

О.В. Гойко, С.І. Мохначов

ТЕХНОЛОГІЯ ОБРОБКИ Й АНАЛІЗУ МЕДИЧНИХ ДАНИХ З УРАХУВАННЯМ ЇХ ОСОБЛИВОСТЕЙ

Національна медична академія післядипломної освіти ім. П.Л. Шупика, м. Київ

Ключові слова: технологія аналізу медичних даних, статистичні методи обробки інформації.

Надано інформацію про особливості та специфіку обробки й аналізу медичних даних. Запропоновано інформаційну технологію, що враховуватиме специфічні вимоги до обробки й аналізу статистичних даних медичних досліджень.

О.В. Гойко, С.І. Мохначов

Технология обработки и анализа медицинских данных с учетом их особенностей

Ключевые слова: технология анализа медицинских данных, статистические методы обработки информации.

Приведена информация относительно особенностей и специфики обработки и анализа медицинских данных. Предлагается информационная технология, которая будет учитывать специфические требования относительно обработки и анализа статистических данных медицинских исследований.

O.V. Goiko, S.I. Mokhnachov

Technology of processing and analysis of medical data in view of their features

Key words: technology of medical data analysis, statistical methods of data processing.

The information is given regarding the features and the specific processing and analysis of medical data. Information technology that will be responsive to the specific requirements for the processing and analysis of statistical data concerning medical researches is offered.

Проблема обробки й аналізу інформації, отриманої в процесі медичної діяльності лікувального закладу, нині є однією з найактуальніших. Невпинне зростання інформаційних масивів і кількісних даних про стан здоров'я пацієнта змушує лікарів замислюватись над тим, як виконати обробку й аналіз медичної інформації з метою правильної діагностики, вибору адекватного лікування, прогнозування перебігу лікування тощо. Аналіз такої інформації без використання сучасних комп'ютерних технологій став просто неможливим, оскільки стримуватиме не лише подальший розвиток медичної науки, але й практичне надання медичної допомоги.

На сьогодні створення і ведення сучасних комп'ютерних баз даних про стан здоров'я пацієнтів, перебіг їх захворювань, лабораторних аналізів і лікування вже не становить складності для фахівців в інформаційній сфері. Складнішою проблемою для них є відсутність ефективної інформаційної технології обробки й аналізу медичних даних, що надала б змогу медику-аналітику знайти не виявлені на перший погляд закономірності та взаємозв'язки між різними показниками в медичних даних, що в кінцевому підсумку дозволить підвищити результативність лікування за рахунок вибору інтенсивності терапії, адекватної стану організму хворого, оціненого за сукупністю виявлених факторів ризику.

МЕТА РОБОТИ

Визначення особливостей обробки й аналізу статистичних даних медичних досліджень, а також удосконалення інформаційної технології, що їх враховуватиме.

Застосування в медицині традиційних статистичних

методів аналізу даних має досить давню історію [2–4]. Визнаючи їх велику користь і важливість, слід відзначити обмеженість їх можливостей стосовно саме медичних даних [5]. Специфіка медичних даних, що застосовується в аналізі причинно-наслідкових зв'язків, не дозволяє коректно використовувати ту велику кількість різних математичних методів, що нині існують і використовуються для обробки й аналізу інформації. Основною причиною цього є те, що класичні статистичні методи добре працюють при перевірці заздалегідь сформульованих гіпотез, а це в медичній практиці вдається зробити досить рідко. Крім того, слід зазначити різні обмеження статистичних методів, що пов'язані з конкретними законами розподілу вибірок, типами даних, обсягом вибірки тощо.

Розроблені наприкінці 1990-х років методи інтелектуального аналізу даних (нейронні мережі різних видів, нечітка логіка, генетичні алгоритми тощо) можуть компенсувати гостроту проблеми обмеженості можливостей статистичних методів. Прикладами універсальних програмних систем, що базуються на цих методах, є Intelligent Miner (США), Knowledge Studio (Канада), SAS (США), MineSet (США), Clementine (Англія), ST: Neural Networks, PolyAnalyst (Росія) тощо. Проте універсальність перетворюється на недолік у випадку медичного застосування, оскільки не передбачає врахування всіх специфічних особливостей медичних даних, пов'язаних з надзвичайно великими обсягами масивів даних і вибором серед них найбільш інформативних ознак, побудові вирішальних правил



з урахуванням ризиків, оцінки валідності інформації тощо.

Інформаційна технологія аналізу медичних даних, що допоможе медикам при вирішенні цих завдань і враховуватиме специфічні вимоги до особливостей обробки й аналізу статистичних даних медичних досліджень, має включати наступні етапи:

1. З метою підвищення ефективності обробки й аналізу медичної інформації доцільно скоротити кількість показників за рахунок виключення сильно корельованих факторів, що слабо впливають на кінцевий результат. Так зване «первинне просіювання» незалежних змінних здійснюється за допомогою методів факторного аналізу та методу головних компонент у класичній і нейромережевій постановці, а також ряду інших методів, що дозволили б редукувати вихідну множину факторів, виключити сильно корельовані фактори і виділити підмножину найперспективніших для здійснення подальшого багатовимірного аналізу.

2. З метою побудови прогностичних моделей необхідно виявити найінформативніші комбінації умовно незалежних факторів на основі використання комплексу покрокових статистичних методів відбору. Для цього доцільно використати декілька покрокових алгоритмів відбору істотних змінних. У якості одного з таких алгоритмів пропонується використовувати алгоритм покрокового включення [1], побудований на основі розрахунку коефіцієнта множинної кореляції між залежною змінною та набором чинників, відібраних на попередньому кроці. Оцінка ефективності кожної з комбінацій розраховується як значення коефіцієнта детермінації: чим він більший, тим краще комбінація істотних змінних пояснює залежну змінну. Задавши межове значення, можна розглядати і досліджувати тільки ті комбінації, в яких значення коефіцієнта детермінації вище межового.

Для відбору комбінацій значущих прогностичних факторів можна використовувати покроковий алгоритм дискримінантного аналізу [1] на основі відстані Махаланобіса, що дає можливість на кожному етапі виявляти такі комбінації факторів, що є найкращими, найзначущими для класифікації та прогнозу.

3. Верифікація відібраних комбінацій найбільш значущих прогностичних факторів повинна здійснюватись шляхом їх використання для класифікації деякої тестової множини даних, не використаних на попередньому етапі.

4. Після відбору однієї або декількох комбінацій найзначущих прогностичних змінних і їх верифікації можна вирішувати завдання прогнозування перебігу захворювання і відповіді організму конкретного пацієнта на терапію. Як і на попередніх етапах інформаційної технології аналізу медичних даних, пропонується використовувати поєднання декількох статистичних і нейромережевих методів, що дасть змогу обрати оптимальний.

ВИСНОВКИ

Інформаційна технологія аналізу медичних даних, що враховуватиме особливості медичних даних і специфічні вимоги до їх обробки й аналізу, дасть змогу медичним працівникам виконувати якісніший аналіз причинно-наслідкових зв'язків, виявляти приховані закономірності та взаємозв'язки між різними факторами, прогнозувати перебіг лікувального процесу, визначати ступінь ризику тощо. Ефективність всебічного аналізу великої кількості різнорідних медико-біологічних, клінічних та епідеміологічних факторів можна забезпечити тільки шляхом комплексного використання методів багатовимірного статистичного й інтелектуального нейромережевого аналізу.

СПИСОК ЛІТЕРАТУРИ

1. *Мінцер О.П.* Оброблення клінічних і експериментальних даних у медицині: навч. посібник / *О.П. Мінцер, Ю.В. Вороненко, В.В. Власов.* – К.: Вища школа, 2003. – 350 с.
2. *Малета Ю.С.* Математические методы статистического анализа в биологии и медицине / *Ю.С. Малета, В.В. Тарасов.* – М., 1982. – 287 с.
3. *Енюков И.С.* Методы, алгоритмы и программы многомерного статистического анализа: Пакет ППСА / *И.С. Енюков.* – М.: Финансы и статистика, 1986. – 232 с.
4. *Стентон Г.* Медико-биологическая статистика / *Г.Стентон.* – М.: Практика, 1999. – 464 с.
5. *Горбань А.Н.* Нейроинформатика / *А.Н.Горбань, В.Л.Дунин-Барковский, А.Н.Кирдин и др.* – Новосибирск: Наука. Сибирское предприятие РАН, 1998. – 462 с.

Відомості про авторів:

Гойко О.В., к. біол. н., доцент каф. медичної інформатики НМАПО ім. П.Л. Шупика.

Мохначов С.І., асистент каф. медичної інформатики НМАПО ім. П.Л. Шупика.

Поступила в редакцію 17.10.2012 г.