существенно упростить работу врача-пульмонолога для постановки своевременного диагноза. Библи. 8, рис. 3.

Ключевые слова: классификатор; метод k-ближайших соседей; звуки дыхания; полиспектральный анализ.

UDC 004.9

A. Poreva, D. Goncharova

National Technical University of Ukraine "Kyiv Polytechnic Institute, Polytehnichna Str, 16, campus 12, off. 413a, Kyiv, 03056, Ukraine.

Research of classifiers' work to optimize diagnoses bronchopulmonary diseases

This article considers the possibility of using classifiers, which are the basis of machine learning to diagnoses bronchopulmonary diseases optimization. The work of a few classifiers was considered and as a result of the research the nearest neighbor method was chosen for the task classifier. As the parameters of this method numerical characteristics of breathing sounds signals were chosen. This characteristics was calculated on the basis of polyspectral analysis. It was found that this classifier is simple to implement

and to operate with the data base of breathing sounds. The resulting accuracy of the classifier is high enough. The algorithm is designed to greatly simplify the work of the doctor-pulmonologist for setting a timely diagnosis. Referense 8, Figures 3.

Keywords: classifier; method k-nearest neighbors; breath sounds; polyspectral analysis.

Reference

1. Zolotih, N. U. Machine Learning and Data Mining // URL: <http://www.uic.unn.ru/~zny/ml/>
2. Vorontsov, K. V. Mathematical methods of training on precedents (machine learning theory) // <http://www.ccas.ru/voron> (Rus)
3. Mitchell, T. (1997). Machine Learning. McGraw-Hill Science/Engineering/Math. ISBN 0-07-042807-7.
4. Data classification by support vector method- URL: <https://habrahabr.ru/post/105220/> (Rus)
5. Poreva, G. S, Karplyuk, Y. S., Makarenkova, A. A., Makarenkov, A. P. (2014). Detection of specific acoustic characteristics of patients with COPD based on spectral analysis of respiratory sounds // Electronics and Communications. Vol. 19, №6(83), Pp. 82-86 (Ukr)
6. Poreva, G. S, Makarenkova, A. A., Karplyuk, Y. S., Goncharenko, A. A. (2014). Application of polyspectral analysis to determine the diagnostic signs in the breathing sounds in patients with COPD - Proceedings of the National Technical University "KhPI" - technologies. Series: New solutions in modern technologies. Kh.: NTU "KhPI". No. 36(1079). P. 200, Pp. 49-55.
7. Poreva, A., Karplyuk, Y., Makarenkova, A., Makarenkov, A. (2015). Detection of COPD's Diagnostic Signs Based on Polyspectral Lung Sounds Analysis of Respiratory Phases 2015 IEEE 35th International Scientific Conference Electronics and Nanotechnology (ELNANO), pp.351-355.
8. k nearest neighbor method in Python, URL: <http://glebmikhaylov.com>.