УДК 616-07:51-053.3

ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ЗАДАЧЕ ДИАГНОСТИКИ СОСТОЯНИЯ ЗДОРОВЬЯ ДЕТЕЙ ПЕРВОГО ГОДА ЖИЗНИ

Черкашина Ю.А.

Национальный исследовательский Томский политехнический университет, Томск, e-mail: cherr999y@mail.ru

Статья является результатом научных исследований сотрудников кафедры Прикладной математики Национального исследовательского Томского политехнического университета и сотрудников Сибирского государственного медицинского университета. Исследования посвящены применению математических методов для проведения процедуры распознавания состояния здоровья детей на основе медицинских данных гормонов крови. В статье подробно рассмотрена процедура неоднородной последовательной процедуры распознавания (НППР). Получены диагностические коэффициенты для отнесения объектов исследования к одному из непересекающихся классов (здоров, болен). Для применения НППР в задачах диагностики состояния организма сформировано две группы детей: обучающая и контрольная. Решение о наличии или отсутствии заболевания у детей контрольной группы определяется при использовании диагностического коэффициента. Приведены результаты исследования. Рассмотрены два примера использования диагностических коэффициентов отнесения объекта одному из двух классов.

Ключевые слова: неоднородная последовательная процедура распознавания, диагностические коэффициенты, диагностика

APPLICATION OF MATHEMATICAL METHODS TO THE PROBLEM DIAGNOSIS OF HEALTH OF INFANTS

Cherkashina Y.A.

National Research Tomsk Polytechnic University, Tomsk, e-mail: cherr999y@mail.ru

The article includes results of scientific results achieved at department of Applied Mathematics at National Research Tomsk Polytechnic University and Siberian State Medical University. Researches are devoted the application of mathematical methods for the procedure of recognition of children's health based on medical data of hormone levels in the blood. The inhomogeneous sequential pattern recognition procedures (ISPRP), used to diagnose the health of children in the first year of life has been considered in the article. Diagnostic coefficients for assigning objects to one of disjoint classes (healthy or sick) were obtained. For diagnostics, using ISPRP was formed two groups: training and control. The decision about the presence or absence of disease in the control group is determined using diagnostic coefficient. Results of the study are given and two examples were considered.

Keywords: heterogeneous sequential pattern recognition procedures, diagnostic coefficients, diagnosis

В последнее время в медицине все большее число исследователей занимаются проблемой диагностирования состояния здоровья детей. Известно, что предрасположенность человека к различным заболеваниям закладывается, в основном, в первый год жизни, поэтому актуальными являются задачи оценки состояния здоровья детей именно в этот период.

Проблемой выявления предпатологических и патологических изменений у грудных детей занимаются врачи, но в дополнение к специальным медицинским методам и подходам можно применять также и математические методы, алгоритмы и модели для оценки состояния здоровья детей [1].

Для постановки диагноза врачу приходится оперировать большим массивом разнородных данных и сложным комплексом методик. В связи с этим актуальной является задача по разработке дополнительных средств обработки показателей, характеризующих состояние детей, подозреваемых в заболевании и выдачи рекомендаций врачу.

Постановка диагноза на протяжении многих лет являлась в определенной мере искусством, помноженным на опыт и интуицию врача, и только с применением математических методов постановка диагноза может быть сформулирована, как математическая задача и автоматизирована.

Успешное решение задачи оценки состояния организма ребенка зависит от квалификации врача, однако современные компьютерные системы значительно ускоряют процесс обработки исходных данных и помогают поставить правильный диагноз.

Поэтому целью работы является обработка медицинских данных детей для оценки состояния их здоровья при помощи математических методов.

Математическая постановка задачи

При решении задач математической статистики существенную роль играет предположение о виде закона распределения наблюдаемой случайной величины X. Методы математической статистики, основанные на этом предположении, называют параметрическими.

Однако у параметрических методов имеются существенные недостатки. Во-

первых, на практике вид распределения наблюдаемой величины очень часто неизвестен. Во-вторых, экспериментальные данные при сборе и обработке информации почти всегда искажаются, что меняет их вид распределения. Поэтому, применяя параметрические методы в условиях такой априорной стохастической неопределенности, необходимо ясно осознавать, что расхождение между параметрической моделью и реальной ситуацией может привести к сильно искаженным или даже неверным результатам.

Следовательно, возникает необходимость в разработке таких статистических процедур, которые, с одной стороны, в ситуации, наиболее благоприятной для параметрических методов, почти не уступали бы им в эффективности, а с другой стороны, были бы малочувствительными к нарушению предположений, лежащих в основе параметрической модели.

Такие методы существуют. Они получили название непараметрических методов, так как не требуют знания закона распределения наблюдаемой случайной величины и используют лишь минимальную априорную информацию. Одним из важных плюсов непараметрических методов является возможность рассмотрение качественных признаков, которые выражаются порядковыми номерами или индексами.

Для отнесения всей совокупности объектов к одному из непересекающихся классов (болен, здоров) можно воспользоваться одним из непараметрических методов, а именно, неоднородной последовательной процедурой распознавания (НППР).

Являясь одномерной, НППР не требует использования сложных методов многомерной статистики. Она также не требует знания законов, которым подчинены эмпирические распределения, и пригодна при любой форме распределений [4]. Еще одним плюсом НППР является то, что данная процедура позволяет использовать неоднородные признаки, а именно, качественные (боль при осмотре, характер боли и т.д.) и количественные (показатели крови).

Для решения задач диагностики с помощью неоднородной последовательной процедуры распознавания следует сформировать две группы детей: обучающую и контрольную [3].

Введем диапазоны изменения показателей. Первый диапазон содержит в себе те значения, которые попадают в интервал, соответствующий медицинским показателям в пределах нормы, второй диапазон, значения, которые не попадают в данный интервал. Для каждого показателя существует своя граница нормы.

Для обучающей группы находятся диагностические коэффициенты по следующей формуле [2]:

$$\text{ДК}(\sigma_{k,j}) = \lg\frac{P\{x_j \in \sigma_{k,j} \mid A1\}}{P\{x_j \in \sigma_{k,j} \mid A2\}}, \qquad (1)$$

где $\sigma_{k,j}$ – k-ый диапазон j-ого показателя, $k = 1..3, j = 1...M$;
x_j – значение j-ого показателя;
$A1$ – событие наличия заболевания;
$A2$ – событие отсутствия заболевания;

$P\{x_j \in \sigma_{k,j} \mid A1\}$ – условная вероятность (вероятность события $x_j \in \sigma_{k,j}$, при условии наступления события $A1$);

$P\{x_j \in \sigma_{k,j} \mid A2\}$ – условная вероятность (вероятность события $x_j \in \sigma_{k,j}$, при условии наступления события $A2$).

Вероятности, стоящие в числителе и знаменателе формулы (1) определяются следующими выражениями:

$$P\left\{x_j \in \sigma_{k,j} \middle| A1\right\} = \frac{P\left\{x_j \in \sigma_{k,j} \cap A1\right.}{P(A1)}. \qquad (2)$$

$$P\left\{x_j \in \sigma_{k,j} \middle| A2\right\} = \frac{P\left\{x_j \in \sigma_{k,j} \cap A2\right.}{P(A2)}. \qquad (3)$$

Решение о наличии или отсутствии заболевания у детей контрольной группы определяется при выполнении следующего неравенства:

$$\text{ДК}_{\text{пор}}(A_2) < \text{ДК}(\sigma_{k,1}) + \text{ДК}(\sigma_{k,2}) + \ldots + \\ + \text{ДК}(\sigma_{k,N}) < \text{ДК}_{\text{пор}}(A_1), \qquad (4)$$

где $\text{ДК}(\sigma_{k,j})$ – значение диагностического коэффициента k-ого диапазона, j-ого показателя.

Существование нескольких состояний (болен, здоров) предполагает наличие диагностических порогов, при достижении которых выносится решение. В неравенстве (4) пороговое значение для состояния $A1$ определяется как:

$$\text{ДК}_{\text{пор}}(A1) = \lg\frac{1-\alpha}{\beta}, \qquad (5)$$

для состояния $A2$:

$$\text{ДК}_{\text{пор}}(A2) = \lg\frac{\alpha}{1-\beta}, \qquad (6)$$

где α – ошибка первого рода (объект из класса $A1$ можно отнести к классу $A2$);
β – ошибка второго рода (объект из класса $A2$ относят к классу $A1$).

Если сумма всех диагностических коэффициентов превысит пороговое значение ДК($A1$) – у ребенка присутствует заболевание, если сумма окажется меньше порогового значения ДК($A2$), ребенок считается здоровым. Если же сумма диагностических коэффициентов оказалась между пороговыми значениями, это означает, что данного

ребенка следует обследовать более тщательно и полученной информации недостаточно для вынесения диагноза.

Таблица 1

Диагностические коэффициенты

Показатель	ДК	
	Вне нормы	В норме
т41	0,228315	– 0,25297
kor1	0,360424	– 0,1549
kor2	0,317672	– 0,14435
kor3	0,291343	– 0,13717
kor4	0,231757	– 0,11869
kor5	0,232314	– 0,1288
kor6	0,232314	– 0,1288
kor7	0,232314	– 0,1288
kor8	0,221849	– 0,12494
kor9	0,243038	– 0,12245
kor10	0,304308	– 0,14078
kor11	0,352183	– 0,14133
kor12	0,437613	– 0,15823
ins5	0,112109	– 0,13397
ins6	0,113404	– 0,0975
ins7	0,384367	– 0,14817
ins8	0,60206	– 0,1549
ins9	0,735954	– 0,14018
ins10	0,754921	– 0,12892
ins11	1,021189	– 0,09084
ins12	1,021189	– 0,09084

Результаты исследования и их обсуждение

Исходная выборка включала 198 детей. Для определения здоров или болен ребенок сформировано две группы: обучающая – 132 ребенка и контрольная – 66 детей.

Диагностические коэффициенты, рассчитанные с использованием неоднородной последовательной процедуры распознавания, представлены в табл. 1. Для того чтобы установить диагноз ребенка из контрольной группы, необходимо в таблице найти значения диагностических коэффициентов, сложить коэффициенты и в зависимости от того, с какой стороны неравенства (4) оказалось значение, вынести решение, больна или здорова женщина.

Пороговые значения состояний «болен», «здоров» представлены в табл. 2.

Таблица 2

Пороговые значения

ДК(А1)	0,97
ДК(А2)	– 1,25

Рассмотрим пример использования диагностических коэффициентов для определения диагноза.

Исходные данные двух детей, один из которых здоров, второй имеет заболевание, представлены в табл. 3.

Таблица 3

Пример распознавания заболевания с использованием НППР

Показатель	Норма	Пример 1 – болен		Пример2 – здоров	
		Значение показателя	ДК	Значение показателя	ДК
т41	[40 120]	125,05	0,228315	88,75	– 0,25297
kor1	[200 400]	494,07	0,360424	273,77	– 0,1549
kor2		483,81	0,317672	268,62	– 0,14435
kor3		475,33	0,291343	264,70	– 0,13717
kor4		468,63	0,231757	262,03	– 0,11869
kor5		463,71	0,232314	260,60	– 0,1288
kor6		460,56	0,232314	260,41	– 0,1288
kor7		459,19	0,232314	261,46	– 0,1288
kor8		459,60	0,221849	263,75	– 0,12494
kor9		461,78	0,243038	267,29	– 0,12245
kor10		465,74	0,304308	272,06	– 0,14078
kor11		471,48	0,352183	278,08	– 0,14133
kor12		479,00	0,437613	285,34	– 0,15823
ins5	[0 11]	8,78	– 0,13397	9,67	– 0,13397
ins6		9,15	– 0,0975	9,78	– 0,0975
ins7		9,54	– 0,14817	9,90	– 0,14817
ins8		9,96	– 0,1549	10,04	– 0,1549
ins9		10,40	– 0,14018	10,20	– 0,14018
ins10		10,86	– 0,12892	10,37	– 0,12892
ins11		11,34	1,021189	10,56	– 0,09084
ins12		11,85	1,021189	10,76	– 0,09084
ДК			4,924182		– 2,86753

В примере 1 ребенок из тестовой выборки имел диагноз, поставленный врачом, – болен. Применяя НППР, сумма всех диагностических коэффициентов равна 4,92, это значение больше верхнего порогового значения ДК($A1$) = 0,97. Следовательно, можно сделать вывод что ребенок в примере 1 болен. Состояние организма определено верно.

В примере 2 ребенок из тестовой выборки имел диагноз – здоров. Применяя НППР, сумма всех диагностических коэффициентов равна – 2,86, это значение меньше порогового, равного – 1,25, следовательно, можно сделать вывод: ребенок в примере 2 здоров.

Для оценки качества распознавания были проверены все дети из тестовой выборки. Качество распознавания с помощью неоднородной последовательной процедуры распознавания составило 81 %.

Заключение

Неоднородная последовательная процедура распознавания имеет большое значение и практическое применение в медицине. Процедура была апробирована на реальных медицинских данных, предоставленных медицинскими работниками. Рассмотрены два примера, демонстрирующие порядок использования НППР. Качество диагностирования, 81 % детей тестовой выборки был правильно отнесен к соответствующему классу, можно считать приемлемым.

Список литературы

1. Гергет О.М., Кочегуров В.А. Решение актуальных медицинских задач математическими методами. – Томск: Изд-во ТПУ, 2002. – 145 с.

2. Гублер Е.В. Вычислительные методы анализа и распознавания патологических процессов. – Л: Медицина, 1978. – С. 296.

3. Дымова А.В., Лукина Е.Ю., Гергет О.М. «Диагностика мастопатии с помощью математических методов» V Международная студенческая электронная научная конференция «Студенческий научный форум». 15 февраля – 31 марта 2013 года – [Электронный ресурс]. Режим доступа – http://www.scienceforum.ru/2013/28/2326, свободный.

4. Марухина О.В. Алгоритмы обработки информации в задачах оценивания качества обучения студентов вуза на основе экспертно-статистических методов: Дис. ... канд. техн. наук. – Томск, 2003. – 164 с.