
ДИАГНОСТИЧЕСКИЕ АЛГОРИТМЫ

Н. С. Мисюк
А. М. Гурленя
В. В. Лозовик

1995

Н. С. Мисюк, А. М. Гурленя, В. В. Лозовик

ДИАГНОСТИЧЕСКИЕ АЛГОРИТМЫ

**Издательство «Вышэйшая школа»
Минск 1970**

615
М65
УДК 616.8-07



70-19953

Диагностические алгоритмы. Мисюк Н. С. и др.
Минск, «Вышэйш. школа», 1970.

Настоящее руководство, рассчитанное на широкий круг врачей и студентов медицинских институтов, преследует цель помочь осуществлению автоматизированного распознавания болезней. Для ознакомления читателя с теоретическими основами построения алгоритмов приводятся краткие сведения из теории вероятностей и случайных чисел. Рассматриваются общие принципы автоматизации диагностического процесса в медицине, основу которого составляет разработка диагностического алгоритма. Излагаются принципы классификации алгоритмов, приводятся наиболее распространенные в медицинской практике диагностические алгоритмы (вероятностные, метод фазового интервала, логического базиса и т. д.) для распознавания болезней при помощи электронных вычислительных машин (ЭВМ). Представлены также алгоритмы, основанные на последовательном статистическом анализе и других методах, которые не требуют для своей реализации применения ЭВМ. Приводятся краткие сведения по кибернетической обработке диагностических данных. В заключительной главе излагаются общие сведения о самообучении диагностических систем, основанном на автоматической коррекции информативных данных самой ЭВМ в процессе ее работы. — Табл. 18. Илл. 4. Библ. в конце глав.

ПРЕДИСЛОВИЕ

Эффективность диагностики заболевания человека повышается как путем углубления знаний о болезнях, так и путем изыскания новых возможностей правильного анализа и оценки информации о больном и его заболевании.

Многие поколения врачей по крупицам собирали сведения о болезнях, воссоздавая их клиническую картину. Без этого было бы невысказимо формирование современной медицины. Это путь старый, но тем не менее не утративший своего значения. Поэтому любое исследование, добывающее новое сведение о давно известном заболевании, новая методика, позволяющая приоткрыть завесу над неразгаданными его тайнами, достойны внимания, так как способствуют углублению наших представлений о болезнях человека. Однако нельзя не признать, что коэффициент полезного действия исследователей, идущих по этому пути, уже не тот, что у наших учителей и их предшественников, так как абсолютное большинство заболеваний описано, а диагностика их в основном разработана.

Трудности распознавания заболеваний подчас обусловлены не столько недостаточностью информации о них, сколько сложностью ее анализа и оценки. Сказанное прежде всего относится к молодым врачам. А между тем многочисленные рекомендации новых и новейших методик ставят перед

ними трудную задачу, обусловленную нарастающим потоком информации, так как анализ и правильная оценка ее усложняются. И то, и другое требуют не только разносторонних знаний, но и большого опыта.

А как быть недостаточно опытному врачу, когда жизнь ставит перед ним сложные диагностические задачи, к решению которых он еще не подготовлен?

Большую помощь ему может принести автоматизация диагностического процесса.

Под автоматизированным распознаванием болезни понимают постановку диагноза посредством определенного диагностического алгоритма, поддающегося математическому описанию. Оно может быть осуществлено электронной вычислительной машиной (ЭВМ) или при помощи специальных таблиц. В первом случае говорят о машинной, а во втором — о табличной диагностике. Этот путь повышения эффективности диагностики заболеваний человека открывает перед клиницистами новые возможности, которые еще предстоит оценить. А это существенно меняет положение врача, которому проще собрать информацию (жалобы, анамнез, объективные данные о состоянии больного и др.), нежели проанализировать, точно оценить ее и принять правильное решение. Автоматизация диагностического процесса, если даже она касается только одного из его этапов — переработки информации, в значительной мере может улучшить работу врача.

Для автоматизированного распознавания болезней при помощи ЭВМ необходимо прежде всего составить программу. В связи с этим надо очертить круг заболеваний, которые предполагается диагностировать таким образом, и избрать алгоритм, согласно которому будут решаться диагностические задачи.

В течение последних лет разработан ряд диагностических алгоритмов. Однако знакомство с ними затруднено тем, что

эти алгоритмы опубликованы в различных статьях и сборниках, подчас изданных небольшими тиражами.

В настоящем пособии, предназначенном для студентов медицинских институтов и врачей, собраны наиболее распространенные алгоритмы, которыми пользуются клиницисты. Надеемся, что оно облегчит знакомство с основами автоматизации диагностического процесса и будет способствовать ее внедрению в клиническую практику.

Отдавая себе отчет, что эта работа не лишена определенных недостатков, мы с благодарностью примем все замечания товарищей, которые пожелают сделать их.

Считаем своим долгом поблагодарить канд. физ.-мат. наук Н. А. Лепешинского за полезные и деловые замечания, сделанные им в процессе работы над пособием.

Минск, 1970 г.

Авторы

1. СЛУЧАЙНЫЕ СОБЫТИЯ. ВЕРОЯТНОСТЬ

Одно из основных понятий теории вероятностей — понятие случайного события.

Представим себе, что проводится серия экспериментов (исследований), результаты которых заранее предвидеть невозможно. Допустим, что нас интересует какой-то определенный результат проводимого эксперимента. Поскольку предвидеть его невозможно, нельзя заранее с уверенностью утверждать, что определенное событие произойдет. Такие события, наступление которых заранее с полной уверенностью предсказать нельзя, называют случайными, так как именно случай определяет, произойдут они или нет.

Так, например, подбрасывая монету непреднамеренным способом, мы не можем с уверенностью утверждать, какой стороной она упадет: выпадение «решки» оказывается таким же случайным событием, как и выпадение «орла».

Отдельные случайные события в теории вероятностей обычно обозначают буквами A, B, C и т. д. Пусть n — общее число событий проводимого эксперимента. Среди них будет какое-либо число событий, благоприятствующих, например, событию A . Обозначим его буквой m . Тогда вероятностью события A называют отношение числа благоприятных событий m к общему числу n . Обозначив вероятность через $P(A)$, будем иметь

$$P(A) = \frac{m}{n}. \quad (1)$$

В примере с подбрасыванием монеты, допустив равновозможность выпадения «орла» или «решки», легко убедиться, что вероятность выпадения «орла» $\frac{1}{2}$.

Остальная часть событий из числа n не вызовет наступления события A . Такие случайные события называют противоположными A . Их число $n - m$.

Обозначив вероятность противоположного события через $g(A)$, получим

$$g(A) = \frac{n - m}{n} = 1 - \frac{m}{n}.$$

Принимая во внимание выражение (1), имеем

$$g(A) = 1 - P(A).$$

Итак, вероятность случайного события A определяется как отношение числа событий, благоприятствующих A , ко всему числу возможных событий. А вероятность события, противоположного A , равняется единице минус вероятность события A .

Вероятность выражается дробью. Наименьшее значение, которое она может принять, есть нуль. Это значит, что данное событие невозможно. Наибольшее значение дроби — единица, указывающая на достоверность данного события. Во всех остальных случаях

$$0 \leq P(A) \leq 1.$$

Поскольку одно из свойств случайных событий состоит в том, что они происходят с одинаковой частотой, то при большом числе наблюдений (экспериментов) относительная частота наступления случайного события должна равняться его вероятности.

2. УСЛОВНАЯ ВЕРОЯТНОСТЬ

Если мы имеем два последовательно случайных события A и B , то возникает вопрос, какова вероятность наступления второго события, если первое уже произошло. Такую вероятность называют условной и обозначают $P(A/B)$. Условной вероятностью события A относительно события B называют отношение вероятности совместного появления событий A и B к вероятности события B :

$$P(A/B) = \frac{P(AB)}{P(B)}.$$

Проиллюстрируем это примером. В группе, состоящей из 82 детей, 46 сделана прививка против кори. Корью заболело 19 детей, в том числе двое подвергшихся прививке.

Обозначим факт прививки через A , а случай заболевания через B . Вся совокупность события насчитывает 82 единицы, совокупность событий, благоприятствующих A , — 46 единиц, а совокупность событий, благоприятствующих B , — 19.

Вероятность, что взятому наугад ребенку была сделана прививка:

$$P(A) = \frac{46}{82}.$$

Вероятность того, что ребенок, взятый наугад, заболел корью:

$$P(B) = \frac{19}{82}.$$

Какова вероятность заболевания ребенка, которому сделали прививку?

В этом случае вся совокупность событий включает только 46 детей, подвергшихся прививке. Совокупность событий, благоприятствующих заболеванию, несмотря на прививку, состоит из 2.

Следовательно, искомая условная вероятность есть

$$P(B/A) = \frac{2}{46}.$$

Можно также вычислить условную вероятность того, что заболевшему ребенку сделана профилактическая прививка:

$$P(A/B) = \frac{2}{19}.$$

3. ЗАВИСИМОСТЬ МЕЖДУ СЛУЧАЙНЫМИ СОБЫТИЯМИ

Одно из основных понятий теории вероятностей — понятие независимости случайных событий.

Если имеет силу равенство $P(A/B) = P(A)$, то говорят, что событие A независимо от события B , ибо его вероятность не зависит от того, произошло ли событие B или не произошло. Точно так же, если $P(B/A) = P(B)$, говорят, что событие B независимо от события A .

Невыполнение вышеприведенных равенств говорит о зависимости этих событий.

4. ВЕРОЯТНОСТЬ ПРОИЗВЕДЕНИЯ ДВУХ СЛУЧАЙНЫХ СОБЫТИЙ

Вероятность произведения двух случайных событий равна произведению вероятности одного события на условную вероятность другого события, вычисленную в предположении, что первое случайное событие уже произошло:

$$P(AB) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B).$$

В рассмотренном ранее примере с прививкой против кори вероятность того, что взятый наугад из всей группы ребенок подвергался прививке и заболел, составляет

$$P(AB) = \frac{46}{82} \cdot \frac{2}{46} = \frac{19 \cdot 2}{82 \cdot 19} = \frac{2}{82}.$$

5. СЛУЧАЙНЫЕ ПЕРЕМЕННЫЕ. ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Рассмотрим теперь понятие случайной переменной.

Допустим, что в результате проводимого эксперимента могут наступать различные случайные события, причем наступлению каждого из них соответствует некоторое однозначное определенное действительное число. Каждому из этих чисел соответствует вероятность, а именно вероятность определенного события. Переменную, принимающую различные числовые значения в соответствии с определенными вероятностями, называют *случайной переменной*.

Различают случайные переменные двух основных классов — дискретные (прерывные) и непрерывные.

Случайная переменная *дискретна*, если совокупность возможных ее значений конечна или по крайней мере поддается исчислению.

Распределение дискретной случайной переменной можно представить двояко. Во-первых, в форме таблицы. В колонках такой таблицы помещают рядом возможные значения случайной переменной и соответствующие им вероятности. Во-вторых, в форме записи распределения случайной переменной с помощью аналитической формулы общего вида

$$P(X) = P\{X = x\}, \quad (2)$$

выражающей вероятность в виде функции, подобранной для случайной переменной, имеющей значение x . Функция вида (2), показывающая вероятность того, что случайная переменная X примет значение x , носит название функции распределения вероятностей этой случайной переменной.

В противоположность дискретным случайным переменным, совокупность возможных значений *непрерывной* случайной переменной не только не конечна, но и не поддается исчислению. Следовательно, непрерывная слу-

случайная переменная может принимать любое действительное значение в некоторых пределах, конечных или бесконечных.

Случайную переменную X называют **непрерывной**, если она имеет форму $F(x)$, непрерывную в каждой точке x , и если для каждого значения x существует

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \geq 0.$$

Функцию $f(x)$ называют **функцией плотности распределения вероятностей**, короче — **плотностью вероятности**. Ее можно истолковать следующим образом: это есть **количество вероятности**, приходящееся на единицу длины интервала $(x; x + \Delta x)$, когда длина интервала стремится к нулю.

6. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ И ДИСПЕРСИЯ

Наиболее полную информацию о распределении случайной переменной дает функция плотности распределения вероятностей. Однако часто возникает необходимость характеризовать распределение случайной переменной с помощью одного-двух числовых показателей, выражающих наиболее существенные свойства этого распределения. К таким характеристикам распределения относятся математическое ожидание и дисперсия.

Если случайная переменная X дискретна, то математическое ожидание этой случайной переменной определяется как сумма произведений отдельных значений, которые может принимать переменная, на соответствующие им вероятности. Математическое ожидание случайной переменной обозначают символом $E(X)$. Следовательно,

$$E(X) = \sum_x x \cdot P\{X = x\},$$

т. е. суммирование здесь распространяется на все возможные значения случайной переменной.

П р и м е р 1. Случайная переменная X может принимать значения $-1, 0, 1$ с вероятностями соответственно $0,2; 0,3; 0,6$. Математическое ожидание случайной переменной X таково:

$$E(X) = -1 \cdot 0,2 + 0 \cdot 0,3 + 1 \cdot 0,6 = 0,4.$$

Если случайная переменная X непрерывна и принимает значения в интервале (a, b) , то в определении математического ожидания сумма сменяется интегралом. Таким образом,

$$E(X) = \int_a^b x \cdot f(x) \cdot dx.$$

Математическое ожидание указывает, каков средний уровень значений, принимаемых случайной переменной. А дисперсия характеризует степень расхождения этих значений и определяется как математическое ожидание квадрата отклонений случайной переменной от ее математического ожидания. Обозначив дисперсию через $D^2(X)$, имеем

$$D^2(X) = E[X - E(X)]^2.$$

Если X — дискретная случайная переменная, то вычисление дисперсии сводится к следующему:

$$D^2(X) = \sum_x [x - E(X)]^2 \cdot P\{X = x\}, \quad (3)$$

причем суммирование распространяется на все возможные значения переменной. Если X есть непрерывная случайная переменная с плотностью вероятности $f(x)$, то для получения дисперсии необходимо вычислить интеграл:

$$D^2(X) = \int_a^b [x - E(X)]^2 \cdot f(x) dx.$$

Пример 2. Вычислить дисперсию случайной переменной из примера 1. Поскольку X есть дискретная переменная по формуле (3) и $E(X) = 0,4$, получим

$$D^2(X) = (-1 - 0,4)^2 \cdot 0,2 + (0 - 0,4)^2 \cdot 0,3 + (1 - 0,4)^2 \cdot 0,6 = 0,656.$$

Считаем, что вышеизложенные краткие сведения из теории случайных функций и теории вероятностей будут достаточны для понимания диагностических алгоритмов.

1. ЭТАПЫ ДИАГНОСТИЧЕСКОГО ПРОЦЕССА

В диагностическом процессе можно выделить несколько этапов, взаимно связанных между собой:

- а) сбор анамнестических данных;
- б) обнаружение субъективных и объективных признаков болезни;
- в) анализ и оценка полученных данных;
- г) определение характера и локализация патологического процесса;
- д) установление прогноза.

Перечисленные этапы имеют большое значение для определения характера заболевания. Однако ценность каждого из них в значительной мере зависит от своеобразия болезней и особенностей больного.

Так, для постановки диагноза мигрени, неврастении или психастении наиболее важны жалобы больного, ибо объективная симптоматика этих заболеваний скудна, а при некоторых разновидностях опухолей (рак нижней губы, шейки матки и др.), синингомиелии и других заболеваниях — данные объективного исследования, так как жалобы у таких больных могут отсутствовать, а начало болезни остается незамеченным. В итоге болезнь обнаруживается «случайно», и больной не в состоянии назвать время ее возникновения.

2. ПЕРВИЧНАЯ МЕДИЦИНСКАЯ ДОКУМЕНТАЦИЯ

Основным первичным медицинским документом, от содержания которого в большой степени зависит точность диагностики, является история болезни. В ней содержатся все сведения о больном и его заболевании (жалобы, анамнез, данные объективного исследования и др.). Она как бы изоморфна заболеванию. Так должно быть. По крайней мере к этому стремится каждый врач, когда заполняет историю болезни. Полноценность этого документа в значительной мере предопределяет качество медицинских документов и возможность их научной обработки.

К сожалению, нельзя не признать, что этот важнейший первичный документ имеет ряд существенных изъянов, затрудняющих возможность его автоматической обработки. Отсутствие стандарта истории болезни, хотя бы для наиболее распространенных заболеваний, импровизация при составлении ее приводят к тому, что в ней нередко содержится информация, не поддающаяся вероятностной оценке. Подчас она насыщена избыточными сведениями, но не содержит данных, необходимых для диагностики и прогнозирования заболевания. Разные врачи при описании одного и того же заболевания используют различную терминологию, а обнаруженные признаки оценивают главным образом качественно («повышено», «понижено» и т. д.), т. е. субъективно.

Следовательно, необходимо создание стандартизированной, унифицированной истории болезни с полным перечнем вопросов, на которые должен ответить врач при обследовании больного. Учитывая, что у некоторых людей болезнь может протекать необычно, нужно предусмотреть в истории болезни графу, куда врач запишет обнаруженные им признаки, обычно отсутствующие у таких больных.

Образцы стандартизированных историй болезней разработаны в ряде лечебных и научно-исследовательских учреж-

дений Советского Союза, например: в Институте хирургии им. А. В. Вишневского — для больных с пороками сердца, в Институте неврологии АМН СССР — для сосудистых больных. Приводим образец стандартизированной истории болезни для больных с острыми нарушениями мозгового кровообращения, разработанный сотрудниками клиники нервных болезней Минского медицинского института и Института неврологии АМН СССР (приложение 1).

Такая история болезни облегчает работу врача, значительно сокращает время на ее заполнение. Наряду с этим врач, заполняя графы, неизбежно должен уделить больше внимания больному, чтобы не пропустить ни одного указанного в истории болезни признака.

При наличии архива стандартизированных историй болезней имеется возможность для вероятностной оценки любого из указанных в ней признаков или сочетаний их. Они становятся сравнимыми между собой и в значительной мере облегчают создание промежуточного носителя информации (перфокарта, перфолента), необходимого для ЭВМ.

3. ФОРМАЛИЗОВАННЫЙ ПЕРВИЧНЫЙ ДОКУМЕНТ

Несмотря на ряд достоинств, у стандартизированной истории болезни есть существенный изъян — невозможность непосредственной обработки при помощи ЭВМ. Такая история болезни не может быть считана ЭВМ. Для этого обязательно нужен промежуточный носитель информации — перфокарта, перфолента. В свою очередь, информация с промежуточного носителя не может быть считана врачом без ЭВМ. Это, естественно, сопряжено с тратой дополнительных средств и удлиняет время обработки первичной документации.

Создание формализованного первичного документа (истории болезни, лабораторные анализы и др.) устраняет не-

обходимость в промежуточном носителе информации. Подобный документ компактен и считывается как врачом, так и ЭВМ. Правда, для этого необходимо специальное считывающее устройство, например «Бланк».

Итак, клиницист должен располагать такой историей болезни, которая может быть заполнена врачом и воспринята вычислительной машиной без всяких дополнительных операций по переносу информации на промежуточный носитель (В. К. Надененко, 1968). Эта история болезни уже будет своеобразным формализованным документом.

Особенно удобна история болезни, в которой использован так называемый «взвешенный код». По сути дела это всего лишь один лист бумаги размером 210×297 мм. Одна сторона его представляет собой универсальное информационное поле, которое делится на ряд колонок, состоящих из набора строк. Каждой клетке такого поля может быть придан любой вес (отсюда — «взвешенный код»), как цифровой, так и алфавитный. Клетка может означать и какое-то понятие (месяц, год, признак болезни, диагноз и др.).

Заполнение такой истории болезни сводится к заштриховыванию карандашом клетки, соответствующей тому или иному сведению. Если, например, у больного болит голова, то зачеркивается клетка, стоящая рядом с признаком «головная боль», и т. д.

Использование формализованной истории болезни позволяет достичь эффекта сразу в четырех направлениях: уменьшается объем истории болезни; сокращается время заполнения; появляется возможность ее обработки на ЭВМ и оперативного получения результатов для практических и научных целей.

При использовании формализованного документа можно применять не только «взвешенный код», но и так называемый «стилизированный шрифт», отличающийся от общепринятых шрифтов некоторыми ограничениями, способствующими

Сибирский институт
Сибирский институт
СИБИРСКИЙ
ЧИТАЛЬНЫЙ ЗАЛ

щими автоматическому распознаванию таких признаков считывающим устройством.

Это устройство представляет собой механизм, обеспечивающий перелистывание листов, отделение их от пачки, транспортировку по одному в считывающий узел и затем в один из приемных карманов. В считывающем узле лист освещается мощным источником света. При этом отраженный свет воспринимается фоточувствительными датчиками, различающими наличие или отсутствие метки. Затем происходит логическая обработка полученной информации, которая передается в ЭВМ (В. К. Надененко, 1968).

Формализованная история болезни разработана сотрудниками кафедры нервных болезней Минского государственного медицинского института (О. М. Вировлянский, А. С. Мاستыкин, Н. С. Мисюк) и Минского проектного бюро завода им. Г. К. Орджоникидзе (Н. Г. Антонова, В. К. Надененко, Н. С. Хасдан). Образец такой истории болезни для больных с острыми нарушениями мозгового кровообращения помещен в приложении 2.

Приведенные образцы историй болезней соответствуют тем требованиям, которые предъявляются к ним: а) обеспечивают простоту и удобство записей; б) легко считываются врачом; в) значительно сокращается время, затраченное на заполнение истории болезни; г) содержащаяся в них информация достаточна для диагностики и назначения рационального лечения; д) информация представлена в форме, удобной для кодирования и последующей обработки с помощью ЭВМ (в одном случае) либо для непосредственного считывания с помощью устройства «Бланк» (в другом случае).

Само собой разумеется, что применение унифицированной истории болезни не исключает возможности обычной описательной формы изложения, например в дневниковых записях (С. А. Колесников, Т. Б. Постнова, 1968).

4. ХРАНЕНИЕ ПЕРВИЧНЫХ ДОКУМЕНТОВ

Первичная документация (история болезни, результаты анализов, рентгенограммы и т. д.) обычно хранится в медицинских архивах в виде оригиналов или заключений специалистов. При существующей форме истории болезней занимают много места, а их разработка сопряжена с рядом трудностей. Использование унифицированных, а тем более формализованных историй болезней устраняет эти недостатки.

Информацию о больном и его заболевании удобно хранить в закодированном виде (перфокарты, перфоленты, магнитные ленты). Наличие систематизированного и кодированного архива способствует обработке его при помощи электронных вычислительных или более простых машин (сортировки, табулятора и т. д.). Таким образом, обеспечивается быстрый и достаточно точный поиск клинического прецедента, получение нужных статистических сведений о частоте встречаемости и величине информационных весов (диагностической ценности) тех или иных признаков, выяснение корреляции между отдельными симптомами и синдромами, процесс самообучения ЭВМ для диагностики, прогнозирования и оценки результатов лечения (С. А. Колесников, Т. Б. Постнова, 1968).

5. СИСТЕМАТИЗАЦИЯ ИНФОРМАЦИИ

Создание унифицированных историй болезней сопряжено с необходимостью систематизации и унификации различных признаков. В связи с тем, что обследование больных и регистрация его результатов в различных клиниках производится по-разному и набор исследуемых признаков неодинаков, истории болезней подчас бывает трудно сопоставить (Н. И. Моисеева, 1967).

Не менее важной представляется необходимость упорядочения терминологии, так как обилие признаков, наличие дублирующих симптомов, разнообразие методики изучения одного и того же явления крайне затрудняют сравнение данных различных исследователей или лечебных учреждений, а прямая передача этих данных на ЭВМ либо невозможна, либо сопряжена с большими трудностями.

Для устранения перечисленных недочетов по каждой медицинской специальности нужно создать единую универсальную схему обследования и описания полученных данных. Так, например, целесообразна разработка единой подробной формы для истории болезни применительно к каждому классу заболеваний, чтобы на основе автоматической информационной системы пользоваться опытом многих клиник (М. Л. Быховский, 1963). Это неизбежно сопряжено с выбором наиболее рациональных методик исследования из числа существующих и разработкой новых. Необходимо только добиться такого положения, чтобы введенные стандартные схемы обследования были приняты повсеместно.

Бережное отношение к накопленному клиническому опыту диктует целесообразность создания четкой схемы регистрации различных признаков, обеспечивающей сопоставление данных, полученных с применением самых различных методик исследования и описанных не по стандарту. Такая схема, например, для больных с поражением центральной нервной системы предложена Н. И. Моисеевой (1967).

6. ЗАДАЧИ МАТЕМАТИЧЕСКОЙ ДИАГНОСТИКИ

Сформулированы четыре основные задачи диагностики (С. А. Колесников, Т. Б. Постнова, 1968):

1. Постановка объективного диагноза на основе обобщения накопленного медицинского опыта.

2. Выяснение достаточности медицинской информации для постановки диагноза.

3. Выработка рекомендаций по наиболее целесообразному маршруту обследования.

4. Оценка эффективности лечения и прогнозирования результатов.

7. ПРИМЕНЕНИЕ МАТЕМАТИКИ ДЛЯ ДИАГНОСТИКИ

Применение математики для диагностики болезней человека прежде всего состоит в том, что с помощью ЭВМ или специальных диагностических таблиц на основании определенного набора признаков ставится диагноз, имеющий наибольшую степень вероятности. Такая диагностическая система может охватить всю сумму сведений о данной группе заболеваний, что может оказаться весьма полезным даже высококвалифицированному врачу.

Понятие «диагностическая система» включает диагностическую таблицу и логику диагноза. Если логика вероятностная, то диагностическая таблица содержит совокупность признаков с указанием их вероятности для каждого входящего в диагностическую систему заболевания (Т. Б. Постнова, 1968).

Так, признак S_j будет абсолютно достоверным для данного заболевания, если он регистрируется в 100% случаев (сыпь при кори, белок в моче при остром нефрите и т. д.). Вероятность этого признака принимается за единицу. В общем случае вероятность любого признака при заболевании D_i будет равна

$$P(S_j/D_i) = M/N, \quad (4)$$

где M — число больных, имеющих признак S_j ;

N — общее число больных с данным заболеванием.

Величина $P(S_j/D_i)$ определяется на основании данных

медицинской статистики, результатов обработки архивного материала и литературных данных. $P(S_j/D_i)$ будет тем достовернее, чем больше число N .

Значение вероятностей признаков, определяемое формулой (4), находится в пределах от 0 до 1 ($0 \leq P(S_j/D_i) \leq 1$).

Так как в медицинской статистике вероятность принято выражать в процентах, то $P(S_j/D_i) = M/N \cdot 100\%$.

Математическая диагностика считает диагноз установленным, если его вероятность не менее 80—90% (Т. Б. Постнова, 1968).

Литература

Антонова Н. Г., Вировлянский О. М., Мастыкин А. С., Мисюк Н. С., Надененко В. К., Хасдан Н. С. О возможности применения системы «Бланк» в невропатологии. — Электронные вычислительные машины в невропатологии. Минск, 1968, 39.

Быховский М. Л. Кибернетика и медицина. — Биологическая и медицинская электроника, вып. 1. М., 1963.

Ивахненко А. Г., Лапа В. Г. Кибернетические предсказывающие устройства. Киев, 1965.

Колесников С. А., Постнова Т. Б. Некоторые вопросы обработки медицинской информации при помощи электронных вычислительных машин. — Вычислительная техника в физиологии и медицине. М., 1968, 95.

Моисеева Н. И. Проблемы машинного диагноза в неврологии. Л., 1967.

Моисеева Н. И. Подготовка медицинских данных для обработки на электронной вычислительной машине с целью установления диагноза. — Вычислительная техника в физиологии и медицине. М., 1968, 111.

Надененко В. К. Способы ввода информации с формализованных документов в электронную вычислительную машину. — Электронные вычислительные машины в невропатологии. Минск, 1968, 37.

Постнова Т. Б. Некоторые вопросы математической диагностики. — Вычислительная техника в физиологии и медицине. М., 1968, 104.

1. ЧТО ТАКОЕ ДИАГНОСТИЧЕСКИЙ АЛГОРИТМ?

А л г о р и т м — совокупность правил, которые определяют содержание и последовательность операций, преобразующих исходные данные в искомый результат. Если он используется для распознавания болезней человека, то его называют диагностическим алгоритмом.

2. КЛАССИФИКАЦИЯ

Диагностические алгоритмы можно разделить по группам, основанным на различных принципах (С. А. Колесников, Т. Б. Постнова, 1968).

Вероятностный метод. Использует значения условной вероятности появления признаков, характерных для данного заболевания, и априорную вероятность. Вероятность различных диагнозов определяется по формуле Байеса.

Метод логического базиса. В его основе лежит рассмотрение синдромов, или комбинаций признаков для данного заболевания. Затем при помощи медицинских сведений исключаются нехарактерные комбинации признаков болезни. Таким образом, диагноз ставится методом исключения из списка заболеваний, синдромы которых не совпадают с синдромами, обнаруженными у больного.

Метод фазового интервала. По идее близок методу логического базиса. Диагноз устанавливает-

ся по минимальному числу несовпадений признаков больного с характерной совокупностью признаков для данного заболевания.

Статистический метод. Состоит в статистическом сопоставлении комплекса признаков, обнаруженных у больного, с комплексами признаков ранее выявленных случаев, хранящихся в архиве или в памяти ЭВМ.

Каждая из перечисленных групп алгоритмов имеет свои разновидности. Возможно комбинирование методов.

3. АЛГОРИТМЫ

1) Вероятностные

а) Распознавание болезней с помощью формулы Байеса

Для установления диагноза широко используется формула Байеса. Приведем один пример, представив при этом диагностическую таблицу, составленную для трех заболеваний и содержащую только четыре признака.

Условно обозначим диагнозы D_1, D_2, D_3 , а признаки — S_1, S_2, S_3, S_4 .

Допустим, что по данным статистики

| | | | | | |
|-----------|----------------------------------|-------|---|----|---|
| при D_1 | S_1 встречается в 90 % случаев | | | | |
| | | S_2 | » | 0 | » |
| | | S_3 | » | 5 | » |
| | | S_4 | » | 60 | » |
| при D_2 | | S_1 | » | 15 | » |
| | | S_2 | » | 80 | » |
| | | S_3 | » | 80 | » |
| | | S_4 | » | 80 | » |
| при D_3 | | S_1 | » | 10 | » |
| | | S_2 | » | 95 | » |
| | | S_3 | » | 90 | » |
| | | S_4 | » | 10 | » |

У больного предполагается одно из указанных заболеваний, которое следует определить.

Допустим, согласно статистике, среди рассматриваемой группы заболеваний диагноз D_1 встречается у 35% больных, D_2 — у 15 и D_3 — у 50% больных.

Эти данные можно представить в виде табл. 1.

Таблица 1

Диагностическая таблица

| Диагноз | Частота встречаемости диагноза $P(D_i)$ | Вероятность признаков | | | |
|---------|---|-----------------------|--------------|--------------|--------------|
| | | $P(S_1/D_i)$ | $P(S_2/D_i)$ | $P(S_3/D_i)$ | $P(S_4/D_i)$ |
| | | S_1 | S_2 | S_3 | S_4 |
| D_1 | 0,35 | 0,9 | 0 | 0,05 | 0,6 |
| D_2 | 0,15 | 0,15 | 0,8 | 0,8 | 0,8 |
| D_3 | 0,5 | 0,1 | 0,95 | 0,9 | 0,1 |

Вероятность диагноза D_i при наличии синдрома S определяется по формуле Байеса.

$$P(D_i/S) = \frac{P(D_i) \cdot P(S/D_i)}{P(S)}, \quad (5)$$

где $P(D_i)$ — вероятность заболевания с диагнозом D_i среди рассматриваемой группы заболеваний;

$P(S/D_i)$ — вероятность появления синдрома S при диагнозе D_i ;

$P(S)$ — вероятность того, что у больного имеется синдром S .

Полагая, что признаки S_1, S_2, S_3 и S_4 независимы, будем иметь

$$P(S/D_i) = P(S_1/D_i) \cdot P(S_2/D_i) \cdot P(S_3/D_i) \cdot P(S_4/D_i); \quad (6)$$

$$P(S) = \sum_{i=1}^3 P(D_i) \cdot P(S_1/D_i) \cdot P(S_2/D_i) \cdot P(S_3/D_i) \cdot P(S_4/D_i). \quad (7)$$

Прежде всего следует вычислить вероятности диагнозов в том случае, где у больного проявились все четыре признака.

Тогда из формул (5)—(7) можно будет найти вероятности диагнозов D_1 , D_2 и D_3 :

$$P(D_1/S) = \frac{0,35 \cdot 0,9 \cdot 0 \cdot 0,05 \cdot 0,6}{0,35 \cdot 0,9 \cdot 0 \cdot 0,05 \cdot 0,6 + 0,15 \cdot 0,15 \cdot 0,8 \cdot 0,8 \cdot 0,8 + 0,5 \cdot 0,1 \cdot 0,95 \cdot 0,9 \cdot 0,1} = 0;$$

$$P(D_2/S) = \frac{0,15 \cdot 0,15 \cdot 0,8 \cdot 0,8 \cdot 0,8}{0,35 \cdot 0,9 \cdot 0 \cdot 0,05 \cdot 0,6 + 0,15 \cdot 0,15 \cdot 0,8 \cdot 0,8 \cdot 0,8 + 0,5 \cdot 0,1 \cdot 0,95 \cdot 0,9 \cdot 0,1} = 0,73;$$

$$P(D_3/S) = \frac{0,5 \cdot 0,1 \cdot 0,95 \cdot 0,9 \cdot 0,1}{0,35 \cdot 0,9 \cdot 0 \cdot 0,05 \cdot 0,6 + 0,15 \cdot 0,15 \cdot 0,8 \cdot 0,8 \cdot 0,8 + 0,5 \cdot 0,1 \cdot 0,95 \cdot 0,9 \cdot 0,1} = 0,27.$$

В итоге наиболее вероятным оказался диагноз D_2 .

Теперь рассмотрим такой случай, когда у больного признак S_1 отсутствует, но имеются все остальные признаки. Здесь вероятность отсутствия признака S_1

$$P(\bar{S}_1/D_i) = 1 - P(S_1/D_i).$$

Далее расчет проводится аналогичным образом, только вероятность $P(S_1/D_i)$ в формулах (6) и (7) заменяется на $P(\bar{S}_1/D_i)$. В итоге получим $P(D_1/S) = 0$; $P(D_2/S) = 0,63$; $P(D_3/S) = 0,37$.

Табл. 2 показывает наибольшую вероятность (0,93%) диагноза D_3 , для которого характерно отсутствие признака S_4 при наличии всех остальных признаков. В некоторых случаях вероятность диагноза невелика. Поэтому требуется проведение дополнительных исследований.

Т а б л и ц а 2

Вероятности диагнозов при наличии
всех признаков и при отсутствии одного из них

| Диагноз | Все признаки | Отсутствие | | | |
|---------|--------------|------------|-------|-------|-------|
| | | S_1 | S_2 | S_3 | S_4 |
| D_1 | 0 | 0 | 0,75 | 0 | 0 |
| D_2 | 0,73 | 0,63 | 0,23 | 0,86 | 0,07 |
| D_3 | 0,27 | 0,37 | 0,02 | 0,14 | 0,93 |

Такие расчеты можно провести и вручную при наличии небольшого числа диагнозов и присущих им признаков. Однако если их много, то вычисление лучше делать на ЭВМ (Р. Ледли, Л. Ластед, 1963; Т. Б. Постнова, 1968).

Теперь следует обратить внимание на зависимость и независимость отдельных признаков, так как это важно для разработки диагностической таблицы.

б) О вероятностной зависимости признаков

Для построения диагностической системы, основанной на вычислении вероятностей заболеваний, в большем числе случаев отдельные признаки вероятности считают независимыми. Это позволяет упростить математические формулы для вычисления вероятности заболеваний.

Свойство вероятностно независимых признаков используют в диагностике для вычисления условной вероятности некоторой системы признаков $B_k (S_1, \dots, S_k)$ при некотором заболевании D_j :

$$P(B_k/D_j) = P(S_1/D_j) \cdot P(S_2/D_j) \dots P(S_k/D_j).$$

Эта формула, называемая формулой умножения, при наличии зависимых признаков приобретает следующий вид:

$$P(S_1, S_2) = P(S_1) \cdot P_{S_1}(S_2), \quad (8)$$

где $P_{S_1}(S_2)$ — условная вероятность признака S_2 , если имеется признак S_1 .

Если рассматривается вопрос о вероятностной зависимости признаков, то следует различать два случая: непосредственную зависимость и опосредствованную.

Непосредственная зависимость возникает при физическом механизме связи, который носит стохастический характер, между рассматриваемыми признаками и является инвариантным по отношению к рассматриваемым состояниям организма.

Примером может служить зависимость между застойными сосками зрительных нервов и повышением внутричерепного давления.

Опосредствованная зависимость возникает между признаками через третий фактор. Если мы возьмем подмножество индивидуумов с указанным третьим фактором или подмножество, не обладающее им, то у каждого из этих подмножеств рассматриваемые признаки будут независимыми, так как для каждого из них третий фактор перестал быть случайной величиной.

Доказательство теоремы: два признака, независимые между собой на подмножестве данного фактора B_j , зависимы на множестве, включающем \bar{B}_j и B_j , если они

информативны для данного фактора (вероятностно связаны с ним) или если

$$P_{B_j}(S_i) \neq P(S_i).$$

Если взять признаки S_1 и S_2 , то

$$\left. \begin{aligned} P(S_1) &= P(B_j) \cdot P_{B_j}(S_1) + P(\bar{B}_j) \cdot P_{\bar{B}_j}(S_1); \\ P(S_2) &= P(B_j) \cdot P_{B_j}(S_2) + P(\bar{B}_j) \cdot P_{\bar{B}_j}(S_2). \end{aligned} \right\} \quad (9)$$

Для совокупности признаков

$$P(S_1, S_2) = P(B_j) \cdot P_{B_j}(S_1, S_2) + P(\bar{B}_j) \cdot P_{\bar{B}_j}(S_1, S_2). \quad (10)$$

Если признаки S_1 и S_2 независимы на подмножествах B_j и \bar{B}_j , то из выражения (10) имеем

$$\begin{aligned} P(S_1, S_2) &= P(B_j) \cdot P_{B_j}(S_1) \cdot P_{B_j}(S_2) + P(\bar{B}_j) \cdot P_{\bar{B}_j}(S_1) \times \\ &\quad \times P_{\bar{B}_j}(S_2); \end{aligned} \quad (11)$$

из выражений (9) — (11) следует, что

$$P(S_1, S_2) \neq P(S_1) \cdot P(S_2).$$

Теорема, таким образом, доказана.

Указанные признаки становятся зависимыми на множестве, включающем B_j и \bar{B}_j . Исключением является только случай, когда

$$P_{B_j}(S_1) = P_{\bar{B}_j}(S_1) = P(S_1)$$

или

$$P_{B_j}(S_2) = P_{\bar{B}_j}(S_2) = P(S_2).$$

В таком случае $P(S_1, S_2) = P(S_1) \cdot P(S_2)$. Однако при этом один из указанных признаков становится неинформативным по отношению к B_j , т. е. вероятностно независимым от третьего фактора.

Вполне очевидно, что если хотя бы для одного состояния или заболевания признаки S_1 и S_2 независимы, то их зависимость при других состояниях является опосредствованной. Если бы их зависимость была непосредственной, то они были бы зависимы при всех состояниях.

Если будет установлено, что зависимость между признаками S_1 и S_2 для каждого состояния B_k опосредствованная, то, подразделив его на два подсостояния (B'_k и B''_k) в соответствии с наличием и отсутствием этого третьего фактора, получим новый набор состояний или заболеваний, когда признаки S_1 и S_2 будут независимыми для каждого из них.

Такой подход, предложенный М. Л. Быховским (1967), может оказаться полезным при построении медицинской памяти системы, приспособленной для применения формулы (8).

в) Логическая схема диагностического процесса

Диагностический процесс, осуществляемый с помощью ЭВМ, сопряжен с необходимостью разработки алгоритма. Он выполняется поэтапно. На каждом этапе производится серия исследований больного, дается оценка ситуации, возникшей в результате этих исследований, а если необходимо, то назначаются новые исследования (М. Л. Быховский, 1968).

Как же выглядит логически схема диагностического процесса? Для уяснения ее приведем табл. 3, в которой по горизонтали даются признаки S_i , а по вертикали — заболевания D_j . При этом в каждой клетке отмечается вероятность признака S_i при заболевании D_j , или $P(S_i/D_j)$, а в левом столбце таблицы — априорные вероятности заболеваний $P(D_j)$.

Логическая схема диагностического процесса

| $P(D_j)$ | D_j | S_i | | | | Примечание |
|----------|-------|-------|-------|--------------|-----|---------------------------|
| | | S_1 | S_2 | S_3 | ... | |
| $P(D_1)$ | D_1 | ... | 1 | ... | ... | ... |
| $P(D_2)$ | D_2 | 0 | ... | $P(S_3/D_2)$ | ... | $\delta_2(S_1, S_2, S_3)$ |
| $P(D_3)$ | D_3 | ... | ... | ... | ... | ... |

Признак S_i , всегда встречающийся при заболевании D_j , должен быть отмечен в этой таблице как 1, а признак, никогда не встречающийся, — как 0. Наряду с этим в графе «Примечание» проставлена группа симптомов δ_j , если, конечно, она существует, полностью определяющая заболевание D_j . Таким образом, если у больного обнаружена группа δ_j , то у него имеется болезнь D_j .

Приведенная таблица — одна из форм концентрации медицинского опыта, памяти в данном классе заболеваний.

Такой логический процесс диагностического мышления, предлагаемый М. Л. Быховским (1968), включает в себя несколько ступеней, а именно: детерминистскую логику, информационно-вероятностную логику и логику фазового интервала.

Допустим, что сведения о больном, полученные при его обследовании, составляют следующую систему признаков: $S_{(x)} = \{x_1, x_2, x_3, x_4, x_5, x_6 \dots\}$, где индекс указывает номер проведенного исследования (номер признака); $x_i = 1$ при наличии признака и $x_i = 0$ при его отсутствии.

Первый этап диагностического процесса начинается с детерминистской логики. Она состоит в том, что на код $S_{(x)}$

последовательно накладываются все коды δ_j , т. е. проводится операция сравнения. Первое совпадение по δ_j указывает номер заболевания D_j и тем самым дает окончательный ответ.

Если этого нет, то исключают все те заболевания D_j , а точнее, их детерминистские коды (нули и единицы в соответствующих клетках таблицы), не совпадающие по соответствующим разрядам с кодом $S_{(x)}$. Это осуществляется путем последовательного наложения в машине всех детерминистских кодов D_j на код $S_{(x)}$. По окончании этой процедуры остается ряд заболеваний, которые возможны при данной клинической картине. На этом заканчивается этап детерминистской логики.

Второй этап диагностического процесса, основанный на вероятностной логике, состоит в выяснении, какое из оставшихся заболеваний имеет место в действительности.

Для этого все признаки заболевания, подлежащие уточнению, разбиваются на независимые группы, после чего в соответствии с проведенными исследованиями выбирается по одному признаку из каждой группы (со своим уточняющим), чтобы вычислить для них a_{ji} , если $x_i = 1$, и \bar{a}_{ji} , если $x_i = 0$. Затем определяется вероятность. Такая операция проводится для всех признаков данной группы с последующим усреднением. Последнее можно проводить и до вычисления вероятностей.

По формуле

$$a_{D_j, S_{(x)}} = a_{D_j, S_1} + a_{D_j, S_2} + \dots + a_{D_j, S_k}$$

для каждого D_j вычисляется информационная мера $a_{D_j, S_{(x)}}$ и логарифм вероятности каждого из заболеваний D_j при данной системе признаков $S_{(x)}$:

$$\log P(D_j/S_{(x)}) = a_{D_j, S_{(x)}} - a_{D_j}. \quad (12)$$

При этом должно быть

$$\sum P(D_j/S_{(x)}) = 1. \quad (13)$$

Для этого нужно выполнить операцию нормализации, приводящую к выражению (13). Результаты, полученные при этом по формуле (12), могут трактоваться как вероятности.

Далее проводится оценка результата вычислений сравнением полученных в формуле (12) вероятностей с некоторыми критериями, порогами T_j . Если для некоторого D_j порог T_j достигнут, то это D_j и является диагнозом. Если же такого D_j нет, то назначается следующее исследование S_{i+1} , относящееся к группе более трудных, которые вносят максимальную информацию в возникшую ситуацию, максимально уменьшая ее неопределенность (энтропию). Неопределенность системы возможных диагнозов оценивается величиной энтропии, предложенной Винером и Шенноном:

$$H(D) = - \sum_j P(D_j/D) \log P(D_j/D).$$

С этой целью вычисляется количество информации, вносимое в среднем в данную ситуацию исследованием S_k . Эта информация измеряется разностью энтропии системы:

$$J(S_k) = H(D) - HS_{(k)}(D),$$

где $H(D)$ — неопределенность (энтропия) системы диагнозов до начала исследования;

$HS_{(k)}(D)$ — неопределенность системы диагнозов после проведения исследования S_k .

В связи с этим

$$HS_{(k)}(D) = - \sum_v P(S_k^v) \sum_j P(D_j/D, S_k^v) \log P(D_j/D, S_k^v), \quad (14)$$

где S_k^v — различные возможные исходы исследования S_k .

Вычисление производится по формуле (14) для различных исследований и назначается то из них S_{i+1} , для которого результат вычислений максимален (М. Л. Быховский, 1968).

2) Информационная мера системы признаков

Информационная мера признака — это количество информации, содержащейся в данном признаке S_i относительно заболевания D_j , т. е.

$$\alpha_{D_j, S_i} = \log \frac{P(D_j/S_i)}{P(D_j)}.$$

Это количество информации равно количеству информации, содержащейся в заболевании D_j относительно признака S_i , т. е.

$$\alpha_{D_j, S_i} = \log \frac{P(S_i/D_j)}{P(S_i)} = \alpha_{S_i, D_j},$$

где $P(D_j/S_i)$ — вероятность заболевания D_j при наличии признака S_i ;

$P(D_j)$ — априорная вероятность заболевания D_j .

По аналогии для системы признаков $D_k(S_1, S_2, \dots, S_k)$ информационная мера равна

$$\alpha_{D_j, D_k} = \log \frac{P(D_j/D_k)}{P(D_j)} \quad (15)$$

или

$$\alpha_{D_j, D_k} = \alpha_{D_k, D_j} = \log \frac{P(D_k/D_j)}{P(D_k)}$$

(М. Л. Быховский, А. А. Вишневский, С. Ш. Харнас, 1961; М. Л. Быховский, 1968).

Весьма важной в проблеме диагностики является возможность вычисления информационной меры системы признаков $D_k(S_1, S_2, \dots, S_k)$ по информационным мерам отдельных признаков.

Показано, что

$$\alpha_{D_j, D_k} = \alpha_{D_j, S_1} + \alpha_{D_j, S_2/S_1} + \dots + \alpha_{D_j, S_k/S_1, \dots, S_{k-1}}, \quad (16)$$

где

$$\begin{aligned} \alpha_{D_j, S_i/S_1, \dots, S_{i-1}} &= \log \frac{P(D_j/S_1, \dots, S_i)}{P(D_j/S_1, \dots, S_{i-1})} = \\ &= \log \left[\frac{P(S_i/D_j)}{P(S_i)} \right]_{S_1, \dots, S_{i-1}} \end{aligned}$$

есть количество информации, содержащееся в признаке S_i , относительно D_j , если известно, что признаки S_1, \dots, S_{i-1} имеются (М. Л. Быховский, 1962).

Выражение (15) показывает дополнительную информацию относительно болезни D_j , содержащуюся в признаке S_i , если известны признаки S_1, \dots, S_{i-1} . Признак S_i называют независимым относительно признаков S_1, \dots, S_{i-1} , если эта дополнительно вносимая им информация равна его информационной мере, т. е. если

$$\alpha_{D_j, S_i/S_1, \dots, S_{i-1}} = \alpha_{D_j, S_i}.$$

В случае независимости всех признаков выражение (16) переходит в равенство

$$\alpha_{D_j, D_k} = \alpha_{D_j, S_1} + \alpha_{D_j, S_2} + \dots + \alpha_{D_j, S_k}. \quad (17)$$

В итоге информационная мера системы признаков вычисляется по формуле (17), если признаки независимы, и по формуле (16), если они зависимы (М. Л. Быховский, 1968).

д) Матричный алгоритм с оценкой веса симптомов

Характеристика. Матричным алгоритмом называется такой алгоритм, который обеспечивает распознавание болезней при помощи специальных таблиц, или матриц, содержащих набор диагнозов и признаков. В зависимости от объема содержащейся в них информации алгоритм может быть реализован вручную или технически в виде релейной схемы, диодной матрицы или программы для ЭВМ. Наиболее

удачным представляется матричный алгоритм, где каждый признак получает оценку, которая учитывается в общей сумме при постановке диагноза (В. В. Парин, Р. М. Баевский, 1966).

Разработка диагностической системы с применением этого алгоритма состоит в том, что прежде всего составляют перечень заболеваний (список диагнозов), которые предполагается распознавать таким образом. Затем на каждое из них заготавливается диагностическая карта, содержащая набор признаков, которые могут иметь место при нем. Далее из всех диагностических карт делается выборка признаков с целью составления общего списка их. В таком списке перечисляются симптомы, встречающиеся при всех заболеваниях, включенных в перечень. Те симптомы, которые могут наблюдаться при разных заболеваниях, включаются в этот список только один раз.

Теперь, когда имеется перечень диагнозов и симптомов, необходимо составить диагностическую матрицу. Для этого на специально разлинованной или обычной миллиметровой бумаге по горизонтали перечисляют диагнозы рассматриваемых заболеваний, а по вертикали — присущие им признаки (симптомы).

Когда это сделано, необходимо определить вес каждого признака для постановки каждого из перечисленных диагнозов. Оценка веса производится согласно табл. 4.

В итоге матрица содержит сотни или тысячи таких оценок, стоящих в местах пересечения линий признак — диагноз. Так, в матрице для дифференциальной и топической диагностики опухолей головного мозга содержится свыше 60 диагнозов, 500 симптомов и более 30 000 оценок веса. Для реализации этой матрицы необходима ЭВМ.

Приводим фрагмент такой матрицы (табл. 5.).

Разумеется, что как диагнозы, так и признаки должны быть закодированы.

Таблица 4

Оценка веса признака

| № п. п. | Характеристика признака | Код |
|---------|---------------------------------------|-----|
| 1 | Симптом очень характерен для диагноза | 3 |
| 2 | Симптом характерен для диагноза | 2 |
| 3 | Симптом не характерен для диагноза | 1 |
| 4 | Симптом безразличен для диагноза | 0 |

Таблица 5

Диагностическая матрица (фрагмент)

| Признак | Диагноз | | | | |
|-------------------------------------|---------------------|------------------|------------------|----------------------|----------------|
| | Опухоль лобной доли | Опухоль мозжечка | Опухоль гипофиза | Опухоль IV желудочка | Опухоль ствола |
| Гемиплегия | 2 | 1 | 1 | 1 | 0 |
| Гемианестезия | 1 | 1 | 1 | 1 | 0 |
| Нистагм | 0 | 3 | 1 | 2 | 2 |
| Слепота | 0 | 0 | 0 | 0 | 1 |
| Застойные соски | 2 | 3 | 1 | 3 | 0 |
| Первичная атрофия зрительных нервов | 0 | 1 | 3 | 1 | 1 |
| Гемианопсия | 1 | 1 | 3 | 1 | 1 |

ЭВМ, запрограммированная для решения диагностической задачи согласно этому алгоритму, в процессе решения делает перебор введенной в нее информации (набор признаков) и выдает на печать диагнозы в порядке убывания значения их вероятности. Наиболее вероятен первый из них, наименее вероятен — последний.

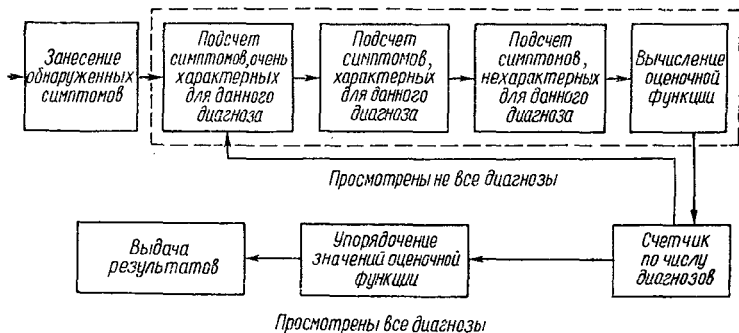
Например:

```
55 07 07 00
32 05 05 00
31 04 06 00
33 03 05 00
```

В этом случае первые две цифры указывают код диагноза, две вторые — число очень характерных симптомов, две третьи — число безразличных признаков и, наконец, две четвертые — число нехарактерных симптомов. При наличии другого типа печатающего устройства перечисляются названия заболеваний (Н. С. Мисюк, О. А. Лисковец, Н. А. Лепешинский, А. С. Мастыкин, 1964; Н. С. Мисюк, Н. А. Лепешинский, А. С. Мастыкин, 1964).

В изложенном логико-вероятностном матричном алгоритме важное место занимает разработка матрицы отношений каждого признака к каждому диагнозу из рассматриваемого класса заболеваний. С помощью этой матрицы для любого набора симптомов, обнаруженных у больного, каждому диагнозу определенным образом можно поставить в соответствие число — значение оценочной функции. Диагноз, для которого ее величина наибольшая, считается наиболее подходящим при обнаруженном наборе симптомов. В случаях, где к наибольшему значению оценочной функции будут близки некоторые другие, потребуется дополнительное исследование для уточнения наиболее вероятного предположения или для подтверждения смешанного характера заболеваний.

Реализация на ЭВМ. Реализация описанного алгоритма на ЭВМ осуществляется по следующей блок-схеме:



При небольшом числе возможных диагнозов этот алгоритм может быть использован для подсчета значений оценочной функции вручную. При подсчете с помощью универсальной ЭВМ номера обнаруженных симптомов для каждого случая заносятся автоматически в запоминающее устройство в любом порядке. Каждый имеющийся номер дает возможность в специально отведенном массиве ячеек памяти машины занести единицу в соответствующий разряд. Так, если в исходном списке 252 симптома, то при 36-разрядных ячейках памяти семь ячеек ($7 \times 36 = 252$) нужно отвести для накопления признаков о наличии симптомов. В связи с этим, если номер обнаруженного симптома равен 110, единица заносится во второй разряд четвертой ячейки из нашего массива.

Для подсчета количества очень характерных, характерных и нехарактерных симптомов для данного диагноза (из числа обнаруженных у больного симптомов) можно

использовать следующий прием, существенно уменьшающий время работы на ЭВМ.

Исходную матрицу, характеризующую отношение симптомов к диагнозам, записывают двумя массивами. Дело в том, что в ней использованы коды 3, 2, 1, 0. В ЭВМ эти коды запишутся 11, 10, 01, 00 соответственно. Каждому диагнозу в матрице соответствует столбец цифр-кодов. Первые (старшие) разряды столбца следует записать в один массив, а вторые (младшие) разряды — в другой. Такая запись дает возможность просто определить количество, например, троек с помощью операции поразрядного умножения.

Приводим пример. В матрице отношений симптомов к диагнозам имеется часть столбца с кодами 3. 2.1.0.3.1.3.2.2. Запишем этот столбец двумя массивами.

Массив старших разрядов: 1.1.0.0.1.0.1.1.1.

Массив младших разрядов: 1.0.1.0.1.1.1.0.0.

Если поразрядно умножить эти два массива друг на друга

$$\begin{array}{r} \times \begin{array}{ccc} 110 & 010 & 111 \\ 101 & 011 & 100 \end{array} \\ \hline \begin{array}{ccc} 100 & 010 & 100 \end{array} \end{array}$$

и подсчитать количество единиц в произведении, то оно укажет на число имеющихся троек в столбце. Логическое умножение последнего результата на содержимое соответствующей ячейки из массива обнаруженных симптомов аналогично дает возможность подсчитать число очень характерных симптомов из числа обнаруженных. Для подсчета двоек вначале массив младших разрядов инвертируется, т. е. выполняется операция поразрядного отрицания по правилу

$$\bar{0} = 1, \bar{1} = 0.$$

- В примере инвертированный массив младших разрядов будет 0.1.0.1.0.0.0.1.1. Теперь остается выполнить операции поразрядного умножения между массивами старших разрядов, инвертированными младшими разрядами, и массивом обнаруженных симптомов и подсчитать число единиц в окончательном результате. Это число дает количество двоек (характерных симптомов) среди обнаруженных симптомов.

Аналогичные схемы можно использовать для комбинации алгоритма поиска клинического процента с логико-вероятностным алгоритмом (Н. А. Лепешинский, 1968).

2) Метод фазового интервала

а) Представление о методе

В основе метода фазового интервала лежит предположение, что состояние организма и его функции можно описать при помощи определенной системы параметров x_1, x_2, \dots, x_n . В таком случае здоровому организму присуща одна система значений этих параметров, а больному организму — другие значения. При допущении, что эти параметры есть оси некоторой системы координат, каждая совокупность их значений является некоторой точкой в этой системе. Исходя из этого, любое состояние организма изображают представляющей точкой в пространстве параметров.

Если в этом пространстве отмечать представляющие точки здоровых людей, то можно получить область, называемую областью нормального состояния (A). Представляющие точки людей с определенным заболеванием также образуют некоторую область (B_1, B_2 и т. д.). Это подчеркивает, что одни и те же заболевания у разных людей классифицируются как близкие, но различные состояния.

Указанные области имеют переменную плотность вероятности. При этом через $p(x_1, x_2, \dots, x_n)$ обозначают плот-

ность вероятности, так что $p dx_1, \dots, dx_n$ будут определять вероятность того, что при данном патологическом состоянии параметры системы находятся между x_1 и $x_1 + dx_1$; x_2 и $x_2 + dx_2$; ...; x_n и $x_n + dx_n$.

В итоге как понятие «норма», так и понятие «заболевание» становятся понятиями статистическими, относящимися к определенной области нормальных или патологических состояний. Эти области могут пересекаться, если имеются патологические состояния, близкие по своей клинической картине.

Мерой данной патологической области может быть сумма дисперсий уклонений от ее центра (математического ожидания) по параметрам x_1, \dots, x_n

$$D_B^2 = \sum_{i=1}^n D_{x_i B}^2 = \sum_{i=1}^n \epsilon (x_{\infty i} - x_{\infty B})^2.$$

Если допустить, что состояние того или иного больного определяется не совокупностью абсолютных значений параметров x_i , а их отклонением от индивидуальной нормы

$$x_i = x_i - x_{i0}$$

или что x_i и $x_{\infty i0}$ в первом приближении независимые случайные величины, то

$$D_{B(x_i)}^2 = D_{B(x_i)}^2 + D_{x_{i0}}^2 = D_{B(x_i)}^2 + D_A^2$$

или

$$D_{B(x_i)}^2 = D_{B(x_i)}^2 - D_A^2. \quad (18)$$

Из формулы (18) следует, что если в качестве координат фазового пространства брать не абсолютные значения параметров x_1, \dots, x_n , а их отклонения от индивидуальной нормы x_i , то область нормального состояния стянется в точку, а мера области патологии уменьшится на меру области нормального состояния, т. е. эти области сузятся. Применение

системы параметров x_i при статистической обработке клинического материала и организации памяти ЭВМ значительно уменьшает области патологии, а это способствует более четкому разграничению областей, т. е. получению исходного материала для более точной диагностики болезней.

Если наблюдение за больным начинается, когда он находится в точке M_1 , то прежде всего следует определить, к какой области принадлежит эта точка (постановка диагноза). Обычно относительно M_1 известна только часть координат (x_1, \dots, x_f) , которые получают при обследовании больного. В связи с тем, что часть координат (параметров) остаются неизвестными, положение точки M_1 недетерминировано. По сути дела, здесь имеется совокупность точек с геометрическим местом (x_1, \dots, x_f) , т. е. некоторой поверхностью M_1 . Если эта поверхность пересекает только одну патологическую область, то, несмотря на недостающее число данных относительно точки M_1 , диагноз будет детерминирован. В случае, где поверхность M_1 будет пересекать несколько областей B_1, B_2, B_3 , можно говорить только о вероятностной оценке того или другого диагноза.

С этой целью можно вычислить соответствующие вероятности как интегралы от $p(x_1, \dots, x_n)$ по M_1 внутри областей B_1, B_2, B_3, \dots

$$\left. \begin{aligned} P(M_1/B_1) &= \int p dx_1, \dots, dx_n; \\ P(M_1/B_2) &= \int p dx_1, \dots, dx_n, \end{aligned} \right\} \quad (19)$$

где $P(M/B)$ — условная вероятность M , если имеет место B .

Обычно интерес представляют обратные вероятности, которые получаются следующим образом. Допустим, что поверхность M_1 пересекает k областей B_1, B_2, \dots, B_k , т. е. при совокупности x_1, \dots, x_f возможно k заболеваний. На основании формулы Байеса

$$P(B_1/M_1) = \frac{P(B_1) \cdot P(M_1/B_1)}{\sum_k P(B_k) \cdot P(M_1/B_k)}.$$

Знаменатель в этой формуле для всех B_k постоянен и носит характер нормирующего множителя:

$$A_1 = \sum_k P(B_k) \cdot P(M_1/B_k).$$

Априорная вероятность болезни B_i равна

$$P(B_i/M_1) = \frac{1}{A_1} P(B_i) \cdot P(M_1/B_i). \quad (20)$$

При помощи формулы (20) вычисляют отношения вероятностей различных заболеваний B_1, \dots, B_n при наличии сведений $M_1(x_1, \dots, x_f)$ и устанавливают последовательность диагнозов в порядке убывания вероятностей.

Если сделано еще одно исследование и получен еще один параметр x , то возникает другое геометрическое место $M_2(x_1, \dots, x_f, x_{f+1})$, пересекающее области B_1, \dots, B_k или некоторые из них.

Согласно формуле (19), $P(M_2/B_1), P(M_2/B_2), \dots, P(M_2/B_k)$ и на основании выражения (20)

$$P(B_i/M_2) = \frac{1}{A_2} P(B_i) \cdot P(M_2/B_i),$$

где $A_2 = \sum_k P(B_k) \cdot P(M_2/B_k).$

Новое исследование и совокупность M_2 могут вызвать другое распределение диагнозов в порядке убывания вероятностей или усилить первое.

Из только что изложенного возникает необходимость введения меры достоверности диагноза. Допустим, что измерения значений параметров $x_1, \dots, x_f, \dots, x_n$ произведены абсолютно точно.

В таком случае следует установить, насколько точно

поставлен окончательный диагноз и когда можно прекратить последовательный процесс определения M_1, M_2, \dots, M_i . Для этого необходимо ввести некоторую количественную оценку.

Предположим данный больной имеет заболевание B , если установлено, что вероятность этого факта лежит в пределах

$$1 \geq P(B/M_i) \geq 1 - a.$$

Величина $1 - a$ называется мерой достоверности диагноза, а величина a — мерой неопределенности.

Величина a определяется уровнем наших медицинских знаний в данный момент и одновременно определяет объем памяти в том смысле, сколь маловероятные комбинации она должна хранить. Неточность в определении x_1, \dots, x_f в свою очередь уменьшает вероятность B/M_i или меру достоверности диагноза (М. Л. Быховский, А. А. Вишневский, С. Ш. Харнас, 1961).

б) Логика диагностического процесса, основанного на фазовом интервале

Если точки в фазовом пространстве описывают состояние организма, траектории — динамику развития, а отдельные области этого пространства — различные заболевания, то что же такое фазовый интервал, на котором основана логика диагностического процесса (М. Л. Быховский, 1968)?

Если состояние организма описывать признаками x_1, x_2, \dots, x_n , каждый из которых имеет два значения (1 — есть признак; 0 — нет признака), то в пространстве параметров x_1, x_2, \dots, x_n состояние данного больного представится некоторой точкой $A(x_1^a, x_2^a, \dots, x_n^a)$. Состояние другого будет представлено точкой $B(x_1^b, x_2^b, \dots, x_n^b)$ и т. д.

Интервалом между точками A и B этого пространства можно назвать выражение

$$D_{AB}^- = |x_1^a - x_1^b| + |x_2^a - x_2^b| + \dots + |x_n^a - x_n^b|. \quad (21)$$

В связи с тем, что любое x_i равно либо 1, либо 0,

$$|x_i^a - x_i^b| = \begin{cases} 0, & \text{если } x_i^a = x_i^b \\ 1, & \text{если } x_i^a \neq x_i^b. \end{cases} \quad (22)$$

Если рассматривать совокупность x_1, x_2, \dots, x_n как n -разрядное двоичное число, то интервал между двумя точками A и B будет равен количеству поразрядных несовпадений (m) в двух n -разрядных двоичных числах $A (x_1^a, x_2^a, \dots, x_n^a)$ и $B (x_1^b, x_2^b, \dots, x_n^b)$:

$$D_{AB} = m. \quad (23)$$

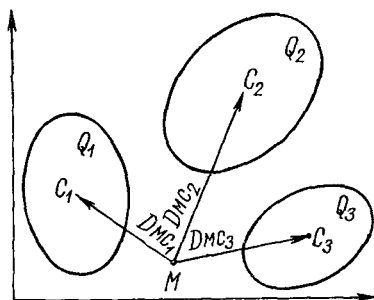
Множеству состояний организма, объединяемых в одну нозологическую форму Q_1 , соответствует множество точек, образующее в фазовом пространстве некоторую область Q_1 . Нозологической форме Q_2 соответствует множество Q_2 и т. д. (рис. 1).

Диаметром множества (D_1, D_2 и т. д.) называют наибольшее расстояние между двумя его точками, а центром множества (C_1, C_2, C_3) — точку области, соответствующую наиболее типичному случаю для данной болезни (см. рис. 1).

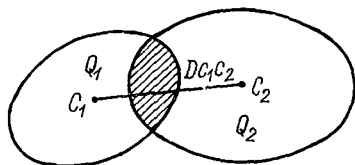
Такой подход позволяет ввести критерий различимости заболеваний. И в самом деле, чем меньше интервал между центрами двух областей, чем больше их диаметр, тем в большей мере эти области перекрывают друг друга, тем большее число точек с равным основанием можно свести к одной из областей (рис. 2). Это, естественно, затрудняет дифференциальную диагностику. В связи с этим вводится критерий различимости двух заболеваний

$$\varepsilon_{ik} = \frac{D_{C_i C_k}}{D_i + D_k} \quad (24)$$

Существует ряд признаков, имеющих для некоторых заболеваний детерминистский характер. Одни из них никогда не встречаются при данном заболевании, а потому их присутствие либо исключает его, либо указывает на наличие другого заболевания. Другие признаки встречаются всегда, следовательно, их отсутствие исключает данное заболевание.



Р и с. 1. Области фазового пространства, соответствующие различным заболеваниям



Р и с. 2. Пересечение фазовых областей

Отсюда следует, что несовпадение по детерминированным признакам делает интервал между данными областями или точкой и областью равным бесконечности. Такие признаки не охватываются формулами (21)—(24) и являются предметом детерминистской диагностической логики (М. Л. Быховский), осуществление которой предшествует как вероятностной логике, так и логике, основанной на принципе фазового интервала. Необходимо считать, что рассматрива-

емое пространство либо не включает детерминированных признаков, либо, при наличии их, имеет место совпадение между всеми рассматриваемыми точками.

Как же выглядит диагностическая логика, основанная на принципе фазового интервала? Допустим, при обследовании больного установлено, что его состояние соответствует некоторой точке M , т. е. какие-то признаки у него есть и соответствующие им x равны 1, другие признаки отсутствуют и эти x равны 0.

Диагностика состоит в определении, к какой области принадлежит точка M . Первоначально можно установить, к центру какой области точка M ближе всего. Для этого по формуле (23) вычисляют интервалы D_{MC_1} , D_{MC_2} , D_{MC_3} (см. рис. 1). Наименьший из них определит ближайшую к точке M область. Для определения принадлежности точки M к данной области вводится понятие диагностического радиуса области R_j . Допустим, что точка M принадлежит к области Q_j , т. е. можно диагностировать Q_j , если $D_{MCj} \leq R_j$.

Известно, что диагностические радиусы области аналогичны порогам T_j в вероятностной логике (М. Л. Быховский) и подобно последним подбираются ЭВМ из максимального отношения количества правильно поставленных диагнозов к общему числу рассматриваемых случаев.

На каждом этапе диагностического процесса приходится иметь дело с неполной системой признаков болезни x_1, \dots, x_ρ ($\rho \leq n$), т. е. с ρ -мерным пространством, в соответствии с тем, что у больного проведено только ρ исследований из n возможных. В связи с тем, что при переходе от ρ -мерной системы к $\rho + 1$ -мерной системе абсолютные длины интервалов изменяются, для сравнения результатов одного этапа с результатами другого предпочтительнее пользоваться не абсолютными, а относительными интервалами, которые равны отношению абсолютного интервала к максимально возможному в данной системе:

$$d_{AB, \rho} = \frac{D_{AB, \rho}}{D_{max, \rho}}$$

или, так как $D_{max, \rho} = \rho$, (максимальное число несовпадений),

$$d_{AB, \rho} = \frac{D_{AB, \rho}}{\rho}.$$

Как было указано выше, в качестве центра области следует брать наиболее типичный случай из имеющейся совокупности. Этот выбор существенно упрощает весь метод, так как при переходе от ρ -мерного к $\rho+1$ -мерному пространству x_1, x_2, \dots, x_ρ координаты центров областей не изменяются, а только добавляется $x_{\rho+1}$ координата.

Выбор из данной совокупности случаев, объединенных одной нозологической формой, наиболее типичного можно осуществить разными способами. Так, в качестве наиболее типичного случая можно взять такой, при котором сумма интервалов от данной точки до всех других точек множества была бы минимальной:

$$\sum_{A_j \subset Q_j} d_{C_j A_j} = \min,$$

либо чтобы максимальным было произведение

$$\prod_{A_j \subset Q_j} (1 - d_{C_j A_j}) = \max.$$

Так как последнее условие значительно чувствительнее, то ему отдают предпочтение.

При вычислении интервала между точками или между данной точкой и центром, где несовпадение по разным признакам имело одинаковый вес, используют формулы (21) и (23). В случае необходимости задача может быть усложнена, если каждому несовпадению будет придан свой вес, а именно $D_{M_1 C_j}$ будет вычислено по формуле

$$D_{M_1 C_j} = \sum_{i=1}^n \beta_{ij} \delta_i$$

где β_{ij} — веса признаков, а $\delta_i = 0$, если признаки x_i совпадают, и $\delta_i = 1$ при несовпадении.

Трудность использования этой формулы состоит в том, что β_{ij} зависит не только от номера признака, но и от номера области, так как для одного заболевания данный признак важный, а для другого — нет, а также от того, что эти веса неизвестны.

Изложенный метод следует использовать не как альтернативу к вероятностной логике, а как дополнение. Поэтому каждый случай желательно подвергать диагностической оценке по вероятностной логике и по методу фазового интервала, а полученные при этом результаты сравнивать (М. Л. Быховский, 1962).

3) Метод логического базиса. Постановка диагноза методом исключения из списка

Этот алгоритм предложен Р. Ледли и Л. Ластедом (1963).

При выработке перечня возможных диагнозов на основании обнаруженных у больного признаков ЭВМ использует аппарат математической логики (исчисление высказываний), что неразрывно связано с понятием комплекса симптом — болезнь (КСБ).

Если комплекс симптомов представляет собой перечень симптомов, которые могут присутствовать или отсутствовать у больного, то комплекс болезней — аналогичный перечень заболеваний. КСБ является перечнем как симптомов, так и болезней, которые могут либо быть, либо отсутствовать у больного.

Логическое значение данных, которые дает клиника в области симптоматологии, состоит в том, что они позволяют свести все множество мыслимых КСБ к значительно меньшему множеству, содержащему только возможные КСБ или совместимые с данными симптоматологии. Обычно КСБ

соответствуют утверждению: если у больного есть определенная болезнь, то у него могут иметь место определенные симптомы.

При логическом анализе списка КСБ удобно пользоваться символической записью. Например, факт отсутствия симптома S_1 и наличия симптома S_2 можно записать в виде \bar{S}_1 и S_2 , а наличие каждого из них — единицей.

Для постановки диагноза рассмотрим случай из двух симптомов (S_1 и S_2) и двух болезней (D_1 и D_2).

| | | | | |
|-------|-----|-----|---|---|
| S_1 | 111 | 111 | 0 | 0 |
| S_2 | 111 | 000 | 1 | 0 |
| D_1 | 100 | 110 | 1 | 0 |
| D_2 | 101 | 101 | 0 | 0 |

Прямоугольной рамкой очерчен столбец, соответствующий рассматриваемому случаю (симптом S_2 есть, а симптома S_1 нет):

| | |
|-------|---|
| S_1 | 0 |
| S_2 | 1 |
| D_1 | 1 |
| D_2 | 0 |

В итоге логический анализ показывает, что у больного с комплексом симптомов \bar{S}_1, S_2 имеется комплекс болезней D_1, \bar{D}_2 , или имеется болезнь D_1 , но нет болезни D_2 .

Если у больного симптомы представлены выражением S_1, \bar{S}_2 , то на указанной выше фигуре этот синдром содержится в столбцах

| | |
|-------|-----|
| S_1 | 111 |
| S_2 | 000 |
| D_1 | 110 |
| D_2 | 101 |

В таком случае для этого больного возможны следующие диагнозы: одновременное наличие D_1 и D_2 , наличие D_1 при отсутствии D_2 , наконец, наличие D_2 при отсутствии D_1 . Отсюда следует, что имеющихся симптомов недостаточно для дифференциальной диагностики этих заболеваний. Поэтому результатом логического анализа будет список возможных диагнозов, соответствующих имеющимся у больного симптомам.

Более формальное представление логического анализа проявляется в следующем виде. Если D_1, D_2, \dots, D_n представляют рассматриваемые болезни, а S_1, S_2, \dots, S_m — признаки, то сведения, полученные при обследовании больного, могут быть выражены булевой функцией

$$E = E[D_1, \dots, D_n, S_1, \dots, S_m].$$

Синдром, обнаруженный у больного, является булевой функцией

$$P = P[S_1, \dots, S_m],$$

и задача постановки диагноза в логическом аспекте сводится к определению функции

$$f = f[D_1, \dots, D_n]$$

из общего перечня возможных диагнозов, удовлетворяющих булеву уравнению (Р. Ледли, Л. Ластед, 1963)

$$E \rightarrow (P \rightarrow f).$$

4) Статистический метод. Поиск клинического прецедента

Данный вид алгоритма предложен и реализован Институтом хирургии им. А. В. Вишневского (И. И. Артоболевский, А. А. Вишневский, М. Л. Быховский, 1962). Он сводится к автоматизированному поиску аналогичного

или тождественного случая, имевшего место в прошлом. Для этого требуется создание информационного массива (в данном случае архива историй болезней). При наличии его рассматриваемый клинический прецедент (признаки заболевания) автоматически сравнивается со случаями информационного массива. Случаи совпадения извлекаются. Если полноты совпадения нет, то находят случаи (или случаи), наиболее близкие к рассматриваемому, и диагноз ставится по аналогии.

Указанные авторы вводят понятие полноты совпадений как отношения $\gamma = \frac{m}{n}$, где m — число общих исследований в двух совпадающих клинических случаях; n — максимальное число исследований (симптомов), принятое для данного класса заболеваний.

Весь информационный массив они предпочитают формулировать в виде двух подмассивов. Один из них содержит синдромы, другой — истории болезней.

Эти подмассивы реализуются на стандартных перфокартах с 80 колонками. В итоге каждому клиническому наблюдению соответствуют две перфокарты. На одной пробита симптоматика, присущая ему; другая содержит диагноз, номер истории болезни и некоторые другие данные.

Сведения, содержащиеся в каждой перфокарте, передаются в электронную схему сравнения. Сюда же подаются данные карты-запроса (перфокарта, содержащая симптоматику исследуемого больного).

Электронная схема сравнения, сопоставляя каждую карту массива с картой запроса, находит совпадающие карты, которые исполнительным механизмом направляются в специальный карман.

По отобранным картам подмассива синдромов автоматически отыскиваются соответствующие им карты подмассива историй болезней, т. е. клинический прецедент.

Техническая реализация предложенной информационной системы была основана на применении счетно-аналитических машин, в частности модифицированного табулятора Т-5, к которому была добавлена специальная электронно-релейная логическая схема (И. И. Артоболевский, А. А. Вишневский, М. Л. Быховский, 1967).

Реализация этого алгоритма может быть упрощена.

При поиске клинического прецедента все симптомы, указанные в историях болезней, в общем «равноправны», а потому возможен такой случай, где по менее характерным признакам для данной болезни будет найден прецедент. В таком случае не исключена диагностическая ошибка. Во избежание ее на кафедре нервных болезней Минского государственного медицинского института использована вероятностная оценка симптомов при данном виде алгоритма (В. В. Лозовик, О. М. Вировлянский, В. К. Пенявский, 1968).

Архив из историй болезней «Острые нарушения мозгового кровообращения» был разбит на подклассы по искомому диагнозу. Каждый подкласс при этом мог беспредельно пополняться.

В каждом подклассе определялась частота появления i -го симптома. Частота $T = \sum_{i=1}^n a_i$, где $a \in \{0, 1\}$: $a = 1$ характеризует наличие симптома, $a = 0$ — отсутствие его; $i = 1, 2, 3, \dots, 300$.

Далее определялся вес каждого симптома в зависимости от частоты его появления $P_i = \frac{T_i}{n}$, где P_i — вес, n — число историй болезней в архиве данного подкласса, $i = 1, 2, 3, \dots, 300$.

Хотя один и тот же симптом часто встречается в различных подклассах данного архива, для одного из подклас-

сов он будет более характерен, нежели для другого. А это и находит свое отражение в определении веса симптомов.

По мере пополнения архива историями болезней веса симптомов уточняются и становятся более достоверными. Таким образом, машина сама объективно по вероятностному принципу оценивает вес каждого симптома для каждого подкласса заболеваний.

Исходя из весовых соотношений по карте-запросу определяется, к какому подклассу относится искомый диагноз. Затем в этом подклассе отыскивается наиболее вероятный прецедент и выдается на печать номер истории болезни, паспортные данные и клинический диагноз прецедента.

Если архив не пополняется, то частота и веса симптомов могут быть записаны на магнитную ленту и храниться в таком виде. При последующих находениях они не вычисляются, а считываются с магнитной ленты, что значительно сокращает время постановки машинного диагноза.

Литература

Артоболевский И. И., Вишневецкий А. А., Быховский М. Л. Автоматическая информационная система отыскания клинического прецедента. — Экспериментальная хирургия и анестезиология, 1962, 3, 3.

Быховский М. Л., Вишневецкий А. А., Харнас С. Ш. Вопросы построения диагностического процесса при помощи математических машин. — Экспериментальная хирургия и анестезиология, 1961, 4, 3.

Быховский М. Л. Вероятностная логика построения самообучающегося диагностического процесса на математических машинах. — Экспериментальная хирургия и анестезиология, 1962, 1, 3.

Быховский М. Л. Вероятностная зависимость признаков в задаче диагностики. — Экспериментальная хирургия и анестезиология, 1967, 2, 75.

Быховский М. Л. Решение задачи диагностики при помощи электронных вычислительных машин. — Вестник АМН СССР, 1968, 5, 41.

Колесников С. А., Постнова Т. Б. Некоторые вопросы обработки медицинской информации при помощи электронных вычислительных машин. — Вычислительная техника в физиологии и медицине. М., 1968, 95.

Ледли Р., Ластед Л. Использование электронных вычислительных машин для обработки медицинских данных в процессе постановки диагноза для отыскания необходимой информации и хранения медицинских записей. — Электроника и кибернетика в биологии и медицине. М., 1963, 315.

Лепешинский Н. А. О реализации логико-вероятностного алгоритма для диагностики болезней человека. — Электронные вычислительные машины в невропатологии. Минск, 1968, 11.

Лозовик В. В., Вировлянский И. М., Пенявский В. К. Введение вероятностной оценки в алгоритм поиск клинического прецедента. — Электронные вычислительные машины в невропатологии. Минск, 1968, 17.

Мисюк Н. С., Лепешинский Н. А., Мاستыкин А. С. Материалы к программированию электронной вычислительной машины для диагностики опухолей головного мозга. Минск, 1964.

Мисюк Н. С., Лепешинский Н. А., Лисковец О. А., Мастыкин А. С. Опыт диагностики опухолей головного мозга при помощи универсальной вычислительной машины «Урал-1». — Журнал невропатологии и психиатрии им. С. С. Корсакова, 1964, 3, 45.

Парин В. В., Баевский Р. М. Введение в медицинскую кибернетику. М.—Прага, 1966.

Постнова Т. Б. Некоторые вопросы математической диагностики. — Вычислительная техника в физиологии и медицине. М., 1968, 104.

1. ТАБЛИЧНЫЕ АЛГОРИТМЫ

Табличными называются алгоритмы, обеспечивающие распознавание заболеваний при помощи специальных диагностических таблиц. Принципы составления таких таблиц могут быть различными.

Различают ряд заболеваний, которые можно диагностировать по простому набору признаков. Примером тому является дифференциальная диагностика геморрагического и ишемического инсультов (табл. 6).

Таблица 6

**Дифференциальная диагностика геморрагического
и ишемического инсультов по простому набору признаков
(вариант первый)**

| Признаки | Геморрагический инсульт | Ишемический инсульт |
|--------------------|-------------------------|---------------------|
| Предвестники | Малохарактерны | Характерны |
| Внезапное развитие | Характерно | Менее характерно |
| Медленное развитие | Нехарактерно | Характерно |
| Утрата сознания | Характерна | Менее характерна |
| Бледность лица | Нехарактерна | Характерна |
| Гиперемия | Характерна | Нехарактерна |

| Признаки | Геморрагический инсульт | Ишемический инсульт |
|-----------------------------|-------------------------|---------------------|
| Артериальная гипотония | Нехарактерна | Характерна |
| Артериальная гипертензия | Характерна | Нехарактерна |
| Гипертермия | Характерна | Нехарактерна |
| Менингизм | Характерен | Нехарактерен |
| Кровь в ликворе | Характерна | Нехарактерна |
| Кровь в ликворе отсутствует | Нехарактерно | Характерно |

Представленный в табл. 6 набор симптомов способствует распознаванию указанных нарушений мозгового кровообращения при условии классического развития болезни и наличии объективных данных, соответствующих табличным. Если представить, что все нарушения мозгового кровообращения будут развиваться в соответствии с указанной схемой, то такую таблицу целесообразнее преобразовать в матрицу-схему «симптомы — болезни», которая будет своеобразным диагностическим алгоритмом, где 1 — наличие симптома, 0 — отсутствие его (табл. 7).

Использование таких таблиц способствует правильной постановке диагноза при условии полного совпадения табличных признаков с признаками симптомокомплекса распознаваемого заболевания. Поэтому они находят широкое применение в дифференциальной диагностике заболеваний, имеющих четкую и типичную клиническую картину, при соблюдении строгих правил сбора информации.

Принципы этой диагностики обладают качеством, позволяющим быстро и правильно исключить признаки, не-

Дифференциальная диагностика геморрагического и ишемического инсультов по простому набору признаков (вариант второй)

| Симптомы | Характер инсульта | |
|-----------------------------|-------------------|-------------|
| | геморрагический | ишемический |
| Предвестники | 0 | 1 |
| Внезапное развитие | 1 | 0 |
| Медленное развитие | 0 | 1 |
| Утрата сознания | 1 | 0 |
| Бледность лица | 0 | 1 |
| Гиперемия | 1 | 0 |
| Артериальная гипотония | 0 | 1 |
| Артериальная гипертония | 1 | 0 |
| Гипертермия | 1 | 0 |
| Менингизм | 1 | 0 |
| Кровь в ликворе | 1 | 0 |
| Кровь в ликворе отсутствует | 0 | 1 |

возможные при данном симптомокомплексе заболевания, и оставить наиболее вероятные.

Диагностическая задача может решаться простым вычислением числа совпадений имеющихся признаков с заданной программой. Большее число совпадений признаков и определит диагноз.

Предположим, что у больного А. выявлены признаки: внезапное развитие заболевания, сознание утрачено, артериальная гипертония, кровь в ликворе. У больного Б.: по-

степенное развитие, сознание утрачено, артериальная гипотония, нормальный ликвор (табл. 8).

Таблица 8

Диагностика геморрагического
и ишемического инсультов
по сумме признаков

| Больные | Сумма признаков | |
|---------|-------------------------|---------------------|
| | геморрагический инсульт | ишемический инсульт |
| А. | 4 | 0 |
| Б. | 1 | 3 |

Из табл. 8 видно, что, суммируя признаки больных А. и Б., у первого больного можно диагностировать геморрагический инсульт, а у второго — ишемический.

Несмотря на определенные достоинства, такие таблицы имеют ряд существенных изъянов, основным из которых является отсутствие оценки веса симптомов, а между тем она может быть очень важна для постановки диагноза.

В связи с этим более совершенны таблицы, в которых учитывается оценка веса каждого признака для постановки любого из рассматриваемых в них диагнозов. Такая таблица, например, разработана для экспресс-диагностики острых нарушений мозгового кровообращения по минимальному объему диагностической информации (Н. С. Мисюк, И. Н. Довнар, 1968) (табл. 9). В ней перечислен 31 признак болезней, что позволяет быстро и с достаточной степенью точности диагностировать один из пяти возможных вариантов нарушения мозгового кровообращения: субарахноидальное кровоизлияние, кровоизлияние в мозг, кровоизлияние в желудочки мозга, тромбоз и эмболия сосу-

**Дифференциальная экспресс-диагностика
острых нарушений мозгового кровообращения**

| № п. п. | Характер нарушения Признаки | Кровоизлияние | | | Инфаркт мозга | |
|------------|-------------------------------------|---------------------------------------|-----------|-----------------------|---------------|---------|
| | | субара- хно- ида- ль- ное | в мозг | в же- лудоч- ки | Гром- боз | Эмболия |
| 1 | Сильная головная боль | 5 | 3 | 0 | 1 | 2 |
| 2 | Психомоторное возбуж- дение | 3 | 1 | 0 | 1 | 2 |
| 3 | Гиперемия лица | 3 | 3 | 2 | 0 | 1 |
| 4 | Бледность лица | 0 | 0 | 0 | 3 | 1 |
| 5 | Сознание сохранено | 2 | 0 | -10 | 3 | 2 |
| 6 | Сознание утрачено быстро | 0 | 3 | 4 | 1 | 2 |
| 7 | Кома | 1 | 2 | 3 | 1 | 0 |
| 8 | Артериальная гипертония | 2 | 3 | 3 | 1 | 0 |
| 9 | Артериальная гипотония | 0 | 0 | 0 | 4 | 1 |
| 10 | Пульс напряжен | 3 | 3 | 3 | 1 | 0 |
| 11 | Сердечная деятельность ослаблена | 0 | 0 | 0 | 4 | 1 |
| 12 | Дыхание расстроено | 0 | 3 | 4 | 0 | 0 |
| 13 | Температура выше 37,4°С | 1 | 2 | 3 | 0 | 0 |
| 14 | Менингизм есть | 5 | 2 | 0 | -3 | 1 |
| 15 | Менингизма нет | 0 | 1 | 0 | 2 | 3 |
| 16 | Гемиплегия | -3 | 5 | 0 | 5 | 3 |
| 17 | Другие очаговые симп- томы | 0 | 3 | 1 | 5 | 3 |
| 18 | Горметония | 0 | 3 | 10 | 0 | 0 |

Продолжение табл. 9

| № п. п. | Характер нарушения Признаки | Кровоизлияние | | | Инфаркт мозга | |
|------------|------------------------------------|-------------------|--------|-------------|---------------|---------|
| | | субарахноидальное | в мозг | в желудочки | Тромбоз | Эмболия |
| 19 | Кровь в ликворе есть | 10 | 4 | 10 | -20 | 0 |
| 20 | Внезапное развитие | 2 | 2 | 1 | 0 | 3 |
| 21 | Медленное развитие | 0 | 0 | 0 | 5 | -0 |
| 22 | Во время, после сна | 0 | 0 | 0 | 5 | 0 |
| 23 | Прием алкоголя | 3 | 3 | 3 | 0 | 0 |
| 24 | Физическое напряжение | 2 | 2 | 2 | 0 | 1 |
| 25 | Эмоциональное напряжение | 2 | 2 | 2 | 0 | 1 |
| 26 | Гипертоническая болезнь | 2 | 3 | 3 | 1 | 0 |
| 27 | Атеросклероз | 2 | 3 | 3 | 4 | 0 |
| 28 | Инфаркт миокарда | 0 | 0 | 0 | 10 | 0 |
| 29 | Ревмокардит | 0 | 0 | 0 | 10 | 0 |
| 30 | Возраст до 50 лет | 3 | 2 | 1 | 0 | 5 |
| 31 | Возраст старше 50 лет | 2 | 3 | 2 | 5 | 0 |
| Сумма | | | | | | |

Диагноз: _____

дов головного мозга. Необходимые сведения могут быть собраны любым врачом.

Каждое из приведенных о больном сведений имеет цифровую оценку. Например, наличие крови в ликворе оценивается при субарахноидальном кровоизлиянии цифрой 10.

Бледность лица при кровоизлияниях получает нулевую оценку, а при тромбозе оценку 3.

Симптомы, не характерные для заболевания, получают отрицательную оценку. Так, наличие крови в ликворе при тромбозе сосудов головного мозга имеет оценку —20. Аналогичные оценки веса даны отсутствию крови в ликворе при субарахноидальном и внутрижелудочковом кровоизлияниях. В итоге такие оценки резко уменьшают вероятность либо исключают постановку указанных диагнозов.

Для постановки диагнозов с помощью приведенной таблицы необходимо уточнить наличие или отсутствие у больного перечисленных в ней признаков. Все, что у больного выявлено, сохраняется. Сохраняются и цифры, стоящие против этих сведений. Например, если у больного имеется головная боль, то соответствующие ей цифры, независимо от характера расстройства мозгового кровообращения, сохраняются. Если же признака нет или проверить его не удалось, то вычеркивают все соответствующие ему цифры. Так, если у больного нет крови в ликворе, то цифры, стоящие против сведения «кровь в ликворе есть», вычеркиваются. Если ликвор не исследован и содержание его остается неизвестным, то цифры, относящиеся к этому сведению, также нужно вычеркнуть. Диагностика проводится только на основе имеющихся сведений.

После того как будут вычеркнуты отсутствующие сведения, по вертикали подсчитывается общая сумма. Наибольшая из них укажет наиболее вероятный диагноз. Так, оптимальная сумма для диагноза субарахноидального кровоизлияния равна 50, кровоизлияния в мозг — 56, кровоизлияния в желудочки мозга — 50, для тромбоза — 39, для эмболии — 42.

Учитывая, что острые нарушения мозгового кровообращения являются всего лишь осложнением основных заболеваний — гипертонической болезни, атеросклероза, инфаркта

миокарда, ревмокардита и других, в таблице перечислены наиболее часто встречающиеся из этих заболеваний; их сохранность после заполнения матрицы позволяет говорить не только о характере нарушения мозгового кровообращения, но и о причине его возникновения.

2. ДИАГНОСТИЧЕСКИЕ АЛГОРИТМЫ, ОСНОВАННЫЕ НА ПОСЛЕДОВАТЕЛЬНОМ СТАТИСТИЧЕСКОМ АНАЛИЗЕ

При дифференциальной диагностике заболеваний некоторые признаки могут отсутствовать. В связи с этим возникает необходимость в разработке методов и правил, мало чувствительных к отсутствию некоторых признаков болезни. Таким требованиям удовлетворяет последовательный статистический анализ А. Вальда (1960), предложенный для дифференциальной диагностики двух сходных заболеваний А. А. Генкиным и Е. В. Гублером (1961, 1962) и нашедший применение для распознавания ряда болезней: разных форм ожоговой болезни (Е. Е. Гублер, Ю. З. Полонский, А. А. Генкин, М. Ю. Кoryтова, 1962—1964; В. И. Легеза, 1964); эпидемического гепатита и хронического холангиогепатита (О. И. Кошиль, Ю. З. Полонский, 1964); сенсорной и моторной афазий при инсультах (Б. В. Иовлев, И. М. Тонконогий, 1966); шизофрении и других сходных состояний при разных формах психических болезней (Б. В. Иовлев, Ф. Ю. Случевский, 1968); облитерирующего эндартериита и ангиоспазмов при некоторых нервных заболеваниях (А. А. Мелодиев, 1968); размягчений и кровоизлияний в мозг (О. А. Балунев, Б. В. Иовлев, Г. З. Левин, И. М. Тонконогий, 1967), распознавания острых заболеваний и угрожающих состояний (Е. В. Гублер, 1968) и др.

По своей математической основе он близок к тем вероятностным методам, которые лежат в основе алгоритмов, используемых для машинной диагностики.

Применяемая для альтернативной диагностики (болезнь *A* или *B*) последовательная статистическая процедура включает следующие операции.

1. Подготовка перечня необходимой симптоматики.
2. Накопление наблюдений с достаточно убедительной верификацией диагноза и составление архива историй болезней.
3. Выделение в архиве историй болезней двух групп:
а) для получения значений диагностических коэффициентов (группа обучения); б) для контроля эффективности математической диагностики.
4. Распределение частот степеней выраженности симптомов для обеих форм заболеваний.
5. Определение отношения частот, отношения правдоподобия, с которыми симптом наблюдается при одном и другом заболевании.
6. Вычисление диагностических коэффициентов.
7. Распределение симптомов в порядке убывания информативности и составление таблицы диагностических коэффициентов.
8. Вычисление пороговых значений суммы диагностических коэффициентов исходя из принятого уровня надежности, с которым требуется выставить диагноз.
9. Сложение диагностических коэффициентов.
10. Постановка одного из двух альтернативных диагнозов в зависимости от того, за какой предел — положительный или отрицательный — вышло значение суммы диагностических коэффициентов.

1) Распределение симптомов

Для выполнения последовательной статистической процедуры в целях дифференциальной диагностики двух болезней *A* и *B* в первую очередь составляется стандартный список

изучаемой симптоматики (анализы крови, мочи, ликвора АД, температура, отдельные симптомы болезни, данные ЭКГ, ЭЭГ, рентгенографии и т. д.).

В архиве историй болезней выделяются две формы болезней — A и B (при этом необходимо, чтобы каждая форма включала не менее 15—20 наблюдений).

После этого производится последовательное распределение симптомов при формах A и B :

$$S_1^A, S_2^A, S_3^A, \dots, S_n^A \\ S_1^B, S_2^B, S_3^B, \dots, S_n^B$$

Эти признаки — случайные величины, зависимость между которыми невелика, но каждый из них содержит определенную информацию о болезни.

Предположим, в качестве формы A взят геморрагический инсульт, а в качестве формы B — ишемический.

Сравнивая распределение первого симптома S_1^A и S_1^B при формах A и B , отмечаем, что в качестве S_1^A и S_1^B используется частота пульса. Общий упорядоченный ряд распределений признака S_1^A и S_1^B при распознаваемых формах A и B представляется в следующем виде:

| | |
|-----|--|
| A | 96; 100; 104; 104; 120; 120; 120; 120; 126; 130; 134 |
| B | 76; 82; 82; 84; 88; 96; 100; 102; 104; 110; 110; 118 |

Для определения порядка расположения одинаковых членов верхнего и нижнего рядов можно использовать таблицу случайных чисел. В этом случае каждому сходному наблюдению присваивается какой-либо номер от 0 до 9. То число, номер которого встретится в таблице случайных чисел раньше, считается большим.

Из табл. 10 видно, что болезнь A имеет более выраженный лейкоцитоз. Несмотря на имеющиеся различия, практически невозможно по количеству лейкоцитов делать выводы о болезни, так как лейкоцитоз является неспецифическим

Пример распределения больных с лейкоцитозом
при болезнях *A* и *B*
(А. А. Генкин, Е. В. Гублер, 1964)

| Больные | Количество лейкоцитов, тыс. | | | | | | | | |
|----------|-----------------------------|-----|--------|---------|---------|---------|---------|---------|-----|
| | 6 | 6—8 | 8,1—10 | 10,1—12 | 12,1—14 | 14,1—16 | 16,1—18 | 18,1—20 | 20 |
| <i>A</i> | 20 | 30 | 50 | 50 | 100 | 200 | 200 | 200 | 150 |
| <i>B</i> | 30 | 40 | 100 | 200 | 300 | 150 | 100 | 50 | 30 |

признаком. Вместе с тем, имея распределение больных по этому признаку, можно заметить, что у больных с лейкоцитозом 13 500 болезнь *A* встречается примерно в три раза реже, чем болезнь *B*. Уже этот показатель дает информацию для дифференциальной диагностики, хотя на основании одного такого признака распознать болезнь нельзя. Диагностика ее возможна в случае подобного распределения других показателей.

2) Оценка различия распределений симптомов

Чтобы оценить степень информативности используемых симптомов, сравнение распределений можно начинать с оценки существенности различий, используя для этого один из непараметрических критериев, в частности критерий *Q* (Розенбаума).

Критерий *Q* основан на сравнении двух упорядоченных рядов наблюдений. Первым считается тот ряд, где максимальная и минимальная величины больше, чем в другом ряду. Подсчитываются *S* — количество наблюдений первого ряда, которое больше максимальной величины второго ряда, и *T* — количество наблюдений второго ряда, которое меньше минимальной величины первого ряда. Если сумма

$Q = S + T$ оказывается достаточно высокой, различия сравниваемых выборок можно считать значительными.

Критическое значение Q для количества наблюдений 11—26 в каждой выборке указано в табл. 11. Минимальная величина Q при $n_1, n_2 > 26$ (n_1 — количество наблюдений первого ряда, n_2 — второго), когда различия следует считать существенными, с $P = 0,05$ составляет 8, а с $P = 0,01$ равна 10.

Таблица 11

Минимальное значение Q для критерия Розенбаума
(по Е. В. Гублеру и А. А. Генкину, 1969)

| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|----|----|----|----|----|------------|----|----|----|----|----|----|----|----|----|----|----|
| 11 | 9 | | | | $P = 0,01$ | | | | | | | | | | | |
| 12 | 9 | 9 | | | | | | | | | | | | | | |
| 13 | 9 | 9 | 9 | | | | | | | | | | | | | |
| 14 | 9 | 9 | 9 | 9 | | | | | | | | | | | | |
| 15 | 9 | 9 | 9 | 9 | | | | | | | | | | | | |
| 16 | 9 | 9 | 9 | 9 | 9 | 9 | | | | | | | | | | |
| 17 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | | | | | | | | | |
| 18 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | | | | | | | | |
| 19 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | | | | | | | |
| 20 | 10 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | | | | | | |
| 21 | 11 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | | | | | |
| 22 | 11 | 11 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | | | | |
| 23 | 11 | 11 | 10 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | | | |
| 24 | 12 | 11 | 11 | 10 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | | |
| 25 | 12 | 11 | 11 | 10 | 10 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | |
| 26 | 12 | 12 | 11 | 11 | 10 | 10 | 10 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|----|----|----|----|----|------------|----|----|----|----|----|----|----|----|----|----|----|
| 11 | 6 | | | | $P = 0,05$ | | | | | | | | | | | |
| 12 | 6 | 6 | | | | | | | | | | | | | | |
| 13 | 6 | 6 | 6 | | | | | | | | | | | | | |
| 14 | 7 | 7 | 6 | 6 | | | | | | | | | | | | |
| 15 | 7 | 7 | 6 | 6 | 6 | | | | | | | | | | | |
| 16 | 7 | 7 | 7 | 7 | 6 | 6 | | | | | | | | | | |
| 17 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | | | | | | | |
| 18 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | | | | | | |
| 19 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | | | | | |
| 20 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | | | | |
| 21 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | | | |
| 22 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | | |
| 23 | 8 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | |
| 24 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 7 | 7 | 7 | 7 | 7 | | |
| 25 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | |
| 26 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 7 | 7 | 7 | 7 | 7 | 7 |

Величина P — вероятность того, что различий между выборками нет; это вероятность ошибочного принятия нулевой гипотезы, т. е. гипотезы об отсутствии различий, когда в действительности различия есть. Если вероятность принятия нулевой гипотезы достаточно мала, делают вывод о значимости (существенности) различий. Считается, что если P не превышает 0,05, наблюдаемые различия неслучайны. Если нулевая гипотеза отвергается, логически можно представить три возможности: первая выборка больше второй; вторая выборка больше первой; первая выборка не равна второй, но мы не знаем, в какую сторону она отли-

чается. Вероятность принятия нулевой гипотезы в последнем случае должна быть в два раза меньше, чем в первом и во втором, для того чтобы мы могли признать различия между выборками значимыми (существенными). Для последнего случая P должно быть не больше 0,025, а для первых двух — не больше 0,05 (Л. С. Каминский, 1964).

В рассмотренном выше упорядоченном ряде распределений (частота пульса) $S = 3$; $T = 5$; $S + T = 3 + 5 = 8$. По таблице определяем, что при $n_1 = 11$, $n_2 = 12$ минимальное значение Q , при котором различия между группами существенны, при $P = 0,05$ равно 7, а при $P = 0,01$ равно 9. Следовательно, в данном случае различия значимы с $P < 0,05$, но $> 0,01$.

Существенность различий может быть определена также по другому непараметрическому критерию (Ван дер Варден, 1960; В. С. Генес, 1964; В. Ю. Урбах, 1964).

Так, для альтернативной диагностики заболеваний A и B , основанной на последовательном статистическом анализе, существенность различия распределения каждого из диагностических признаков может определяться по критерию χ^2 , который является мерой расстояния между распределениями.

$$\chi_j^2 = N_A^{(j)} N_B^{(j)} \left[\sum_{i=1}^{S_j} \frac{1}{A_i^{(j)} + B_i^{(j)}} \left(\frac{A_i^{(j)}}{N_A^{(j)}} - \frac{B_i^{(j)}}{N_B^{(j)}} \right)^2 \right],$$

где $N_A^{(j)}$ и $N_B^{(j)}$ — число больных в группах A и B , данные которых используются для построения распределений j -го признака;

$A_i^{(j)}$ и $B_i^{(j)}$ — частоты появления больных в i -градации j -го признака для сравниваемых групп больных;

S_j — число градаций j -го признака.

Эта мера позволяет получить для каждого признака вероятность того, что соответствующие распределения для болезни *A* и *B* различны, и упорядочить признаки по их информативности (диагностической ценности). Более информативными будут признаки, распределения которых обнаруживают наибольшие различия.

Если существенных различий между распределениями симптома при формах *A* и *B* ни один из критериев не выявит, такой симптом считается малоинформативным и в дальнейшем не используется. В то же время этот симптом может нести некоторую информацию о различиях между сравниваемыми формами. Увеличение числа наблюдений в процессе диагностики пополняет распределения, способствуя выявлению существенных различий. Поэтому при недостатке симптомов целесообразно сохранить для дальнейшей работы и те из них, распределения которых при формах *A* и *B* различаются с уровнем существенности 0,05.

3) Диапазоны распределений симптомов

Для дальнейшего сравнения распределений шкала изменения симптома разбивается на диапазоны, в которых представлены отношения частот болезней *A* и *B*.

При большом количестве наблюдений (свыше 100 для каждой формы) шкала изменений симптома делится на 5—10 равных диапазонов. При числе наблюдений, меньшем 100, иногда могут быть получены диапазоны, неудобные по ряду соображений. В этом случае необходимо руководствоваться определенными принципами. Для сохранения информативности симптома выделяются диапазоны с максимальным различием частот форм *A* и *B* (табл. 12).

Каждый диапазон должен включать не менее 5—7 наблюдений для каждой из двух форм болезни и не меньше одного наблюдения наименее частой (в данном диапазоне)

Диапазоны распределений симптомов

| Показатели | Форма болезни | Диапазоны | | | | Итого |
|------------------------------|---------------|-------------------|----------------|--|------------------------------------|-------|
| | | < 75 | 75 — 84 | 85 — 104 | ≥ 105 | |
| Частота пульса при инсультах | A | 56; 58; 60; 70 | 78; 80; 82; | 87; 89; 90; 91; 100; 102; 103 | 112 | |
| | B | 72; 72 | 74; 78 | 85; 86; 88; 90; 98; 100 | 116; 120; 122; 140; 130; 155 | |
| Число наблюдений | A | 4 | 3 | 7 | 1 | 15 |
| | B | 2 | 2 | 6 | 6 | 16 |

формы. Благодаря этому снижается вероятность недостоверных различий, которые возможны при малом числе наблюдений. Если в каком-либо диапазоне окажется много наблюдений одной формы и будут отсутствовать наблюдения другой (например, $A/B = 55/0$), то нужно искусственно заменить 0 единицей: $A/B = 55/1$.

Необходимо избегать резких различий между отношением двух форм болезни в соседних диапазонах. Если в одном диапазоне $A/B = 1/5$, а в соседнем $4/1$, то лучше выделить промежуточный диапазон, где отношение A/B близко к единице.

Частоты высчитываются как частное от деления числа признаков в диапазоне на общее число признаков в каждой из двух форм. Отношение частот A/B в каждом диапазоне высчитывается в процентах, если общее число наблюдений в каждой из двух форм принято за 100%. Если же это число принимается за единицу, то отношения частот A/B высчитываются в долях единицы.

4) Определение числа правдоподобия

Принцип дифференциальной диагностики двух форм болезни на базе вероятностной оценки признаков состоит в определенном подходе к отношению частот A/B . Если оно равно (табл. 13) $2,16$ ($0,27 : 0,125 = 2,16$), болезнь A в $2,16$ раза вероятнее болезни B . При отношении $A/B = 0,17$ ($0,06 : 0,375 = 0,17$) гипотеза A в $0,17$ раза правдоподобнее гипотезы B , т. е. менее правдоподобна. Поэтому отношение A/B называется отношением правдоподобия.

Аналогично вычисляются отношения правдоподобия для каждого диапазона симптомов S_2, S_3, \dots, S_n .

Таблица 13

Характеристика отношений частот
двух заболеваний

| Показатели | Форма болезни | Диапазоны (частота пульса) | | | | Итого |
|------------------------------|---------------|----------------------------|---------|----------|-------|-------|
| | | < 75 | 75 — 84 | 85 — 104 | ≥ 105 | |
| Число наблюдений | A | 4 | 3 | 7 | 1 | 15 |
| | B | 2 | 2 | 6 | 6 | 16 |
| Частоты | A | 0,27 | 0,20 | 0,47 | 0,06 | |
| | B | 0,125 | 0,125 | 0,375 | 0,375 | |
| Отношение частот | A/B | 2,16 | 1,60 | 1,24 | 0,17 | |
| Диагностические коэффициенты | | + 33 | + 20 | + 9 | - 75 | |

5) Вычисление диагностических коэффициентов

Для удобства обработки данных в процессе математической диагностики логарифмы отношений вероятностей умножаются на 100, что позволяет заменить умножение этих отношений сложением целых дву- или трехзначных чисел.

Величина $\lg A/B \cdot 100$ называется диагностическим коэффициентом. В табл. 13 представлены значения диагностических коэффициентов для дифференциальной диагностики двух форм заболеваний, где одним из признаков является частота пульса.

6) Информативность симптомов

Диагностические коэффициенты со знаком плюс указывают на большую вероятность болезни A , со знаком минус — болезни B . Диагностическая информативность признаков (информация о преобладании вероятности одного из двух диагнозов) прямо пропорциональна величине диагностического коэффициента. В то же время информативность каждого диапазона зависит от частоты нахождения в этом диапазоне больных с формой A и B . Отсюда даже при большом диагностическом коэффициенте, но с редким попаданием больных в этот диапазон диагностическая и информативность его сравнительно невелика.

Информативность каждого диапазона выражается формулой

$$J = k(P_A + P_B),$$

где k — диагностический коэффициент диапазона;

P_A и P_B — вероятности попадания в этот диапазон форм A и B .

Информативность всего симптома J_x определяется суммой информативности его диапазонов $J_x = J_1 + J_2 + J_3 + \dots + J_n$.

Пример: информативность симптома S_1 (частота пульса) по данным табл. 13:

$$J_1 = 33 (0,27 + 0,125) = 13,0$$

$$J_2 = 20 (0,20 + 0,125) = 6,5$$

$$J_3 = 9 (0,47 + 0,38) = 8,1$$

$$J_4 = 75 (0,067 + 0,38) = 33,5$$

$$\underline{J_{x_1} = J_1 + J_2 + J_3 + J_4 = 61,1}$$

После определения информативности каждого симптома их распределяют в порядке ее убывания. При этом учитываются и сроки их получения. Симптомы, полученные раньше при постановке диагноза, правильнее и использовать раньше (в порядке убывания информативности), чем симптомы, выявленные позднее.

В соответствии с теорией последовательного анализа (А. Вальд, 1960; А. А. Генкин, 1962) рассчитывают допустимые вероятности ошибок первого и второго рода (α и β). Ошибкой первого рода (α) считается ошибочный диагноз болезни B , если на самом деле у больного болезнь A ; ошибкой второго рода (β) — ошибочный диагноз болезни A , если у больного имеется болезнь B . Ошибки первого и второго рода могут оказаться неравноценными.

Для определения болезни A пороговое произведение отношений правдоподобия рассчитывают по формуле $\frac{1-\alpha}{\beta}$.

При допустимом уровне ошибок первого и второго рода, равном 0,1 (в среднем одна ошибка на 10 ответов), пороговое произведение равно 9. Это означает, что вероятность болезни A в 9 раз превышает вероятность наличия болезни B .

Пороговое произведение для определения болезни B рассчитывается по формуле $\frac{\alpha}{1-\beta}$.

Так как отношения правдоподобия представлены в логарифмическом виде, то формулы для расчета порогов суммы диагностических коэффициентов приобретают общий вид:

$$\text{болезнь } A: 100 \cdot \lg \frac{1-\alpha}{\beta};$$

$$\text{болезнь } B: 100 \cdot \lg \frac{\alpha}{1-\beta}.$$

После составления диагностической таблицы приступают к практической реализации ее для диагностики. Для этого измеряют величину каждого признака, указанного в таблице, в порядке диагностической информативности и определяют допустимый уровень ошибок и диагностический порог. Суммируют диагностические коэффициенты для каждого признака, пока не будет достигнут порог с плюсом или минусом. При достижении или превышении порога со знаком плюс диагностируют болезнь *A*, при достижении порога со знаком минус диагностируют болезнь *B*.

Если симптомы использованы, но необходимого порога достигнуть не удастся, можно утверждать, что имеющейся диагностической информации недостаточно для постановки диагноза с данным уровнем надежности, и ответ считается неопределенным:

$$\frac{\alpha}{1-\beta} < \frac{S_1^A(S_1^0)}{S_1^B(S_1^0)} < \frac{1-\alpha}{\beta}.$$

При высоких диагностических порогах бывает много неопределенных ответов. В этом случае иногда целесообразно понизить уровень ошибок, чтобы получить определенный ответ, однако при этом может быть допущена диагностическая ошибка. В другом случае, чтобы получить определенный ответ с достаточным уровнем надежности, обращаются за дополнительной информацией.

Диагностические таблицы, составленные на основе по-

сдoleвательной статистической процедуры, обладают большой точностью, простотой и позволяют быстро поставить диагноз. В качестве примера можно сослаться на табл. 14, предложенную Е. В. Гублером и др. (1964) для определения вероятности глубокого ожога, при составлении которой был использован последовательный статистический анализ по Вальду.

Таблица 14

Диагностическая таблица, составленная на основе последовательной статистической процедуры

| | |
|---|--|
| 1. Пульс (ударов в минуту) | 7. Артериальное давление (минимальное) |
| $\frac{< 90}{-133} \quad \frac{90 - 120}{-5} \quad \frac{> 120}{+77}$ | $\frac{< 50}{+77} \quad \frac{50 - 90}{-16} \quad \frac{> 90}{-48}$ |
| 2. Площадь ожога (%) | 8. Палочкоядерные лейкоциты (%) |
| $\frac{20 - 30}{-73} \quad \frac{35 - 50}{+15} \quad \frac{55 - 75}{+67} \quad \frac{80 - 95}{+78}$ | $\frac{0 - 2000}{-33} \quad \frac{2000 - 5500}{+3} \quad \frac{> 5500}{+81}$ |
| 3. Температура утренняя (градусов) | 9. Лейкоциты (тыс.) |
| $\frac{< 36,5}{+54} \quad \frac{36,5 - 37,5}{0} \quad \frac{> 37,5}{-77}$ | $\frac{< 13}{-44} \quad \frac{13 - 33}{-9} \quad \frac{> 33}{+85}$ |
| 4. Температура вечерняя (градусов) | 10. Нейтрофилы (%) |
| $\frac{< 36,5}{+39} \quad \frac{36,5 - 37,5}{+32} \quad \frac{37,5 - 38,5}{-14}$ | $\frac{< 75}{-106} \quad \frac{75 - 85,5}{-16} \quad \frac{> 85,5}{+25}$ |
| $\frac{> 38,5}{-58}$ | 11. Лимфоциты (%) |
| 5. Артериальное давление (максимальное) | $\frac{< 6}{+20} \quad \frac{6 - 13}{-1} \quad \frac{> 13}{-62}$ |
| $\frac{< 100}{+35} \quad \frac{100 - 140}{-39} \quad \frac{> 140}{+9}$ | 12. Цветной показатель |
| 6. Юные формы лейкоцитов (%) | $\frac{< 0,89}{-12} \quad \frac{0,89 - 1,1}{+30} \quad \frac{> 1,1}{-22}$ |
| $\frac{0}{-29} \quad \frac{0,5 - 1}{+12} \quad \frac{\geq 1,5}{+57}$ | |

Последовательный статистический анализ можно использовать не только для дифференциальной диагностики двух болезней, но и для выделения одной болезни из многих, хотя он в этом случае не всегда может обладать оптимальными свойствами.

Применение последовательного статистического анализа дает возможность различать болезни подчас даже не по специфическим признакам, позволяя наиболее полно реализовать информацию, полученную при обследовании больного, и за счет этого дополнить опыт врача новыми методами.

Будучи близким к методам, лежащим в основе составления диагностических программ для ЭВМ, последовательный статистический анализ обладает весьма важным достоинством — не требует применения вычислительной техники.

Для дифференциальной диагностики с помощью таблицы у обследуемого необходимо определить величину первого диагностического показателя (частоту пульса) и по диапазону, в который он попадает, найти диагностический коэффициент. Аналогичная процедура продельвается и со вторым показателем (площадь ожога). Далее эти коэффициенты суммируются. В случае, если не удастся достигнуть пороговой величины ± 127 , указанные выше операции продельваются с другими признаками. При достижении пороговой величины процедура заканчивается, и выносится соответствующее решение. Если сумма всех значений больше $+127$ (вероятность ошибки 5%), то имеется глубокий ожог, при значениях, меньших -127 , глубокого ожога нет.

Если же в результате обработки всех 12 диагностических признаков уровень порога не может быть достигнут, то, очевидно, диагностической информации формы ожоговой болезни недостаточно для правильной дифференциальной диагностики с заданным уровнем надежности.

В процессе реализации этой таблицы для дифференциальной диагностики по существу производится кодирование

взятых признаков значениями «+», «н», «—», при этом каждому из знаков придается свой вес в зависимости от признака, что является не чем иным, как вероятностно-матричной логикой.

Таким образом, использование диагностических таблиц, составленных на большом клиническом материале с помощью последовательного статистического анализа, позволяет во многом облегчить и ускорить решение диагностической задачи. Такие таблицы могут стать основой для дифференциально-диагностических программ ЭВМ.

7) Проверка диагностической таблицы

Для оценки эффективности диагностической таблицы необходимо провести ее проверку. Она производится на группе больных, которые использовались при ее составлении, и на специальной контрольной группе.

Проверяя каждый случай болезни с точно установленным диагнозом, его рассматривают, как если бы он не имел диагноза. Диагноз устанавливается с помощью проверяемой таблицы. Результаты этой проверки могут быть представлены в форме трех ответов: определенный правильный (табличный диагноз совпадает с истинным), определенный ошибочный и неопределенный. Каждый из них рассчитывают в процентах по отношению к общему числу ответов.

Результаты этой проверки могут быть разными. При малой коррелированности симптомов количество ошибок не выходит за пределы заранее намеченного уровня и зависит от информативности и числа симптомов, число же неопределенных ответов может быть любым — от 0 до 100%. При высокой коррелированности симптомов процент ошибок может быть выше намеченного уровня. Определить допустимую степень коррелированности симптомов пока не представляется возможным, хотя теоретически она должна равняться 0.

Вместе с тем при высокой коррелированности симптомов процент ошибок при проверке на специальной группе часто не превышает расчѐта, подтверждая хорошие качества диагностических таблиц. Более надежный результат получается при корреляции таблицы на специальной проверочной группе больных.

3. АЛГОРИТМ РАЗЛИЧЕНИЯ

Для ранней диагностики болезней и функциональных состояний организма (утомление, судорожное состояние, гипоксия и т. д.) необходимо преодолеть трудности, связанные прежде всего с отсутствием значительных отклонений физиологических параметров организма от нормы. В этом случае возникает потребность в правильной регистрации анализа и сведении к минимуму ошибок различения.

Анализ диагностической информации может быть проведен на основании статистических методов количественного анализа физиологических процессов (биоэлектрические явления, дыхание, пиломоторика и т. д.).

Распределения, характеризующие функциональные состояния организма, находят путем рассмотрения интервала, предшествующего предпатологическому и патологическому состоянию, признаки которого известны и четко выражены.

С этой целью предложен алгоритм различения (А. А. Генкин и Г. М. Зараковский, 1964), при помощи которого можно диагностировать умственное утомление, используя для этого показатель ЭЭГ (случай с одним лишь распределением). В качестве этого показателя берется лишь один параметр Δ_T — уровень асимметрии деятельности фаз ЭЭГ за время T .

Распределения этого показателя для умственной деятельности субъекта или группы субъектов в неутомленном состоянии — $\delta_1(\Delta_T)$; в период умственного утомления — $\delta_2(\Delta_T)$.

Измеряя значения уровня асимметрии ЭЭГ, усредняемые за последовательные отрезки времени T :

$$\Delta_{T_1}, \Delta_{T_2}, \Delta_{T_3}, \dots, \Delta_{T_k},$$

требуется вынести решение об умственном утомлении. В этом случае можно поставить диагноз утомления, когда оно не наступило, и не учесть его наступления. Ошибки эти не равнозначны. Вторая ошибка более серьезна.

Вероятность первой α и второй β указанных ошибок является характеристикой надежности диагноза, а значения α и β устанавливаются заранее.

Различение основывается на анализе отношения правдоподобия (вероятности). Отношение правдоподобия для первого отрезка времени Δ_{T_1} вычисляется по формуле

$$Z_1 = \frac{\delta_2(\Delta_{T_1})}{\delta_1(\Delta_{T_1})}.$$

Если $Z_1 < 1$, то есть при данном значении асимметрии ЭЭГ (Δ_{T_1}) более вероятно неутомленное состояние, значит, утомление не наступило.

Отношение правдоподобия Z_2 для второго отрезка времени Δ_{T_2}

$$Z_2 = \frac{\delta_2(\Delta_{T_2})}{\delta_1(\Delta_{T_2})}.$$

Если $Z_2 < 1$, то рассматривается третий отрезок времени, где аналогичным образом вычисляется Z_3 , и т. д.

Если на каком-то отрезке времени окажется, что $Z_k > 1$, то с этого момента проводится последовательный статистический анализ по Вальду с установленными значениями α и β .

При

$$Z_k = \frac{\delta_2(\Delta_{T_k})}{\delta_1(\Delta_{T_k})} > \frac{1 - \alpha}{\beta}$$

диагностируется умственное утомление. Если же

$$\frac{\alpha}{1-\beta} < Z_k < \frac{1-\alpha}{\beta},$$

то берется следующий отрезок времени $k + 1$ и вычисляется отношение правдоподобия

$$Z_{k+1}^{\text{послед}} = \frac{\delta_2(\Delta T_k) \cdot \delta_2(\Delta T_{k+1})}{\delta_1(\Delta T) \cdot \delta_1(\Delta T)}.$$

Если при каком-то S окажется, что

$$\frac{\alpha}{1-\beta} < \frac{\delta_2(\Delta T_k) \cdot \delta_2(\Delta T_{k+1}) \dots \delta_2(\Delta T_{k+s})}{\delta_1(S_{T_k}) \cdot \delta_1(\Delta T_{k+1}) \dots \delta_1(\Delta T_{k+s})} < \frac{1-\alpha}{\beta},$$

то еще нет основания для принятия решения о наличии утомления.

При первом отрезке времени m , когда

$$Z_{k+m}^{\text{послед}} > \frac{1-\alpha}{\beta},$$

ставится диагноз умственного утомления с уровнем надежности α, β . Если

$$Z_{k+m}^{\text{послед}} < \frac{\alpha}{1-\beta},$$

то повторяется процедура с Z_1 . Несмотря на большие дисперсии распределений, для диагностики утомления с вероятностью ошибок $\alpha = \beta = 0,001$ необходимо 30—60 сек.

Описанный алгоритм различения реализуется и для других параметров, имеющих разные распределения; в то же время его можно использовать для диагностики патологических состояний по многим признакам, имеющим, однако, недостаточную информативность.

4. НЕКОТОРЫЕ МЕТОДЫ КИБЕРНЕТИЧЕСКОЙ ОБРАБОТКИ ДИАГНОСТИЧЕСКИХ ДАННЫХ

Создание специальных диагностических алгоритмов, не требующих применения сложных технических средств, представляет важную и перспективную задачу, так как ЭВМ с большим штатом обслуживающего персонала доступны в основном крупным научно-исследовательским центрам. Отсюда возникает необходимость в первичном анализе и синтезе накопленных медициной и постоянно поступающих сведений. Для рационального их использования признаки болезней классифицируются на определенные группы, отыскивается мера их информативной значимости. Для этого разработаны методы, доступные каждому врачу и не требующие сложной вычислительной техники (В. С. Генес, 1964, 1967, 1968).

Приходится признать, что не все признаки заболеваний в равной мере могут быть использованы для диагностики. Часть отклонений от нормы может одинаково встречаться при разных болезнях. Другая группа признаков обнаруживается только при данном заболевании, но редко. Важность таких признаков для диагностики несомненна, но отсутствие их еще не отвергает диагноза. Третья группа встречается при одной болезни чаще, чем при другой. Поэтому использование их для диагностики предполагает количественное выражение их значимости.

Как осуществляется подобная процедура?

Прежде всего в соответствии с медицинскими руководствами составляются перечни всех признаков, характерных для определенных заболеваний, подлежащих дифференциальной диагностике. Далее делается выборка за определенный промежуток времени историй болезней. На каждого больного заполняется карточка, где указываются все показатели при поступлении или за определенный период лечения.

Карточки раскладываются по группам и отдельным формам заболеваний. В каждой группе болезней признаки сортируются по величине или качеству отклонений от нормы. Если показатель имеет количественную оценку (количество лейкоцитов, РОЭ, температура и т. д.), его можно выразить графически (рис. 3).

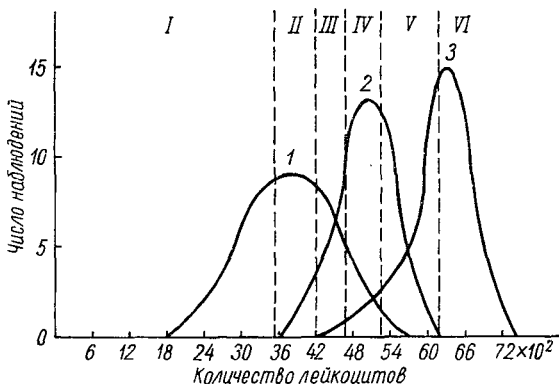


Рис. 3. График определения показателей состояния здоровья, пригодных для диагностики заболеваний: I—VI — зоны; 1 — болезнь А; 2 — болезнь В; 3 — болезнь С (В. С. Генес, 1965)

На оси абсцисс откладываются величины показателя, на оси ординат — число случаев болезни. В результате получается графическое изображение распределения случаев каждого заболевания по величине изучаемого показателя. Если показатель приемлем для различения болезней, на графике могут быть выделены зоны, характерные только для одного заболевания (зоны I, VI), зоны превышения частоты одного заболевания над частотой другого (зоны II,

IV, V). Уровень этого превышения определяется по табл. 15 и выражает надежность отнесения каждой зоны к тому или иному заболеванию.

Таблица 15

Таблица для определения уровня превышения частоты одного заболевания над частотой другого (В. С. Генес, 1968)

| Общее число наблюдений в зоне | Уровни значимости, % | | |
|-------------------------------|----------------------|----|----|
| | 75 — 90 | 95 | 99 |
| 12 | 8 | 9 | 11 |
| 13 | 9 | 10 | 12 |
| 14 | 9 | 11 | 13 |

В первом столбце табл. 15 представлено общее число наблюдений в зоне, в остальных — число наблюдений одного из заболеваний, необходимое для определения зоны, характерной для этого заболевания, с уровнем значимости 75—90, 95 и 99%.

При непригодности показателя (признака) для различения болезней характеристические зоны на графике не выделяются.

Для количественной оценки каждого признака, пригодного для дифференциальной диагностики, строится табл. 16. В ней представлены отдельные болезни (А, В, С) или их формы и градации показателя, соответствующие границам зон, выделенных на графике (см. рис. 3).

В центре зоны указано число больных (22, 18, 12, 4). Под ним помещается процент с его средней ошибкой (39 ± 7),

Количественная оценка информативной
способности показателя (В. С. Генес, 1968)

| Бо- лезнь | З о н ы | | | | | | Всего боль- ных | Приходится на характеристи- ческие зоны | |
|---------------------------------|-------------------------|------------------------|-------------------------|------------------------|------------------------|-------------------------|-----------------------|---|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 | | слу- чаев | % |
| | до 3450 | 3450 — 4040 | 4050 — 4640 | 4650 — 5240 | 5250 — 6140 | свыше 6140 | | | |
| A | 100 ± 2 22 39 ± 7 | 82 ± 8 18 32 ± 6 | 46 ± 10 12 8 ± 4 | 11 ± 5 4 | 0 | 0 | 32 ± 4 56 100% | 40 | 71 ± 6 |
| B | 0 | 18 ± 8 4 8 ± 4 | 50 ± 10 13 24 ± 6 | 74 ± 8 25 46 ± 7 | 34 ± 8 12 22 ± 6 | 0 | 31 ± 4 54 100% | 25 | 46 ± 7 |
| C | 0 | 0 | 4 ± 4 1 2 ± 2 | 15 ± 6 5 8 ± 4 | 66 ± 8 23 37 ± 6 | 100 ± 2 34 53 ± 6 | 37 ± 4 63 100% | 57 | 90 ± 4 |
| Всего боль- ных в зоне | 22 | 22 | 26 | 34 | 35 | 34 | 173 | 122 | 70 ± 3 |

который вычисляется от общей численности больных (56) с данной формой болезни А. Над числом больных в каждой зоне располагается процент (100 ± 2), который это число составляет от всех больных, попавших в данную зону (последняя строка таблицы). Этот процент соответствует вероятности диагноза данной болезни. Проценты и их вероят-

ностные средние ошибки определяются по готовой таблице. Эта вероятность вычисляется по формуле Байеса.

В последних двух столбцах таблицы представлены суммы случаев и их проценты, приходящиеся в каждом заболевании на их характеристические интервалы.

Суммарный процент различающихся случаев каждого заболевания является мерой информативной ценности показателя по всем его признакам. Эта мера представлена для всей группы в соответствующем столбце последней строки. Она действительна в пределах конкретного контингента больных, имеющегося в медицинском учреждении. Эта оценка может использоваться и в других медицинских учреждениях с аналогичным соотношением контингентов больных по заболеваниям. Если же частота заболеваний в этих учреждениях различна, то указанная оценка неприменима для других учреждений. В этом случае можно пользоваться лишь частотой признака при каждом заболевании для более сложного расчета вероятности заболевания по формуле Байеса, в которую входят частоты отдельных заболеваний.

Качественные показатели анализируются аналогично, только вместо линейного графика строится столбиковый — с заполнением столбиков на каждую градацию качества для отдельного заболевания.

Так как для основного числа заболеваний проведение дифференциальной диагностики по одному показателю невозможно, необходимо проанализировать информативный вес различных комплексов показателей.

Для дифференциальной диагностики заболеваний по ряду показателей необходимы признаки, которые были бы как можно меньше связаны друг с другом и имели бы наибольший информативный вес. С этой целью составляется кодировочная таблица, где строки представлены отдельными признаками, а столбцы — характеристическими зонами.

На пересечении столбцов и строк выделяются натуральные границы зон.

На каждый комплекс в карточки больных заносится специальный код в виде многоразрядного числа. В этом случае каждый показатель комплекса представлен отдельным разрядом этого числа. Число разряда определяет номер зоны-показателя. Если комплекс закодирован, например, числом 4232 и состоит из четырех показателей, то первый слева показатель находится в четвертой зоне, второй показатель — во второй и т. д.

Для двух показателей строится двухкоординатная таблица (табл. 17).

Таблица 17

Кодировочная таблица (В. С. Генес, 1968)

| Показатели | З о н а | | | | | |
|----------------|---------|----------------|----------------|----------------|----------------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Лейкоциты | < 3450 | 3450 — 4040 | 4050 — 4640 | 4650 — 5240 | 5250 — 6150 | > 6150 |
| Частота пульса | < 55 | 55 — 65 | 66 — 86 | > 86 | | |

В ячейках на пересечениях столбцов и горизонтальных строк таблицы выделяются колонки для отдельных заболеваний, куда заносятся номера историй болезней больных с данным комплексом показателей. В отдельной графе записываются порядковые номера больных без разделения по заболеваниям. По таблице определяется степень характеристики каждой ячейки для определенных болезней. Для каждого заболевания суммируются числа случаев в характерных для него ячейках, определяется их процент и средняя

ошибка. Эта характеристика является мерой информативной значимости комплекса показателей каждой болезни. Процент различных случаев всех заболеваний служит мерой информативной значимости комплекса в отношении данной группы заболеваний.

При малом числе наблюдений может применяться метод обобщения. Для этого суммируется количество больных с определенным заболеванием в рядом лежащих ячейках таблицы, а затем всей зоны, где суммировались ячейки. Лучшим является комплекс показателей, который дает больший процент различаемых случаев.

Если комплекс велик, то для показателей строятся сложные таблицы, а вместо многокоординатных таблиц заполняются отдельные многокоординатные карточки (табл. 18).

Таблица 18

Многокоординатная карточка (В. С. Генес, 1968)

| Болезни | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| A | | B | | C | | D | | Всего |
| № п. п. | № и. б. | № п. п. | № и. б. | № п. п. | № и. б. | № п. п. | № и. б. | № п. п. |
| 1 | 1522 | 1 | 1615 | | | | | 1 |
| 2 | 1641 | | | | | | | 2 |
| 3 | 1723 | | | | | | | 3 |
| 4 | 1800 | | | | | | | 4 |
| 5 | 1814 | | | | | | | 5 |

Вероятность $A = 84 \pm 14\%$.

Дополнительный показатель: гемокультура.

Лечение: пропись 81.

Благоприятный исход: через 3 — 4 дня состояние 11111231

Следующий этап диагностической системы — составление каталогов: алфавитных (на показатели состояния здоровья, болезни, сложные методы обследования); номерных (на комплексы показателей, лечебные мероприятия).

Карточки показателей состояния здоровья отражают методики исследования, характеристические зоны для различных заболеваний, вероятность заболевания при каждом симптоме. Данные на каждую группу дифференцируемых между собой заболеваний заносятся отдельно.

Карточки каталога на отдельные болезни включают перечень всех форм заболеваний; распределение заболеваний по месяцам; список признаков болезней для каждой формы; список показателей состояния здоровья, учитываемых при дифференциальной диагностике, в том числе и таких, которые не изменяются при данной форме заболевания; номер комплекса показателей, минимально необходимых и достаточных для диагностики.

Карточки на сложные методы исследования (ЭЭГ, ЭКГ и т. д.) должны включать методические указания о способах измерения; списки болезней, при которых данная методика имеет диагностическую ценность; отдельные показатели с характеристическими зонами для разных болезней.

Карточки номерного каталога на комплексы показателей должны содержать перечень показателей с указанием способа их измерения; кодировочные таблицы для разных групп болезней; номера многокоординатных таблиц для оценки результатов исследования.

В карточках номерного каталога на прописи лечения содержатся указания по лечению заболеваний в соответствии с многокоординатными карточками.

Работа врача в направлении описанной диагностической системы осуществляется следующим образом. После сбора анамнеза и осмотра больного врач выбирает из алфавитного каталога болезней карточку предполагаемого у больного за-

болевания. По ней определяется номер комплекса показателей, необходимых для диагностики. Уточнив по карточкам из номерного каталога, что еще нужно для врачебного обследования, он определяет эти показатели у больного и кодирует их в многоразрядное число, оценивая по соответствующей многокоординатной таблице. Если результат получается однозначный, то у больного устанавливается диагноз соответствующей болезни. В то же время по номерам историй болезней, записанным в данной ячейке, можно просмотреть клинические прецеденты с теми же проявлениями симптомов.

При нахождении в ячейке еще и других болезней необходимо проведение дополнительных исследований, наименование которых указано в ячейке. По карточке алфавитного каталога на показатели определяется способ измерения и характеристические зоны для дифференциации заболеваний, оставшихся в данной ячейке при исследовании по комплексу показателей. После дополнительного исследования результат оценивается по таблице характеристических зон. Диагноз считается установленным при однозначном результате. Если же остаются другие болезни, то продолжается дальнейшее исследование в соответствии с указаниями карточки. В случае отсутствия последних вопрос о диагнозе решается по степени вероятности каждого заболевания.

Для проверки, нет ли у больного еще каких-либо других болезней, сличают обнаруженные симптомы с перечнем симптомов этого заболевания. Если симптомы, не указанные в нем, отсутствуют, то возможность других болезней исключается. В случае обнаружения таких симптомов отыскиваются соответствующие карточки в алфавитном каталоге показателей состояния здоровья и по ним устанавливаются возможные дополнительные заболевания. Далее производится процедура, аналогичная описанной выше, до тех пор, пока будет найдено объяснение симптомов.

Из анализа рассмотренной диагностической системы видно, что отдельные ее звенья представляют не что иное, как диагностический алгоритм, не требующий применения электронных вычислительных машин. Эта система позволяет весьма экономично проводить процесс диагностики и назначать дифференцированное лечение. Кроме того, врач может получать указания о методах контроля за эффектом лечения.

После окончания лечения или верификации диагноза на операции или секции полученные данные заносятся на карточки алфавитного каталога как показатели (в характеристические зоны) или в ячейки многокоординатных таблиц, благодаря чему диагностическая система, непрерывно пополняясь и совершенствуясь, обобщает опыт врачей конкретного медицинского учреждения. Такая система весьма полезна для начинающего врача, который не получил еще достаточных сведений о разнообразии форм заболеваний, а также и для опытного клинициста, диагностические возможности которого основаны на личном опыте и интуиции.

Наличие упорядоченных и количественно выраженных правил и ограничений для осуществления диагностического процесса помогает устранить ряд диагностических ошибок при условии широкого использования алгоритмов во врачебной практике и в процессе преподавания.

Внедрение математических методов в медицину должно прийти на смену диагностическому процессу, основанному на знаниях, сохранившихся в памяти врача, и эмпирическом подходе к оценке данных при помощи элементарных логических операций. Создание диагностических матриц-таблиц в то же время способствует созданию дифференциально-диагностических программ для ЭВМ.

Литература

Балуев О. А., Иовлев Б. В., Левин Г. З., Тонконогий И. М. О решении задач дифференциальной диагностики размягчений и кровонезлияний с помощью последовательного статистического анализа. — Журнал невропатологии и психиатрии им. С. С. Корсакова, 1967, 9, 1300.

Вальд А. Последовательный анализ. М., 1960.

Ван дер Варден. Математическая статистика. М., 1960.

Генкин А. А. Выступление при обсуждении докладов. — Биологические аспекты кибернетики. М., 1962, 231.

Генкин А. А., Гублер Е. В. Применение последовательного анализа для дифференциальной диагностики и использование этого метода для различения двух форм ожоговой болезни. — Применение математических методов в биологии, сб. 3. Л., 1964, 174.

Генкин А. А., Зараковский Г. М. Некоторые пути автоматизации дифференциальной диагностики функциональных состояний организма и начальных стадий болезней. — Кибернетика в клинической медицине. Л., 1964, 73.

Генес В. С. Таблицы достоверных различий между группами наблюдений по качественным показателям. М., 1964.

Генес В. С. Некоторые простые методы кибернетической обработки данных диагностических и физиологических исследований. М., 1967.

Генес В. С. Некоторые подходы к обработке, хранению и использованию медицинской информации. — Вычислительная техника в физиологии и медицине. М., 1968, 84.

Гублер Е. В., Полонский Ю. З. Применение вероятностно-статистического анализа для оценки тяжести ожогового шока и прогноза ожоговой болезни. — Тезисы докладов 14-й научной сессии Института хирургии им. А. В. Вишневского. М., 1962.

Гублер Е. В. Перспективы применения математических методов и электронных вычислительных машин при диагностике последствий ожогов и травм. — Биологическая и медицинская электроника. М., 1963, 3, 35.

Гублер Е. В., Полонский Ю. З. Автоматизация распознавания заболеваний по минимуму диагностической информации с помощью диагностических таблиц. — Кибернетика в клинической медицине. Л., 1964, 68.

Гублер Е. В., Полонский Ю. З., Генкин А. А., Корытова М. Ю. Распознавание форм ожоговой болезни при помощи дифференциально-диагностических таблиц. — Экспериментальная хирургия и анестезиология, 1964, 5, 17.

Гублер Е. В. Опыт применения последовательной диагностической

процедуры. Некоторые перспективы развития вычислительных методов дифференциальной диагностики. — Математические методы в психиатрии и неврологии. Л., 1968, 62.

Гублер Е. В., Генкин А. А. Применение критериев непараметрической статистики для оценки различий двух групп наблюдений в медико-биологических исследованиях. М., 1969.

Иовлев Б. В., Тонконогий И. М. О применении последовательного статистического анализа для решения задач дифференциальной диагностики. — Журнал невропатологии и психиатрии им. С. С. Корсакова, 1966, 5, 708.

Иовлев Б. В., Случевский Ф. И. Последовательный статистический анализ и изучение вопросов дифференциальной диагностики в психиатрии. — Математические методы в психиатрии и неврологии. Л., 1968, 65.

Каминский Л. С. Статистическая обработка лабораторных и клинических данных. М., 1964.

Кошль О. И., Полонский Ю. З. Применение последовательного статистического анализа для дифференциальной диагностики стертых и безжелтушных форм эпидемического гепатита и обострений хронического холангиогепатита. — Кибернетика в клинической медицине. Л., 1964, 91.

Легеза В. И. Применение последовательного статистического анализа для дифференциальной диагностики трех форм ожоговой болезни. — Кибернетика в клинической медицине. Л., 1964, 96.

Мелодиев А. А. Применение метода последовательного статистического анализа для дифференциальной диагностики облитерирующего эндартериита и ангиоспазмов при некоторых нервных заболеваниях. — Математические методы в психиатрии и неврологии. Л., 1968, 71.

Мисюк Н. С., Довнар И. Н. Дифференциальная диагностика острых нарушений мозгового кровообращения при помощи таблицы-матрицы. — Электронные вычислительные машины в невропатологии. Минск, 1968, 32.

Урбах В. Ю. Биометрические методы. М., 1964.

Возможность самообучения — одно из важнейших свойств кибернетических систем. В процессе его возникает новая информация, способствующая целенаправленному поиску наиболее оптимального режима работы. Более того, самообучающаяся кибернетическая система способна самостоятельно изменять свой алгоритм для более быстрого достижения поставленной цели. Так, свойство самообучения позволяет автоматически накапливать опыт машинной обработки и оценки данных, учитывать ошибки диагностики. В случае необходимости ЭВМ вносит в свою память исправления, касающиеся порогов значений отдельных показателей, их вероятностей и др.

Программа самообучения строится так, чтобы свести к минимуму число диагностических ошибок. Процесс самообучения осуществляется на основе архивного материала и в результате ввода данных о больных при диагностике их заболеваний.

В основе алгоритма самообучения могут лежать пробные постановки диагноза с различными весами, придаваемыми каждому признаку. Естественно, что наиболее удачный набор весов будет определяться по совпадению машинного диагноза с клиническим или окончательным (верифицированным) диагнозом. Результативность процесса самообучения в значительной степени зависит от числа про-

анализированных случаев. Весьма важным является то, что его скорость у ЭВМ значительно выше, чем у человека. Отсюда вытекает необходимость быстреего накопления материала для машинной обработки с последующим самообучением ЭВМ (В. В. Парин, Р. М. Баевский, 1966).

Для наглядности приводим алгоритм самообучения, описанный М. Л. Быховским (1962, 1969).

Качество распознавания болезни D_j при помощи ЭВМ определяется отношением

$$K = \frac{N_j - \bar{N}_j}{N_{0j}},$$

где N_{0j} — число обследованных больных с заболеванием D_j ;
 N_j — число больных, которым машина поставила правильный диагноз;
 \bar{N}_j — число больных, которым машина поставила ошибочный диагноз.

Чем следует руководствоваться, чтобы на основании теории вероятностей можно было поставить диагноз или признать, что на материале имеющихся данных поставить диагноз невозможно?

В том случае, где установлено, что вероятность одного из заболеваний близка к 1, а остальных — к 0, вопрос относительно ясен. Иное дело, если вероятность одного равна 0,4, другого — 0,35, третьего — 0,2 и т. д. Или вероятность одного равна 0,4, а других — 0,05. Несомненно, что во втором случае трудно сказать, что первое заболевание является искомым диагнозом. В третьем случае, наоборот, это сделать значительно легче.

Становится очевидным, что для постановки диагноза важна не только вероятность его, но и то, что разность между вероятностью этого заболевания и следующего за ним не меньше некоторой минимальной. Эти два подхода объединяются одним критерием — порогом L .

Примем, что для каждого заболевания L_j такое, что если вероятность одного из них $P(D_j/D_s) \geq L_j$ (D_j — болезнь, а D_s — система признаков), то это заболевание имеет место в действительности. Если $L_j > 0,5$, то все остальные вероятности (даже в сумме) окажутся меньше 0,5.

Если принять такой метод выбора порогов, при котором $L > 0,5$, то он создает гарантированную разность вероятностей между первым по списку заболеванием и последующим. При этом число N_j будет практически близким к нулю.

Общее качество машинной диагностики для всех заболеваний данного класса предлагается определять критерием

$$K = \sum_j \beta_j K_j, \quad (25)$$

где β_j — весовые коэффициенты, указывающие на то, какие заболевания ЭВМ должна распознавать лучше.

В приведенном случае K служит критерием избирательного качества, где весовые коэффициенты β_j указывают характер предпочтения.

Если все β равны друг другу ($\beta_1 = \beta_2 = \dots = \beta_m = 1$), то будет иметь место критерий равномерного качества

$$K = \sum K_j. \quad (26)$$

Если значения информационных коэффициентов зафиксированы, то критерий качества является функцией выбранных порогов $K(T_1, \dots, T_m)$. Для максимизации критерия качества следует реализовать на ЭВМ процесс самообучения, состоящий из пробных постановок диагноза, вычислений по формуле (25) или (26) критерия качества и подбора таких значений T_1, \dots, T_m , при которых рассматриваемый критерий максимален. Описываемый процесс самообучения проводится на материале архива историй болезней и текущих данных клиники.

Процесс нахождения оптимальной совокупности (T_1, \dots, T_m) состоит в том, что сначала определяется оптимальное значение T_1 при фиксированных T_2, \dots, T_m , затем — T_2 при фиксированных T_1, T_3, \dots, T_m и т. д. Определение каждого частного оптимума сводится к многократным пробным прогонкам фактического материала по формуле (25) или (26) каждый раз с изменением данного T_j на ΔT_j , пока не будет установлен максимум (рис. 4).

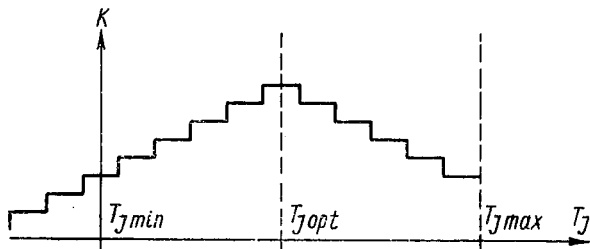


Рис. 4. Определение оптимальной совокупности

Если $L_{jmin} \geq 0,5$, то выбор каждого из T_j не зависит от значений других порогов и на этом процесс заканчивается. В противном случае после того, как будут пройдены все m шагов, необходимо вновь вернуться к величине T_1 и найти ее второе оптимальное значение. Это же повторяют с величиной T_2 и т. д. до тех пор, пока последующее оптимальное значение каждого порога не будет минимально отличаться от предыдущего.

Поиск может быть облегчен установлением границы для T_j , а именно: T_{jmax} имеет место при $L_j = 1$ или

$$T_{jmax} = \log \frac{1}{P(D_j)};$$

наряду с этим T_{jmin} имеет место при $L_j = P(D_j)$ (выбор диагноза D_j делается априорно), т. е.

$$T_{j\min} = \log \frac{P(D_j)}{P(D_j)} = 0.$$

По сути дела минимальное значение $T_{j\min}$ определяется величиной $L_{j\min} = 0,5$. При этом обеспечивается должная минимальная разность между вероятностью заболевания, принятого в качестве диагноза, и вероятностью следующего по порядку заболевания. Одновременно с этим соблюдается принятый принцип классификации нозологических форм, а именно: все D_j несовместимы и единственно возможны. Таким образом,

$$T_{j\min} = \log \frac{1}{2P(D_j)}.$$

Поиск максимума выгоднее начинать с $T_{j\min}$ или близкого к нему, постепенно его уменьшая.

Описанный процесс самообучения может охватывать не только подбор порогов, но и оптимизацию хотя бы наиболее важных информативных коэффициентов α_{ji} .

В итоге при наличии фактического материала (например, тысяча или более историй болезней) описанный выше процесс позволит произвести самообучение ЭВМ (подбор T_j и α_{ji}) так, что в пределах этого фактического материала ЭВМ будет наилучшим образом диагностировать заболевания (М. Л. Быховский, 1962).

Литература

Быховский М. Л. Вероятностная логика построения самообучающегося диагностического процесса на математических машинах. — Экспериментальная анестезиология, 1962, 1, 3.

Быховский М. Л. Теоретические основы обучения в диагностических системах. — Машинная диагностика и информационный поиск в медицине. М., 1969, 51.

Парин В. В., Баевский Р. М. Введение в медицинскую кибернетику. М.—Прага, 1966.

Министерство здравоохранения БССР
Вторая клиническая больница г. Минска
Клиника нервных болезней

ИСТОРИЯ БОЛЕЗНИ
БОЛЬНОГО С НАРУШЕНИЕМ МОЗГОВОГО КРОВООБРАЩЕНИЯ

Ф. И. О. _____

Пол м./ж. Возраст _____ лет. Семейное положение _____

Образование _____ Профессия, должность _____

Место работы _____

Адрес больного и родственников _____

Кем (чем) доставлен больной _____

На какой день после инсульта _____

Диагнозы:

1. Направившего учреждения _____

2. Предварительный _____

3. Окончательный _____

4. Сопутствующие заболевания _____

5. Осложнения _____

6. Оперативное вмешательство (наименование) _____

| | | | |
|--|---|--|---|
| Санобработка: полная частичная обмывка не было | Исход болезни: выздоровление значительное улучшение улучшение без перемен ухудшение смерть | Трудоспособность: восстановлена утрачена временно утрачена постоянно направлен(а) на амбулаторное лечение выдан б/лист № _____ на _____ дней, в т. ч. зачтено дней _____ до поступления в больницу | Поступил _____ Выбыл _____ Переведен в _____ Проведено в больнице _____ койко-дней |
| Транспортировка: на носилках на каталке может идти | (дата) _____ (час) _____ | | |

Зав. отделением _____

Лечащий врач _____

Главный врач _____

Осмотр дежурного врача

Назначения

Время _____

Дежурный врач _____

Примечание. При заполнении истории болезни наличие признака обозначать знаком (+) или подчеркивать, согласно указаниям в тексте, отсутствие — (—). Выраженность признака оценивать следующим образом: норма — (0); усиление, увеличение, повышение легкое — (+ 1); умеренное — (+ 2); резкое — (+ 3); ослабление, уменьшение, снижение легкое — (— 1); умеренное — (— 2); резкое — (— 3). В случае необходимости дается описание.

Для тяжелого больного дневник ведется по форме № 2, для остальных — по форме № 1. В последнем случае перечисляются признаки болезни и ежедневно оцениваются по указанной шкале.

| № п. п. | Признаки | I. Жалобы | II. Предвестники |
|---------|------------------------------|-----------|------------------|
| 1. | Головная боль | | |
| | а) общая | | |
| | б) локальная (слева, справа) | | |

| № п. п. | Признаки | I. Жалобы | II. Предвестники |
|------------|--|-----------|------------------|
| | в) постоянная | | |
| | г) приступообразная | | |
| | д) сильная | | |
| | е) умеренная | | |
| | ж) слабая | | |
| | з) больше беспокоит (утром, днем, вечером) | | |
| | и) усиливается в горизонтальном положении | | |
| 2. | Чувство тяжести в голове | | |
| 3. | Шум в ушах | | |
| 4. | Шум в голове | | |
| 5. | Головокружение | | |
| | а) системное | | |
| | б) несистемное | | |
| 6. | Тошнота | | |
| 7. | Рвота | | |
| 8. | Икота | | |
| 9. | Нарушение глотания | | |
| 10. | Расстройство речи | | |
| 11. | Сонливость | | |
| 12. | Преходящее расстройство сознания | | |

| № п. п. | Признаки | I. Жалобы | II. Предвестники |
|------------|---|-----------|------------------|
| 13. | Общие судорожные припадки | | |
| 14. | Припадки джексоновской эпилепсии | | |
| 15. | Расстройство зрения | | |
| 16. | Диплопия (при взгляде влево, вправо, вверх, вниз) | | |
| 17. | Рука — парез, паралич; слева, справа | | |
| 18. | Нога — парез, паралич; слева, справа | | |
| 19. | Гемипарез, гемиплегия; слева, справа | | |
| 20. | Парестезия — лицо, рука, нога; слева, справа | | |
| 21. | Гемигипестезия, гемианестезия; слева, справа | | |
| 22. | Расстройство походки | | |
| 23. | Приливы к лицу, чувство жара в нем | | |
| 24. | Боли в области сердца | | |
| 25. | Носовые кровотечения | | |
| 26. | Боль при поворотах головы и кивании | | |
| 27. | | | |
| 28. | | | |

| | | | |
|-----|--|--|--|
| 29. | | | |
| 30. | | | |

III. Нарушение мозгового кровообращения развилось

| | | |
|-----|---|--|
| 31. | Быстро | |
| 32. | Постепенно | |
| 33. | Скачкообразно | |
| 34. | Симптомы прогрессивно нарастали | |
| 35. | Внезапно, но симптомы быстро регрессировали | |
| 36. | Внезапно, но симптомы регрессировали медленно | |
| 37. | | |
| 38. | | |
| 39. | | |

IV. Нарушение мозгового кровообращения наступило в _____ 19 ____ г. (час и дата)

| | | |
|-----|------------------------------------|--|
| 40. | Во время физического напряжения | |
| 41. | Во время эмоционального напряжения | |
| 42. | При перегревании организма | |
| 43. | В связи с приемом алкоголя | |
| 44. | Во время или сразу после сна | |
| 45. | На фоне инфекции | |
| 46. | В связи с травмой головы | |
| 47. | | |
| 48. | | |
| 49. | | |

V. Динамика нарушения мозгового кровообращения

| | |
|-----|-----------------------------|
| 50. | Первые признаки |
| | |
| | |
| 51. | Признаки, возникшие позднее |
| | |
| | |
| | |

VI. Заболевания, отмеченные в анамнезе

| № п. п. | Наименование болезни | Заболел (дата) | Выздоровел (дата) |
|---------|-----------------------------------|----------------|-------------------|
| 52. | Гипертоническая болезнь | | |
| 53. | Артериальная гипертония | | |
| 54. | Артериальная гипотония | | |
| 55. | Атеросклероз | | |
| 56. | Диабет | | |
| 57. | Травмы головы | | |
| 58. | Гриппозная инфекция | | |
| 59. | Облитерирующий эндартериит | | |
| 60. | Ревматизм | | |
| 61. | Аневризмы сосудов головного мозга | | |

| № п. п. | Наименование болезни | Заболел (дата) | Выздоровел (дата) |
|------------|---|-------------------|----------------------|
| 62. | Эндокардит септический | | |
| 63. | Мерцательная аритмия | | |
| 64. | Инфаркт миокарда | | |
| 65. | Болезни крови | | |
| 66. | Хронический алкоголизм | | |
| 67. | Расстройства мозгового кровообращения (когда, какие, сколько) | | |
| | | | |
| | | | |
| 68. | | | |
| 69. | | | |
| 70. | | | |
| 71. | | | |

VII. Осложнения прежних заболеваний

| | |
|-----|---------------------------------|
| 72. | Каких болезней, какие (описать) |
| | |
| | |
| | |

VIII. Заболевания настоящие

| № п. п. | Наименование болезни | Дата возникновения заболевания |
|------------|-----------------------------------|-----------------------------------|
| 73. | Гипертоническая болезнь | |
| 74. | Артериальная гипертония | |
| 75. | Артериальная гипотония | |
| 76. | Атеросклероз | |
| 77. | Диабет | |
| 78. | Травмы головы | |
| 79. | Гриппозная инфекция | |
| 80. | Облитерирующий эндартериит | |
| 81. | Ревматизм | |
| 82. | Аневризмы сосудов головного мозга | |
| 83. | Эндокардит септический | |
| 84. | Мерцательная аритмия | |
| 85. | Инфаркт миокарда | |
| 86. | Болезни крови | |
| 87. | Хронический алкоголизм | |
| 88. | | |
| 89. | | |
| 90. | | |

IX. Данные объективного обследования

| № п. п. | Признаки |
|------------|---|
| 91. | Положение больного: активное, пассивное, вынужденное |
| 92. | С о з н а н и е: ясное, помраченное, оглушенное, сопор, кома, психомоторное возбуждение; нарушено (кратковременно, до 12 часов, до 1 суток, 2 суток, 3 суток, свыше 3 суток) |
| 93. | К о ж н ы е п о к р о в ы: бледные, гиперемированные, багрово-синюшные, цианотичные, покрыты потом, «гусиная кожа» и др. (описать) |
| 94. | Л и м ф а т и ч е с к и е ж е л е з ы: не изменены, увеличены (какие? описать) _____ |
| 95. | О р г а н ы д ы х а н и я: границы легких, данные перкуссии, аускультации и др. (описать) а) _____ |
| | б) число дыханий в 1 минуту |
| | в) ритм дыхания: ритмичное, аритмичное |
| | г) характеристика дыхания: нормальное, учащенное, замедленное, поверхностное, глубокое, стертородное, дыхание Чейн — Стокса, дыхание Биота, дыхание Куссмауля и др. (описать) |
| 96. | С е р д е ч н о - с о с у д и с т а я с и с т е м а: а) данные осмотра, пальпации, перкуссии (описать) _____ |

б) аускультативно: норма, приглушение тонов, систолический шум на верхушке, диастолический шум на верхушке, акцент второго тона на аорте, акцент второго тона на легочной артерии, мерцательная аритмия, систолический шум на магистральных сосудах, другие изменения (описать) _____

в) пульсация на артериях: общая сонная, внутренняя сонная, поверхностная височная, бедренная, артерии тыла стопы и др. _____ слева, справа, есть, нет _____

г) пульс: ритмичный, аритмичный, напряженный, удовлетворительного напряжения, слабого напряжения, хорошего наполнения, удовлетворительного наполнения, слабого наполнения, нитевидный, не пальпируется _____

Число ударов в 1 минуту _____, дефицит пульса _____
ударов

д) АДД _____; АДС _____; в височных артериях справа _____; слева _____ (в мм рт. ст.).

97

Органы пищеварения: (описать) _____

| | |
|----|---|
| 98 | Мочеполовая система: (описать) _____ |
| | _____ |
| | _____ |
| 99 | Функция тазовых органов: стул — в норме, задержка, непроизвольный; мочеиспускание — в норме, задержка, недержание, ишурия парадокса и др. |

X. Нервная система

| | |
|-----|---|
| 100 | Нарушение высших корковых функций: снижение памяти, снижение интеллекта, расстройство эмоциональной сферы, нарушение схемы тела, аграфия, акалькулия, апраксия, афазия (моторная, сенсорная, амнестическая), дезориентировка во времени, дезориентировка в пространстве и др. (описать) _____ |
| | _____ |
| | _____ |

| 101 | Черепно-мозговые нервы: | Слева | Справа |
|-----|---|-------|--------|
| | I пара (расстройство обоняния) | | |
| | II пара (нарушение остроты и полей зрения) | | |
| | III пара (птоз, мидриаз, расходящееся косоглазие) | | |
| | IV пара (диплопия при взгляде вниз) | | |
| | V пара (порции: двигательная, чувствительная; | | |
| | ветви: первая, вторая, третья) | | |
| | VI пара (сходящееся косоглазие, диплопия) | | |

| | | | |
|-----|---|-------|--------|
| | VII пара (парез, паралич мимической мускулатуры; периферический, центральный) | | |
| | VIII пара (нарушения слуховые, вестибулярные) | | |
| | IX—X пара (парез, паралич мягкого нёба, дисфагия, гнусавый оттенок голоса и др.) | | |
| | XI пара (парез, паралич мышц; кивательной, трапецевидной) | | |
| | XII пара (атрофия языка, глоссоплегия, дизартрия, девиация языка) | | |
| | Чем проявляется поражение черепно-мозговых нервов? Синдромы: бульбарный, псевдобульбарный и др. _____ | | |
| | _____ | | |
| | _____ | | |
| 102 | З р а ч к и: | Слева | Справа |
| | миоз | | |
| | мидриаз | | |
| | анизокория | | |
| | реакция на свет: прямая, содружественная, на конвергенцию; живая, вялая, отсутствует | | |
| | синдром Горнера | | |
| | синдром Аргаиль — Робертсона | | |
| | _____ | | |

| | | | |
|-----|---|-------|--------|
| 103 | Острота зрения: ОД — ; ОС — | | |
| 104 | Поля зрения: | Слева | Справа |
| | а) концентрическое сужение | | |
| | б) скотома: положительная, отрицательная; | | |
| | в) гемианопсии: верхняя, нижняя, бина- зальная, битемпоральная, квадрантная | | |
| | | | |
| | | | |
| 105 | Глазное дно: | | |
| | 1) сужение артерий | | |
| | 2) извитость артерий | | |
| | 3) расширение вен | | |
| | 4) симптом «медной проволоки» | | |
| | 5) геморрагии | | |
| | 6) застойный сосок зрительного нерва | | |
| | 7) отек соска зрительного нерва | | |
| | 8) нейроретинит | | |
| | 9) атрофия соска зрительного нерва: первичная, вторичная | | |
| | 10) ретинальное давление | | |

| Нормальные | Слева | Справа | Патологические | Слева | Справа |
|--------------------|-------|--------|-----------------------|-------|--------|
| Корнеальный | | | Орального автоматизма | | |
| Конъюнктивальный | | | Маринеско — Радовичи | | |
| Нижнечелюстной | | | Хватательные | | |
| Мягкого нёба | | | Кистевой Бехтерева | | |
| Глоточный | | | Бабинского | | |
| Назопальпебральный | | | Клонус коленной чашки | | |
| Карпорадиальный | | | Клонус стопы | | |
| Двуглавой мышцы | | | Гордона | | |
| Брюшной верхний | | | Шеффера | | |
| Брюшной средний | | | Россолимо | | |
| Брюшной нижний | | | Мендель — Бехтерева | | |
| Коленный | | | Жуковского | | |
| Ахиллов | | | Защитные на руках | | |
| Кремастерный | | | Защитные на ногах | | |
| Анальный | | | | | |

107 Мышечный тонус: нормальный, повышен, понижен; спастичность, экстрапирамидная ригидность, горметония; слева, справа

108 Двигательные расстройства

а) гиперкинезы: дрожание (старческое, паркинсоническое, интенционное), гемибаллизм и др. (описать)

б) парез, паралич: руки, ноги, гемипарез, гемиплегия; справа, слева

в) нарушение координации движений: мозжечковое, от расстройства чувствительности, вестибулярное и др. _____

г) походка: нормальная, гемипаретическая, атактическая (от расстройства чувствительности, мозжечковая, вестибулярная), замедленная, мелкими шажками, астазия-абазия, не ходит и др. _____

109 Эпилептические проявления (описать) _____

110 Расстройство чувствительности: а) характер расстройства — болевой, температурной, тактильной, глубокой, вибрационной;

б) степень расстройства: анестезия, гипестезия, гиперестезия, гиперпатия;

в) локализация — половина лица, половина туловища, рука, нога; слева, справа. Другие виды расстройства чувствительности (описать) _____

111 Оболочечные симптомы (перечислить) _____

Течение болезни

| № п. п. | Патологические признаки | Даты осмотров | | | | | | | |
|---------------|-------------------------|---------------|--|--|--|--|--|--|--|
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| Подпись врача | | | | | | | | | |

| Дата | Дневник | Назначения |
|------|---------|------------|
| | | |

СОДЕРЖАНИЕ

| | |
|---|----|
| Предисловие | 3 |
| I КРАТКИЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ ГЛАВА И ТЕОРИИ СЛУЧАЙНЫХ ФУНКЦИЙ | |
| 1. Случайные события. Вероятность | 6 |
| 2. Условная вероятность | 8 |
| 3. Зависимость между случайными событиями | 9 |
| 4. Вероятность произведения двух случайных событий | 9 |
| 5. Случайные переменные. Плотность распределения вероятностей | 10 |
| 6. Математическое ожидание и дисперсия | 11 |
| II ОБЩИЕ ВОПРОСЫ АВТОМАТИЗАЦИИ ДИАГНОСТИЧЕСКОГО ГЛАВА ПРОЦЕССА | |
| 1. Этапы диагностического процесса | 14 |
| 2. Первичная медицинская документация | 15 |
| 3. Формализованный первичный документ | 16 |
| 4. Хранение первичных документов | 19 |
| 5. Систематизация информации | 19 |
| 6. Задачи математической диагностики | 20 |
| 7. Применение математики для диагностики | 21 |
| Л и т е р а т у р а | 22 |
| III ГЛАВА ДИАГНОСТИЧЕСКИЕ АЛГОРИТМЫ ДЛЯ ЭВМ | |
| 1. Что такое диагностический алгоритм? | 23 |
| 2. Классификация | 23 |

| | |
|---|----|
| 3. Алгоритмы | 24 |
| 1) Вероятностные | 24 |
| а) Распознавание болезней с помощью формулы Байеса | 24 |
| б) О вероятностной зависимости признаков | 27 |
| в) Логическая схема диагностического процесса | 30 |
| г) Информационная мера системы признаков | 34 |
| д) Матричный алгоритм с оценкой веса симптомов | 35 |
| 2) Метод фазового интервала | 41 |
| а) Представление о методе | 41 |
| б) Логика диагностического процесса, основанного на фазовом интервале | 45 |
| 3) Метод логического базиса. Постановка диагноза методом исключения из списка | 50 |
| 4) Статистический метод. Поиск клинического прецедента | 52 |
| Л и т е р а т у р а | 55 |

IV ДИАГНОСТИЧЕСКИЕ АЛГОРИТМЫ, НЕ ТРЕБУЮЩИЕ ГЛАВА ПРИМЕНЕНИЯ ЭВМ

| | |
|---|----|
| 1. Табличные алгоритмы | 57 |
| 2. Диагностические алгоритмы, основанные на последовательном статистическом анализе | 64 |
| 1) Распределение симптомов | 65 |
| 2) Оценка различия распределений симптомов | 67 |
| 3) Диапазоны распределений симптомов | 71 |
| 4) Определение числа правдоподобия | 73 |
| 5) Вычисление диагностических коэффициентов | 74 |
| 6) Информативность симптомов | 74 |
| 7) Проверка диагностической таблицы | 79 |
| 3. Алгоритм различения | 80 |
| 4. Некоторые методы кибернетической обработки диагностических данных | 83 |
| Л и т е р а т у р а | 93 |

V О САМООБУЧЕНИИ ДИАГНОСТИЧЕСКИХ СИСТЕМ ГЛАВА

| | |
|---------------------|---------|
| Л и т е р а т у р а | 99 |
| Приложение 1 | 100 |
| Приложение 2 | вкладка |

*Мисюк Николай Семенович,
Гурленя Арнольд Михайлович,
Лозовик Вячеслав Владимирович*

ДИАГНОСТИЧЕСКИЕ АЛГОРИТМЫ

Редактор *А. Шалковская*
Обложка художника *Г. Кричевского*
Художественный редактор *Л. Медведева*
Технический редактор *М. Кислякова*
Корректор *Е. Польская*

АТ 23783. Сдано в набор 1/XII 1969 г. Подписано к печати 23/VII 1970 г. Бумага 70×108¹/₃₂. типогр. № 2. Печ. л. 3,75 (5,25)+1 вкл. Уч.-изд. л. 5, 82. Изд. № 69—68. Тип. зак. 2087. Тираж 1500 экз. Цена 21 коп.

Издательство «Высшая школа» Государственного комитета Совета Министров БССР по печати. Редакция литературы по естественным наукам. Минск, ул. Кирова, 24. Полиграфический комбинат им. Я. Коласа Государственного комитета Совета Министров БССР по печати. Минск, ул. Красная, 23.

21 в.

51
8448

75665-

1871