

УДК: 519.233.3

© Пожидаев В.Ф., Пожидаев И.В., Топорков Д.В., 2012

ПОСЛЕДОВАТЕЛЬНЫЙ АНАЛИЗ В ВЫБОРЕ НЕОДНОЗНАЧНЫХ РЕШЕНИЙ

Пожидаев В.Ф., Пожидаев И.В., Топорков Д.В.

Восточнoукраинский национальный университет им. В. Даля; Харьковский национальный университет имени В.Н. Каразина; Национальный авиационный университет

Введение. Медицинскую диагностику организма можно проводить с помощью математических методов и вычислительной техники. Среди математических методов наиболее постепенно развивается байесовский подход. Поскольку байесовский метод принятия решений имеет строгое математическое обоснование, на его основе возможна реализация механизма вывода, позволяющего решать задачи диагностики, которые характеризуются необходимостью учета неопределенности. Байесовский метод позволяет учесть эту неопределенность. Байесовский подход изложен в книгах Н. Бейли [1], А. Вальда [3], Е.В. Гублера [4, 5], Е.В. Гублера и А.А. Генкина [6], Л. Ластеда [7], С.А. Айвазяна, В.М. Бухштабера, И.С. Енюкова и Л.Д. Мешалкина [8] и др. Применению байесовского подхода в диагностике посвящены многочисленные статьи [9, 10, 11 – 16].

Цель исследования – иллюстрация байесовского подхода к решению задачи дифференциальной диагностики.

Материалы и методы. Допустим, что рассматривается ограниченная группа d различных заболеваний (D_1, D_2, \dots, D_d) и что каждый больной страдает только одним из них (т. е. из этого упрощенного описания исключается более сложный случай сочетания нескольких заболеваний). Допустим также, что имеется список s признаков, симптомов или результатов лабораторных анализов (на основе которых обычно получают необходимую информацию), обозначаемых S_1, S_2, \dots, S_s . Часто бывает удобно рассматривать всю совокупность признаков, симптомов и т. д.; для этой цели используется символ S без индекса. Пока будем считать, что все симптомы дискретны, т. е. каждый из них может либо присутствовать, либо отсутствовать.

Мы имеем:

1) для любого данного заболевания D_j вероятность $P(S|D_j)$ наблюдения определенного симптомокомплекса S .

2) вероятность $P(D_j)$ того, что больной, выбранный случайным образом, страдает заболеванием D_j (статистические данные).

Далее последователь сообщают информацию о симптомах, наблюдаемых у данного больного. Эти данные характеризуют определенный симптомокомплекс S . Для предварительного диагноза необходимо знать вероятность каждого заболевания при данном

симптомокомплексе, т. е. $P(D_j | S)$. Если заболевание D_j диагностируется с большой уверенностью, то вероятность $P(D_j | S)$ будет относительно велика, а все остальные вероятности будут значительно меньше (возможно, в 10 или 100 раз).

Если же два заболевания D_1 и D_2 кажутся почти одинаково возможными, а все остальные исключаются, то каждая из вероятностей $P(D_1|S)$ и $P(D_2|S)$ будет равна примерно 0,5, а все остальные вероятности будут близки к нулю и т. д.}. Известно, что апостериорная вероятность $P(D_j|S)$ пропорциональна произведению априорной вероятности $P(D_j)$ на функцию правдоподобия $P(S|D_j)$, т. е.

$$P(D_j|S) \sim P(D_j) P(S|D_j) \quad (1)$$

Так как сумма всех апостериорных вероятностей для различных заболеваний должна быть равна единице, более точно выражение (1) можно записать в виде:

$$P(D_j | S) = \frac{P(D_j)P(S | D_j)}{\sum_{j=1}^d P(D_j)P(S | D_j)}, j=1..d. \quad (2)$$

По существу это формула Бейеса, позволяющая вычислить вероятность справедливости некоторой гипотезы на основании ее априорной вероятности и некоторых эмпирических данных. Чаще всего принимается допущение о том, что отдельные симптомы имеют независимые распределения.

Если считать, что s симптомов статистически независимы, то можно записать соотношение

$$P(S | D_j) = \prod_{i=1}^s P(S_i | D_j), j=1..d \quad (3)$$

и на основе имеющегося материала с достаточно высокой точностью вычислить вероятности $P(S_j|D_j)$.

Ясно, что вычисление данных для каждого нового больного с помощью уравнений (2) и (3) требует большого объема хотя и элементарных, но утомительных и поглощающих много времени арифметических вычислений. Поэтому для этой цели крайне желательно и даже необходимо применять электронную вычислительную технику. В связи с этим возникает важный вопрос использования компьютеров для решения этих

задач. Поскольку число поддающихся наблюдению симптомов, и особенно объем лабораторных тестов, непрерывно увеличивается, вскоре может наступить такой момент, когда даже опытный врач не сможет воспользоваться всей той информацией, которую он будет иметь в своем распоряжении. Таким образом, мы сталкиваемся с парадоксальной ситуацией — потенциально имеется возможность получить наиболее точные результаты, но на самом деле использовать их затруднительно из-за наличия очень большого количества информации. Применение вычислительной техники для хранения и обра-

ботки исходных данных, а также для вычисления вероятности верного решения приобретает в этой ситуации решающее значение.

Выводы: При наличии постоянно пополняемой динамической базы данных байесовский подход к анализу принадлежности параметров системы к тому или иному кластеру дает все более надежные результаты. Этот факт можно успешно использовать при диагностике состояния в том случае, когда точные методы измерения недоступны, т. е. в случае неразрешимого контроля любой системы.

ЛИТЕРАТУРА:

1. **Бейли Н.** Математика в биологии и медицине / Н. Бейли. — М.: Мир, 1970. — Гл. 11.
2. **Вальд А.** Статистические решающие функции / А. Вальд // Позиционные игры / под ред. Н.Н. Воробьева и Н.Н. Врублевской. — М.: Наука, 1967. — С. 300-522.
3. **Гублер Е.В.** Информатика в медицине / Е.В. Гублер. — М.: Финансы и статистика, 1991.
4. **Ластед Л.** Введение в проблему принятия решений в медицине / Л. Ластед. — М.: Мир, 1971.
5. Honours Year Project Report. Bayesian Learning with Incomplete Information in Large Medical Databases: An Application of the EM Algorithm / By Quek Hui Nar, School of Computing National University of Singapore, 1998/99.
6. **Gregory S. Cunningham.** Bayesian estimation of regularization parameters for deformable surface models / Gregory S. Cunningham, Andre Lehovich, Kenneth M. Hanson. — Los Alamos National Laboratory, University of Arizona, Dept. of Applied Mathematics, 1999.
7. Адаптация метода Байеса к медицинской диагностике / **А.Н. Повалихин, А.В. Стребуков, Г.Г. Устинов, А.В. Гайнер, Л.М. Жмудяк** // Международная конференция молодых ученых по математическому моделированию и информационным технологиям: программа и тезисы докладов, 17 ноября 2002 г. — Новосибирск, 2002. — С. 77.
8. Использование двумерных распределений для диагностики по методу Байеса / **А.Н. Повалихин, А.В. Стребуков, М.Л. Жмудяк, А.Л. Жмудяк** // Материалы пятой городской научно-практической конференции молодых ученых, 20-21 ноября 2003 г. — Барнаул, 2003. — С. 339-340.

Пожидаев В.Ф., Пожидаев И.В., Топорков Д.В. Последовательный анализ в выборе неоднозначных решений // Украинський медичний альманах. — 2012. — Том 15, № 2. — С. 129-130.

Рассматривается последовательный байесовский анализ к пошаговому методу формирования возможного диагноза исследуемого состояния биологической системы.

Ключевые слова: диагноз, симптом, вероятность заболевания.

Пожидаєв В.Ф., Пожидаєв І.В., Топорков Д.В. Послідовний аналіз у виборі неоднозначних рішень // Український медичний альманах. — 2012. — Том 15, № 2. — С. 129-130.

Розглянуто послідовний байесівський аналіз до покрокового методу формування можливого діагнозу досліджуваного стану біологічної системи.

Ключові слова: діагноз, симптом, ймовірність захворювання.

Pojidaev V.F., Pojidaev I.V., Toporkov D.V. A successive analysis in the choice of ambiguous solution // Український медичний альманах. — 2012. — Том 15, № 2. — С. 129-130.

The successive is Bayes, an analysis to the incremental method of forming possible diagnosis of the probed biological system is considered.

Key words: diagnosis, symptom, probability of disease.

Надійшла 14.01.2012 р.
Рецензент: проф. Ю.М.Вовк