

УДК 004:62-52:004.03

ВИКОРИСТАННЯ НЕПАРАМЕТРИЧНИХ МЕТОДІВ В ЗАДАЧАХ АВТОМАТИЗОВАНОГО МЕДИЧНОГО ДІАГНОСТУВАННЯ

Комлева Г. О.

к.т.н., доцент каф. СПЗ Комлева Н.О.

Одеський Національний Політехнічний Університет, УКРАЇНА

АНОТАЦІЯ. Виконано аналіз статистичних методів, що дозволяють проводити класифікацію об'єктів медичного діагностування на підставі значень параметрів референтних груп, у яких не висуваються апріорні припущення про характер розподілу досліджуваних даних та не передбачається будь-яких припущень про співвідношення параметрів розподілу аналізованих величин.

Вступ. Задача медичного діагностування може бути вирішена з використанням статистичного аналізу даних. Статистика в медицині є одним з інструментів аналізу експериментальних даних і клінічних спостережень, а також мовою, за допомогою якого повідомляються отримані математичні результати. Крім того, математичний апарат широко застосовується в діагностичних цілях, вирішенні класифікаційних завдань [1]. У рамках даної роботи механізми діагностування розглянуто на прикладі пульмонологічної системи DiaSpectrEx [2].

Мета роботи. Метою роботи є підвищення результативності роботи лікарів шляхом додавання нової функціональності у медичну діагностичну систему DiaSpectrEx, що дозволяє класифікувати апріорно невідомий об'єкт з використанням даних у малих референтних групах з заздалегідь невідомим законом розподілу, або з таким, що не є Гаусовим.

Основна частина роботи. Традиційно для діагностування стану об'єкта та його класифікації використовуються різні методи, кожен з яких має свої переваги та особливості застосування. Основними з них є класифікація за допомогою дерев рішень; байєсівська класифікація; класифікація за допомогою штучних нейронних мереж; класифікація методом опорних векторів; статистичні методи, зокрема, лінійна регресія; класифікація за допомогою методу найближчого сусіда; класифікація СBR-методом; класифікація за допомогою генетичних алгоритмів [3]. Розглянутий у даній роботі аналіз базується на дослідженні наборів медичних даних, одержуваних в результаті спеціалізованих обстежень пацієнтів, які входять в комплекс заходів для проведення неінвазивного пульмонологічного діагностування.

Задача, яку вирішує система DiaSpectrEx, є вузькоспеціалізованою, тому повних аналогів знайдено не було. Але так як діагностування відбуватиметься з використанням статистичних методів, частковими аналогами програмної системи DiaSpectrEx можна вважати інші статистичні програми: SPSS, STADIA, STATISTICA [4].

Якщо при рішення задачі діагностування є необхідність вибору одного з лише двох можливих діагнозів, використовується метод Манна-Уїтні, який використовує непараметричний статистичний критерій для оцінки різниці між двома вибірками за рівнем будь-якої ознаки, вимірюваної якісно, та дозволяє виявити відмінності в значенні параметра між малими вибірками. В іншому випадку використовується метод Краскела-Уолліса, призначений для перевірки рівності медіан декількох вибірок. Цей критерій є багатовимірним узагальненням критерію Уилкоксона-Манна-Уїтні. Критерій Краскела-Уолліса є ранговим, тому він інваріантний по відношенню до будь-якого монотонного перетворення шкали вимірювання. Таким чином були перевірені нульові гіпотези про те, що відповідні діагностичні ознаки не показують відмінності між класифікаційними групами. У тих випадках, коли відповідна нульова гіпотеза була відхилена, приймалася альтернативна гіпотеза про існування відмінностей між групами по досліджуваній діагностичній ознаці. Розглянемо ці методи більш детально. Для застосування U -критерію Манна-Уїтні потрібно виконати такі дії [5].

1. Скласти єдиний ряд з даних з обох вибірок, розставивши їх елементи за ступенем наростання ознаки і приписавши меншому значенню менший ранг. Загальна кількість рангів N

вийде рівною сумі рангів n_1 та n_2 вибірок, де n_1 – кількість елементів в першій вибірці, а n_2 – кількість елементів у другій вибірці.

2. Розділити єдиний ранжируваних ряд на два, що складаються відповідно з одиниць першої і другої вибірок. Підрахувати окремо суму рангів, що припали на долю елементів першої вибірки, і окремо – на частку елементів другої вибірки. Визначити більшу з двох рангових сум (T_x), відповідну вибірці з n_x елементами.

3. Визначити значення U -критерію Манна-Уїтні за формулою:

$$U = n_1 + n_2 + \frac{n_x(n_x + 1)}{2} - T_x \quad (1)$$

4. По таблиці для обраного рівня статистичної значущості визначити критичне значення критерію для даних n_1 і n_2 . Якщо отримане значення U менше табличного або дорівнює йому, то визнається наявність суттєвої різниці між рівнем ознаки в розглянутих вибірках (приймається альтернативна гіпотеза). Якщо ж отримане значення U більше табличного, приймається нульова гіпотеза. Достовірність відмінностей тим вище, чим менше значення U .

Критерій Краскела-Уолліса призначений для оцінки розходжень одночасно для трьох і більше вибірок та дозволяє встановити, що рівень ознаки змінюється при переході від групи до групи, але не вказує напрямок цих змін. Він є продовженням U -критерію на більшу кількість вибірок та у якості обмеження потребує, щоб число спостережень в кожній вибірці бути не менше 3. Для застосування Краскела-Уолліса потрібно виконати наступні дії [6].

1. Об'єднати усі вибірки в одну загальну вибірку, визначивши ознаку, за якою виконується групування.

2. Виконати ранжування цієї вибірки.

3. Підрахувати суму рангів для кожної групи.

4. Знайти $H_{екс}$ за формулою:

$$H_{екс} = \frac{12}{N(N+1)} * \sum \frac{T_j^2}{n} - 3(n+1) \quad (2)$$

де N – загальна кількість обстежуваних в об'єднаній групі, n – кількість досліджуваних в кожній групі, T – суми рангів по кожній групі.

5. Визначити критичні значення і відповідний їм рівень значущості по таблиці критичних значень критерію H Краскела-Уолліса. Якщо $H_{екс} > H_{кр}$, то гіпотеза H_0 відкидається.

Висновки. Таким чином, у систему DiaSpectrEx була додана нова функціональність, що надала можливість проведення діагностування з використанням непараметричних статистичних методів. При розробці системи була обрана архітектура системи «клієнт-сервер», що дозволило розподілити бізнес-логіку програми та зняти навантаження з серверу, та шаблони Model-View-Controller та Layers. Програмною мовою серверу була обрана Java з використанням фреймворку Spring, програмною мовою клієнта – JavaScript у зв'язці з AngularJS, який дозволяє реалізувати MVC-структуру на клієнті.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Комлева Н.О. Построение системы диагностических признаков с использованием метода дискриминантного анализа в офтальмологических исследованиях. – Радиоелектронні і комп'ютерні системи. – Харків «ХАІ», 2010. – Вип. 6 (47). – С. 250 – 253.
2. Chernenka K.S., Tymchenko V.I., Komleva N.O. Decision support System for Automated Medical Diagnostics // Electrotechnic and Computer Systems. – Kiev: Science and Technology, 2016. – No. 23(99). – P. 65 – 72.
3. Касюк С.Т. Первинний, кластерний, регресійний і дискримінантний аналіз даних спортивної медицини на комп'ютері: учеб.-метод. посібник / С.Т.Касюк. – Челябинськ: Уральська Академія, 2015. – 160 с.
4. Дубнов П.Ю. Обработка статистической информации с помощью SPSS – М.: ООО «Издательство АСТ»: Издательство «НТ Пресс», 2010. – 221 с.
5. Гублер Е. В., Генкин А. А. Применение непараметрических критериев статистики в медико-биологических исследованиях. – Л., 2003. – 187 с.
6. Реброва О.Ю. Статистический анализ медицинских данных / О.Ю. Реброва. – М.: Медиа Сфера, 2006. – 310 с.