

ПРОГРАММА ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ

А.М. Жмудяк, д.т.н.; А.А. Жмудяк, к.т.н.

(Алтайский государственный технический университет им. И.И. Ползунова, г. Барнаул,
L_jmoudiak@hotmail.com)

Диагностика базируется на расчете вероятностей болезней. Рассмотрено использование многомерных распределений, многократно определенных диагностических признаков, маркеров вирусных гепатитов и др. Созданная программа верно проводит дифференциальную диагностику механической и паренхиматозной желтух в 97 % случаев.

Ключевые слова: программа диагностики, диагностика, вероятность, формула Байеса, диагностические признаки, паренхиматозная желтуха, механическая желтуха.

Разработанные методика и программа диагностики [1] были развиты и использованы для дифференциальной диагностики паренхиматозной и механической желтух. Тестирование на 564 историях болезней показало, что компьютерный диагноз верен в 97 % случаев.

Паренхиматозная желтуха обусловлена вирусными (вирусные гепатиты А, В, С, D, G, TT и др.) и иными заболеваниями печеночной клетки, которые лечатся терапевтически. Механическая желтуха является следствием перекрытия (обтурации) желчных путей, например, камнями или опухолью, и зачастую требует срочного хирургического вмешательства. Эти совершенно разные по природе и тактике лечения заболевания имеют ряд схожих проявлений и затрудненную дифференциальную диагностику: процент диагностических ошибок колеблется от 8 до 30, что обуславливает практическую востребованность программы.

Методика компьютерной диагностики базируется на расчете вероятностей болезней по формуле Байеса. Для диагностики использовались 36 дискретных и непрерывно распределенных *диагностических признаков* (ДП): регистрационно-биографические данные (пол, возраст и т.п.), данные анамнеза, симптомы заболевания, клинические анализы, данные инструментальных обследований пациента и др.

На основании собранной БД были сформированы распределения ДП: отдельно по паренхиматозной желтухе и отдельно по механической. По распределениям определены выборочные плотности относительных частот (далее – плотности вероятностей) непрерывных ДП и выборочные относительные частоты (далее – вероятности) дискретных ДП. Эти плотности и вероятности подставляются в формулу Байеса. По вероятностям и плотностям вероятностей используемых ДП с помощью формулы Байеса вычисляются вероятности паренхиматозной желтухи и механической. При некоторых сочетаниях маркеров дополнительно вычисляются вероятности и диагностируются формы паренхиматозной желтухи, например, вероятность острого вирусного гепатита В.

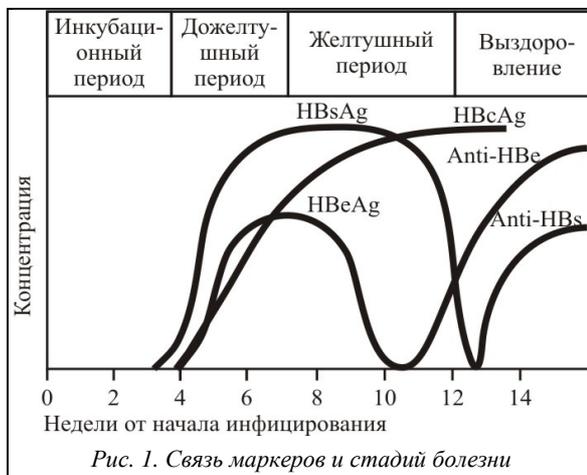
Всего рассматривается девятнадцать маркеров вирусных гепатитов. Необходимость анализа комбинаций маркеров обосновывается хотя бы сле-

дующим. Маркер HBsAg определяется не только в острой и хронической стадиях гепатита В, но и у практически здоровых людей – носителей этого антигена. По оценке вирусологов, в мире около 400 миллионов людей являются носителями HBsAg. (Информация о маркерах взята из работ [2–5]). По собранной и взятой из литературы статистике определены вероятности комбинаций положительных и отрицательных значений маркеров как при механической и паренхиматозной желтухах, так и при гепатитах А, В и С.

Следование медицинским таблицам диагностических значений маркеров затрудняет описание всех возможных их сочетаний, поэтому разработана методика описания динамики маркеров. Приведем элементы этой методики для гепатита В. Связь маркеров и стадий болезни отражена в методике с ориентировкой на приведенную на рисунке 1 схему.

Сочетание положительных и отрицательных маркеров, определенных у больного, в программе сравнивается с динамикой маркеров. В зависимости от ситуации это является основанием для диагноза «острый вирусный гепатит В» или для расчета вероятности этого диагноза, а также для расчета вероятностей механической и паренхиматозной желтух.

В известных работах при байесовской диагностике ДП считаются взаимонезависимыми, хотя понятно, что в едином организме ДП в общем случае зависят друг от друга.



За время болезни и лечения ДП изменяются и определяются не один раз, хотя обычно при диагностике математическими методами используются ДП, определенные только один раз, в первый день болезни. ДП, определенные в другие дни, в расчетах не используются. То есть не учитывается динамика болезни – важнейший при постановке диагноза фактор.

Однако авторам удалось учесть измерения одного и того же ДП в другие дни: повторные измерения одного и того же ДП были математически рассмотрены как новые отдельные ДП. То есть величина ДП, например, билирубина в 1-й день – это один ДП, билирубин во 2-й день – это математически другой ДП, билирубин в 5-й день – опять новый признак и т.д. Затем была найдена и учтена взаимозависимость между ДП, определенными в разные дни. Таким образом удалось учесть данные многократно определенных ДП, то есть динамику болезни.

При создании описываемой методики была поставлена задача включения медицинских знаний в чисто математический подход. В качестве одного из решений этой задачи предложено использовать медицинские знания для уточнения и дополнения используемой статистики (собранной и используемой БД). Имеющиеся распределения ДП всегда получены (построены) по ограниченному выборкам. Медицинские знания позволяют эти распределения уточнить и дополнить. Таким уточнением и дополнением статистики в математический подход являются знания врача о физиологии и патофизиологии, течения болезни, воздействия лекарств и т.п.

В подавляющем числе публикаций описывается расчет вероятностей болезней по одномерным распределениям. Переход к многомерным распределениям снял бы многие проблемы. Но сбор данных, достаточных для построения многомерных распределений, в большинстве случаев неосвоен. Вместе с тем, используя доступную статистику, в ряде случаев можно по одномерным распределениям построить многомерные за счет медицинских знаний вследствие понимания взаимосвязи симптомов и анализов.

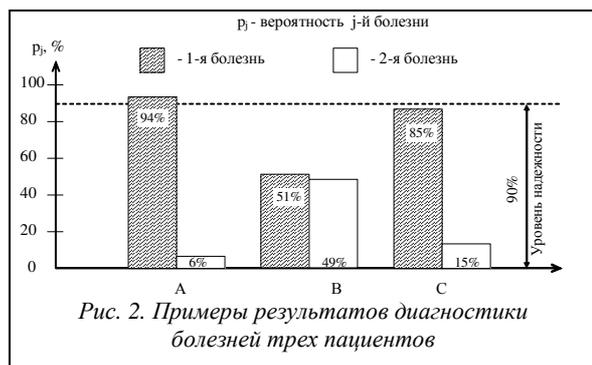
Переход к многомерным распределениям важен, как минимум, для диагностики с расчетом вероятности болезни по формуле Байеса. Диагностические признаки заболевания взаимосвязаны. Многомерное распределение органически содержит в себе взаимосвязь признаков заболевания, поэтому построение таких распределений избавляет от необходимости учета взаимосвязности признаков.

Предельный случай – многомерное распределение всех признаков заболевания $q_j(a_1, a_2, a_3, \dots, a_m)$, где j – номер болезни, a_i ($i=1, 2, 3, \dots, m$) – признак заболевания. Если бы удалось построить такое распределение, отпала бы необходимость в

расчете вероятности болезни по формуле Байеса. Действительно, определив у больного конкретные величины признаков a_i : $a_i=a_{01}$, $a_i=a_{02}$ и т.п., сразу по зависимости $q_j(a_1, a_2, a_3, \dots, a_m)$ получаем плотность вероятности q_i . После этого сравниваются q_i у распознаваемых болезней $j=1, 2, 3, \dots, n$ и ставится диагноз; диагнозом считается болезнь, в распределении которой q_i больше.

Рассмотрим теперь практическую диагностику с расчетом вероятностей болезней по формуле Байеса. Результатом диагностики конкретного пациента являются вероятности каждой болезни. Сумма вероятностей болезней равна единице. По рассчитанным вероятностям ставится диагноз. Если болезней две и вероятность первой болезни (p_1) не ниже вероятности второй болезни (p_2), диагнозом может считаться первая болезнь.

Рассмотрим случай, когда $p_1=51\%$, $p_2=49\%$ (рис. 2, В). Здесь вероятности болезней практически неразличимы и делать однозначный вывод, по мнению авторов, нельзя. В другом случае, когда $p_1=94\%$, $p_2=6\%$, можно более уверенно говорить о диагнозе (рис. 2, А).



Если вероятность одной из болезней не только не ниже вероятностей других болезней, но и превышает установленный заранее уровень надежности диагноза, выбор делается в пользу этой болезни. Установим уровень надежности, равный 90%. Для примера А получаем диагноз – болезнь 1, для примеров В и С диагноз неопределенный, то есть программа отказывается сделать выбор в пользу первой болезни и требует больше данных. Такой подход к оценке вероятностей позволяет перевести часть неправильно поставленных диагнозов в область неопределенных ответов. Одновременно некоторые правильные диагнозы также становятся неопределенными (как в примере С). Представляется, что важнее сделать меньше ошибок в диагнозе, чем большему числу пациентов поставить верный диагноз.

Созданная методика реализована в виде универсальной компьютерной программы, которая используется для диагностики механической и паренхиматозной желтух; точность диагностики – 97% правильных, 1% неправильных и 2% неопределенных диагнозов. Программа написана на

языке C++ *Builder 6.0*, имеет интуитивно понятный интерфейс, время диагностики одного пациента около 10 секунд. Программный продукт позволяет вести БД больных и заниматься научными исследованиями. Врач может смотреть динамику изменения ДП больного на фоне графика изменения ДП при обычном течении каждой из болезней, анализировать гистограммы ДП в разные дни заболевания, получать не только диагноз, но и влияние каждого ДП на него и т.п. Программа внедрена в трех больницах Алтайского края.

Разработанная методика диагностики сравнивалась с лучшими методами диагностики: дискриминантный анализ, деревья классификации, нейронные сети. Напрямую сравнить методы не удалось. Чтобы воспользоваться известными программами, нужно для каждого больного иметь абсолютно все ДП, чего в реальной больнице не бывает. (Авторская программа работает с реальными ДП, определенными не в каждый день; она функционирует и тогда, когда определена только часть ДП.) В известные программы авторы ввели свои методические наработки, после чего эти программы в 90 % случаев стали показывать верный диагноз. Затем эти программы научили учитывать динамику заболеваний. Только после этого дискриминантный анализ и деревья классификации достигли точности диагностики 92 % и 93 %, а у ней-

ронных сетей правильных диагнозов стало 96 % (у авторской методики – 97 %). То есть качественную диагностику нейронные сети показали только с использованием авторских наработок, включая учет динамики. Количество неверных диагнозов у нейронных сетей – 3 %, у авторской методики – 1 %.

В ходе сравнения было предположено, что нейросетевой (искусственных нейронных сетей) и вероятностный подходы могут быть взаимоусилены совмещением, состоящим в использовании в качестве исходных данных для искусственных нейронных сетей не исходных статистических данных, а их вероятностных характеристик.

Литература

1. Жмудяк М.Л., Повалихин А.Н., Стребуков А.В., Гайнер А.В., Жмудяк А.Л., Устинов Г.Г. Диагностика заболеваний методами теории вероятностей. Барнаул: Изд-во АлтГТУ, 2006. 168 с.
2. Филиппова Н.Н., Комарова М.В. Вирусные гепатиты. URL: www.labdiagn.h1.ru/hepatit.shtml (дата обращения: 07.08.08).
3. Аммосов А.Д. Гепатит В. URL: <http://www.vectorbest.ru/brosh/gepb.htm> (дата обращения: 05.08.08).
4. Хазанов А.И. Функциональная диагностика болезней печени. М.: Медицина, 1988. 304 с.
5. Harrison's principles of Internal Medicine. 16th Edition. Editors: D.L. Kasper, E. Braunwald, A.S. Fauci, S.L. Hauser, D.L. Longo, J.L. Jameson. McGraw-Hill, Medical Publishing Division, New York, 2005, 2783 p.

РАЗРАБОТКА АЛГОРИТМА И ПРОГРАММЫ УПРАВЛЕНИЯ ПРОЦЕССОМ ЭПИДЕМИИ

*Е.А. Андреева, д.ф.-м.н. (Тверской государственной университет);
Н.И. Овсянникова (Поморский государственный университет, г. Архангельск,
natmat68@mail.ru)*

Рассматривается задача поиска оптимального управления эпидемией с помощью вакцинации, карантина и информационно-просветительской программы «Здоровье» в неоднородном сообществе, состоящем из четырех возрастных групп.

Ключевые слова: оптимальное управление эпидемией, алгоритм численного решения, программное управление.

Существует немало программных продуктов, позволяющих прогнозировать процесс эпидемии, строить динамику распространения инфекции в однородном сообществе, зная начальные данные и интенсивность управления. Данный программный продукт в отличие от предыдущих учитывает три различных вида управления, направленных на погашение инфекции сразу в n различных группах.

Рассмотрим динамику управляемого процесса распространения эпидемии в неоднородном сообществе, состоящем из n возрастных групп, в виде следующей системы дифференциальных уравнений:

$$\begin{cases} \dot{x}_i(t) = -x_i(t)(1-g_i(t))\sum_{j=1}^n \beta_{ij}y_j(t) - \mu_i x_i(t) + \Lambda_i - v_i(t)x_i(t), \\ i = \overline{1, n} \\ \dot{y}_i(t) = x_i(t)(1-g_i(t))\sum_{j=1}^n \beta_{ij}y_j(t) - \mu_i y_i(t) - \tilde{\mu}_i y_i(t) - \gamma_i y_i(t) - u_i(t)y_i(t), \\ i = \overline{1, n} \\ x_i(0) = x_i^{(0)}, y_i(0) = y_i^{(0)}, \\ x_i(t) \geq 0, y_i(t) \geq 0, t \in [0, T], \end{cases} \quad (1)$$

где $x_i(t)$ и $y_i(t)$ – число подверженных заболеванию и инфицированных в i -й группе ($i = \overline{1, \dots, n}$) в момент t ; $x_i(t)\sum_{j=1}^n \beta_{ij}y_j(t)$ – функция, характери-