УДК 538

АНАЛИЗ АЛЬТЕРНАТИВ И ВЫБОР ДИАГНОСТИЧЕСКИХ ГИПОТЕЗ. ЧАСТЬ I. ПРАВИЛО БАЙЕСА И МЕТОДЫ СТАТИСТИЧЕСКОЙ ТЕОРИИ ПРИНЯТИЯ РЕШЕНИЙ

Бескровный И.М.

Российский национальный исследовательский медицинский университет, e-mail: beskrovny2000@mail.ru

В настоящее время известен широкий спектр методов и моделей эффективного выбора альтернатив, объединяемых в рамках дисциплины, именуемой «Теория принятия решений» [2, 5]. Большинство описанных моделей методологически обоснованы и значительная их часть проверена на практике. Тем не менее, ряд базовых моделей в их классической постановке имеют определенные ограничения, сужающие возможную область их применения и снижающих достижимую эффективность их использования. В частности речь идет о методе Байеса. Цель настоящей работы – предложить подход, позволяющий снять указанные ограничения путем модификации известных моделей, придающей им определенную универсальность. При этом повышается возможность их использования при анализе диагностических гипотез и выборе альтернативных стратегий лечения.

Ключевые слова: альтернативы, диагностические гипотезы, правило Байеса

THE ANALYSIS OF ALTERNATIVES AND CHOICE OF DIAGNOSTIC HYPOTHESES. PART I. RULE BAYS AND METHODS OF THE STATISTICAL THEORY OF DECISION-MAKING

Beskrovnyj I.M.

Russian national research medical university, e-mail: beskrovny2000@mail.ru

Now the wide spectrum of methods and models of an effective choice of the alternatives united within the limits of discipline, called «the decision-making Theory» [2, 5] is known. The majority of the described models are methodologically proved also their considerable part it is checked up in practice. Nevertheless, in their classical statement the certain restrictions narrowing possible area of their application and reducing achievable efficiency of their use have a number of base models. In particular it is a question of method Bays. The purpose of the present work - to offer the approach, allowing to remove the specified restrictions by the updating of known models giving to them certain universality. Possibility of their use thus raises at the analysis of diagnostic hypotheses and a choice of alternative strategy of treatment.

Keywords: alternatives, diagnostic hypotheses, rule Bays

Основным современным статистическим механизмом доказательной медицины является байесовский метод, который широко используется в различных экспертных системах [1, 4]. При этом во многих случаях разработчики этих систем, а тем более – пользователи не задаются вопросом о системных основах этого метода и допустимости его применения в отдельных ситуациях. Между тем, традиционное использование симптомов для вероятностной диагностики (метод Байеса), пригодно для статистики, определяющей на множестве больных точность диагноза, но никак не для определения конкретной болезни у конкретного индивидуума. Существенным ограничением применимости этого метода в дифференциальной медицинской диагностике является тот факт, что расчётные соотношения, используемые в нём, справедливы лишь при независимости каждого из симптомов из наблюдаемой совокупности [1].

Строго говоря, уже одно это обстоятельство делает Байесовский метод неприменимым для решения задач выделения наиболее вероятного заболевания из совокупности других, имеющих взаимопересекающиеся совокупности симптомов. Кроме того, неудовлетворительно проработаны системные обоснования определения априорной вероятности. Поэтому можно утверждать, что методологическая неопределённость в этом вопросе сводит на нет мнимую достоверность получаемых этим методом результатов.

Цель настоящей работы – дать обоснование приведенным выше утверждениям и предложить альтернативные подходы к решению задач дифференциальной диагностики. Предлагаемые модели базируются на сочетании Байесовского метода, в тех пределах, где его применение не противоречит законам теории вероятностей, и методов статистической теории принятия решений.

Начнём с краткого изложения Байесовского подхода (БП) и по ходу изложения укажем на его уязвимые места. (Уязвимые только с точки зрения его применимости для задач выбора наиболее вероятных альтернатив в медицинской диагностике).

Базовым понятием БП является условная вероятность. Условная вероятность со-

бытия d при наличии признака s – это вероятность того, что событие d наступит при условии, что наступило событие s. То есть, вероятность того, что пациент действительно страдает заболеванием d, если у него обнаружен *только* симптом s. *В* традиционной теории вероятностей для вычисления условной вероятности события d при наступлении события s используется следующая формула:

$$P(d|s) = (d \wedge s)/P(s). \quad (1)$$

Такая же формула используется для вычисления условной вероятности события s при наличии события d

$$P(s|d) = (d \wedge s)/P(d). \quad (2)$$

Как видно из этих соотношений, условная вероятность определяется в терминах совместимости событий. Она представляет собой отношение вероятности совпадения событий d *и* s *к* вероятности появления события s (или d). Из формулы (2) следует, что

$$P(d \wedge s) = P(d|s)P(d) = P(s|d)P(s), \quad (3)$$

откуда следует $P(s|d)P(s) = P(d|s)P(d)$.

Разделив обе части на $P(s)$, получим *правило Байеса в* простейшем виде:

$$P(d|s) = (s|d)P(d)/P(s). \quad (4)$$

Это правило, которое иногда называют *инверсной формулой для условной вероятности,* позволяет определить вероятность $P(d|s)$ появления события d при условии, что произошло событие s через известную условную вероятность $P(s|d)$. *В* полученном выражении $P(d)$ — априорная вероятность наступления события d, а $P(d|s)$ — апостериорная вероятность, т.е. вероятность того, что событие d имело место, если известно, что событие s свершилось. Для систем, основанных на знаниях, формула эта гораздо удобнее, чем исходная формула вычисления условной вероятности.

Однако, сложность получения нужной информации явилась причиной негативного отношения многих специалистов по искусственному интеллекту к вероятностному подходу вообще. Это негативное отношение подкреплялось тем, что в большинстве классических работ по теории вероятностей понятие вероятности базируется лишь на понятии *объективная частотность* (частота появления при достаточно продолжительных независимых испытаниях).

Существует мнение, что эти базовые предположения небесспорны с точки зрения практических приложений. Сторонники такого подхода придерживаются *субъективистской* точки зрения на определение вероятности, которая позволяет иметь дело с *оценками* вероятностей совместного появления событий, а не с действительной «объективно» измеренной частотой. Такой взгляд на вещи связывает вероятность смеси событий с субъективной верой в то, что событие действительно наступит.

Например, врач может не иметь статистических данных о том, какая часть пациентов, жалующихся на боль в груди, страдает инфарктом миокарда. Но, на основании собственного опыта он может оценить, у какой части его пациентов, страдающих этим заболеванием, встречался такой симптом. Следовательно, он может оценить значение вероятности $P\{$*боль в груди* | *инфаркт миокарда*$\}$. Субъективный взгляд на природу вероятности тесно связан с правилом Байеса по следующей причине. Предположим, мы располагаем достаточно достоверной оценкой вероятности $P\{s|d\}$, где s означает симптом, а d – заболевание. Тогда по формуле Байеса можно вычислить вероятность $P\{d|s\}$. Оценку вероятности $P(d)$ можно взять из публикуемой медицинской статистики, а оценить значение $P(s \mid d)$ врач может на основании собственных наблюдений. Кроме того, как будет показано в дальнейшем, достаточно обоснованная оценка может быть получена при использовании методов статистической теории принятия решений.

На практике, вычисление $P(d|s)$ не вызывает затруднений, когда речь идет о единственном симптоме, т.е. когда имеется множество заболеваний D *и* множество симптомов S. При этом, для каждого члена d_i из множества D нужно вычислить условную вероятность того, что у пациентов, страдающих этим заболеванием, наблюдался *только один определенный* симптом из множества S. Тогда для вычисления условной вероятности одного из заболеваний используется формула Байеса в следующем виде

$$P(d_i \mid s) = \frac{P(d_i)P(s \mid d_i)}{\sum_{i=1}^{n} P(d_i)P(s|d_i)}. \quad (5)$$

Однако, если в множестве D имеется m членов, а в множестве S – n членов, то потребуется вычислить $mn + m + n$ оценок вероятностей. Это отнюдь не простая работа. Но ситуация значительно усложняется, если мы попробуем включить в процесс составления диагноза не один симптом, а несколько. В этом случае правило Байеса необходимо использовать в более общей форме, в которой оно имеет вид

$$P\{d|s_1 \wedge \ldots \wedge s_k\} = P(s_1 \wedge \ldots \wedge s_k|d)P(d)/P(s_1 \wedge \ldots \wedge s_k). \quad (6)$$

Для применения формулы в таком виде потребуется вычислить $(mn)^k + m + n^k$ оценок вероятностей. Это очень много, даже при небольшом значении D. Эти оценки вероятностей требуются нам по той причине, что в общем случае для вычисления $P(s_1 \wedge \ldots \wedge s_k)$ нужно предварительно вычислить произведения вида

$$P(s_1|s_2 \wedge \ldots \wedge s_k)P(s_2|s_3 \wedge \ldots \wedge s_k)\ldots P(s_k). \quad (7)$$

На практике ситуацию можно несколько упростить, если предположить, что некоторые симптомы независимы друг от друга. Тогда объем вычислений существенно снижается. Независимость любой пары симптомов s_i, и s_j означает, что

$$P(s_i) = P(s_i|s_j), \quad (8)$$

из чего следует соотношение

$$P(s_i \wedge s_j) = P(s_i)P(s_j). \quad (9)$$

Если *все* симптомы независимы, то объем вычислений будет таким же, как и в случае учета при диагнозе единственного симптома.

Конечно, предположение о полной независимости всех наблюдаемых симптомов является совершенно необоснованным. Исходя из опыта медицинской практики, очень трудно найти пример, когда какое-нибудь заболевание сопровождалось бы наличием единственного симптома (разве, что кровотечение из пулевого отверстия или ножевого ранения). На самом деле, практически любое заболевание сопровождается группой симптомов, которые никак нельзя считать независимыми. А если добавить реальную возможность наблюдения еще нескольких сиптомов, характерных для другого (а то и нескольких) сопутствующих заболеваний, то картина усложняется до чрезвычайности.

Но, на практике разработчиков многочисленных экспертных систем эти сложности мало смущают, ибо, как известно – «… Нормальные герои всегда идут в обход…» (К/ф «Доктор Айболит»). Поэтому, в большинстве случаев, используется предположение о наличии *условной независимости.* Это означает, что пара симптомов s_i *и* s_j считается независимой, если имеются какие-либо дополнительные свидетельства на этот счет или фундаментальные знания E. Тогда можно считать справедливым соотношение

$$P(s_i|s_j,E) = P(s_i|E). \quad (10)$$

и использовать его, невзирая на явную необоснованность и, не заботясь о наличии дополнительных свидетельств или, тем более, фундаментальных знаний.

Между тем, столь грубое пренебрежение не является столь уж необходимым. Значительного упрощения процедуры можно достигнуть, используя другие достаточно эффективные подходы, не требующие введения столь жёстких ограничений, как при формальном использовании правила Байеса. В частности весьма перспективным представляется использование для целей дифференциальной диагностики таблиц решений, являющихся основным инструментом статистической теории принятия решений [2, 5].

Теория статистических решений представляет в общем случае лицо, принимающее решение (ЛПР), как лицо, производящее выбор из совокупности альтернатив $A = \{a_1, a_2, \ldots, a_m\}$ при условии, что заданы:

1. Набор состояний окружающей среды $S = \{s_1, s_2, \ldots, s_n\}$, факты наступления которых не поддаются управлению, но вероятности их наступлений $P = \{p_1, p_2, \ldots, p_n\}$ известны (то есть состояние s_1 будет иметь место с вероятностью p_1 и т.д.).

2. Задана матрица платежей в терминах затрат или выигрышей V_{ij}, ассоциированных с каждой парой «альтернатива – состояние окружения».

Таким образом, теория статистических решений является теорией проведения статистических наблюдений, обработки этих наблюдений и их использования. Результат решения задачи ЛПР определяет по одному из критериев принятия решения. Для того чтобы прийти к однозначному и по возможности наиболее выгодному варианту решению, необходимо ввести оценочную (целевую) функцию. При этом по каждой стратегии ЛПР (V_i) вычисляется некоторый результат $E(V_i)$, характеризующий все последствия этого решения. Из массива i результатов принятия решений ЛПР выбирает элемент $E(V_{ij})$, который наилучшим образом отражает мотивацию его поведения.

В тех случаях, когда ЛПР известны вероятности состояний окружающей среды, применяется критерий максимального математического ожидания выигрыша. Платёжная матрица дополняется столбцом, каждый элемент которого представляет собой значение математического ожидания выигрыша при выборе соответствующей стратегии ЛПР.

Если все перечисленные выше элементы матрицы платежей заданы, то предпочтительная альтернатива может быть определена вычислением математического ожидания ценности $E(V_i)$ каждой из рассматриваемых альтернатив по формуле:

$$E(V_i) = \sum_{i=1}^{n} p_i V_{ij}. \quad (11)$$

и последующим выбором альтернативы, имеющей наибольшую ожидаемую ценность

$$E(V^*) = \max_i E(V_i). \qquad (12)$$

Следовательно, теория статистических решений сводит проблему анализа альтернатив к анализу содержащейся в них *мотивировки и* предлагает считать наиболее предпочтительной ту альтернативу, для которой математическое ожидание мотивировки имеет наибольшее значение.

Таблица 1

Матрица решений статистической игры

Окружающая среда		$s_1\ s_2\ \dots\ s_n$	Ожидаемая ценность альтернативы
Вероятности наступления		$p_1 p_2 \dots p_n$	
Альтернативы	Альтернатива a_1 Альтернатива a_1 … Альтернатива a_1	$v_{11}\ v_{12} \dots v_{1n}$ $v_{21}\ v_{22} \dots v_{2n}$ … … … … $v_{m1}\ v_{m2} \dots v_{mn}$	$E(V_i) = \sum_{i=1}^{n} p_i V_{ij}$

Однако, в общем случае, многие ситуации выбора не могут быть удовлетворительно описаны моделью анализа альтернатив в классической постановке вследствие множественности критериев, которые подлежат учету в этой модели. Поэтому достаточно широкое распространение получили модифицированные (или расширенные) модели анализа альтернатив.

Расширенные модели анализа альтернатив были получены на базе классической модели путем внесения следующих изменений:

1. Заменой n состояний окружающей среды набором целевых показателей $G = \{G_1, G_2 \dots, G_n\}$, которые ЛПР желает достигнуть;

2. Заменой платежной матрицы матрицей вероятностей p_{ij}, характеризующих шансы на достижение каждого из целевых состояний при реализации каждой из исследуемых альтернатив;

3. Заменой вектора вероятностей реализации состояния окружающей среды на вектор, элементами которого является относительная оценка или кардинальный ранг каждого из назначаемых целевых показателей $W = \{w_1, w_2, \dots, w_n\}$.

Таким образом, в модифицированной модели ЛПР освобождается от необходимости устанавливать количественную меру мотивировки для каждой пары «альтернатива – состояние окружения», задавая платежную матрицу. Вместо этого он оценивает мотивировку, содержащуюся в каждой из альтернатив, на основе ожидаемого вклада этой альтернативы в достижение целей более высокого порядка, мотивировка которых задана извне в виде вектора W. При решении задач медицинской диагностики вкладом альтернативного заболевания «в достижение некоторой цели» можно вероятность появления сопутствующего симптома. Естественно, что в ряде случаев это существенно упрощает проблему для ЛПР, освобождая его от ответственности за назначение оценок мотивировки и сводя задачу просто к логическому анализу.

Структура модифицированной модели представлена табл. 2. В соответствии с этой моделью наиболее предпочтительной альтернативой является та, которая доставляет максимум полезности (или мотивировки):

$$E(V^*) = \max_i \sum_{j=1}^{n} p_{ijj} w\ . \qquad (13)$$

По существу, в расширенной модели используется та же самая платежная матрица, как и классической постановке. Действительно, в обоих случаях вклад каждого ij-го элемента матрицы в ожидаемую ценность i-й альтернативы определяется произведением величины выигрыша (проигрыша) на его вероятность.

Разница состоит в том, что в классической постановке вероятность наступления j-го состояния среды одинакова для всех альтернатив, а величина ij-го выигрыша (проигрыша) специфична для каждой i-й строки. В модифицированной модели наоборот – выигрыш (проигрыш) по каждому j-му целевому показателю является постоянным, а вероятность получения этого выигрыша (проигрыша) специфична для каждой i-й альтернативы.

В представленном виде табл. 2 решений становится удобней в использовании и получила достаточно широкое распространение. Например, таблицу можно использовать для выбора наиболее вероятного заболевания из группы болезней, имеющих сходные совокупности симптомов [3].

Для этого:

1. В качестве альтернатив a_i рассматриваются те заболевания, при которых характерными являются наблюдаемые симптомы (гипотезы d_i);

2. В качестве целевых показателей G_j принимаются сами наблюдаемые симптомы s_j;

3. В качестве вероятностей p_{ij} в данном применении проставляются условные вероятности $P\{s_i|d_j\}$, то есть, вероятности того, что при i-м заболевании наблюдается j-й симптом.

В такой постановке ЛПР исходит из того, что гипотеза d_i (при решении задачи диагностики заболеваний альтернативы называют гипотезами), обладающая наибольшей ожидаемой ценностью, является наиболее вероятной. Сформированная таким образом модель представлена в табл. 3. Обратим внимание, что вероятности $P\{s_i|d_j\}$, которые проставляются в ячейки платёжной матрицы, не обязательно разыскивать в каких-то статистических справочниках. Они определяются на основе накопленного опыта наблюдения за больными с различными заболеваниями, то есть, содержат достаточно достоверную информацию. Естественно, эта информация в значительной мере субъективна, поскольку точность оценки зависит от квалификации врача.

Таблица 2

Модифицированная матрица решений

Оценки целевых показателей (мотивировки)		w_1	w_2	…	w_n	Ценность альтернативы $E(V_i)$
Целевые показатели		G_1	G_2	…	G_n	
Альтернативы	Альтернатива a_1	p_{11}	p_{11}	…	p_{1n}	$E(V_i)=\sum_{j=1}^{n} p_{ij} w_j$
	Альтернатива a_2	p_{11}	p_{11}	…	p_{2n}	
	Альтернатива a_1	…	…	…	…	
	Альтернатива a_m	p_{11}	p_{11}	…	p_{mn}	

Таблица 3

Матрица выбора диагностических гипотез

Оценки степени выраженности симптомов (мотивировки к выбору)		w_1	w_2	…	w_n	Ранги достоверности гипотезы d_i
Набор симптомов		S_1	S_2	…	S_n	
Альтернативы	Гипотеза d_1	$P\{s_1\|d_1\}$	$P\{s_2\|d_1\}$	…	$P\{s_n\|d_1\}$	$Q(d_i)=\sum_{j=1}^{n} p_{ij} w_j$
	Гипотеза d_2	$P\{s_1\|d_2\}$	$P\{s_2\|d_2\}$	…	$P\{s_n\|d_2\}$	
	Гипотеза d_i	…	…	…	…	
	Гипотеза d_m	$P\{s_1\|d_m\}$	$P\{s_2\|d_m\}$	…	$P\{s_n\|d_m\}$	

Это может сыграть негативную роль при недостаточной квалификации врача. Зато, квалифицированный врач способен выявить не только наличие определённых симптомов, но и тонкие нюансы степени их выраженности, такие характер кашля (глухой, звонкий, надрывный), степень твёрдости живота, локализация болевых ощущений и т.п.

Квалифицированный врач может также учесть и такие важные показатель как возраст, общее состояние здоровья, перенесенные ранее заболевания, наследственность и т. п. Ясно, что столь тонкие различия невозможно формализовать в виде каких-то статистических данных. Кроме того, в особо сложных случаях эти оценки могут приниматься консилиумом, что ещё больше повысит их надёжность по сравнению с оценками, полученными из статистических справочников.

Полученные на основе табл. 3 оценки рангов могут использоваться двояко. Во-первых, они могут послужить основой для выбора наиболее вероятной гипотезы d^*. Этот выбор осуществляется исходя из соотношения

$$Q(d^*) = \max_i \sum_{j=1}^{n} p_{ijj} w \ . \qquad (14)$$

В тех случаях, когда значение $Q(d^*)$ заметно превышает значения остальных оценок вычислительные процедуры можно завершить и переходить при необходимости к проведению дополнительных анализов по профилю выявленного заболевания, или непосредственно к лечебным процедурам. Но, бывает, что разница между вычисленными ранговыми оценками невелика. Так в одной из работ (к сожалению, ссылка на источник утрачена) приводилась таблица вероятностей для группы заболеваний по признаку «Острый живот», полученная авторами при использовании разработанной ими программы:

Холецистит	0,787
Внематочная беременность	0,655
Панкреатит	0,321
Аппендицит	0,218

и на основании этой таблицы авторами был сделан «уверенный» вывод о выявленном заболевании холециститом. Видимо, авторы оказались в плену кажущейся «точности» полученных оценок, отображенных с тремя знаками после запятой. Но, что на самом деле означают эти вероятности? Не более того, что при наблюдаемой совокупности симптомов Холецистит может подтвердиться в 75–80 случаях из ста. А второй диагноз, окажется справедливым для 60–70 пациенток из ста. Ясно, что в такой ситуации отдавать явное предпочтение первому диагнозу будет не корректным. Если различие между соседними по значимостям ранговых оценок гипотезами недостаточно велико, то необходимо использование дополнительных методов дифференциальной диагностики.

В качестве такового предлагается использовать байесовский подход в сочетании с таблицами принятия решений. Предлагаемая процедура распознания наиболее достоверной гипотезы о заболевании, проводимая с использованием правила Байеса, таблиц решений и при участии опытного врача – диагноста, описывается ниже. Используемые для этого таблицы решений (табл. 4, 5) будут иметь структуру модифицированной матрицы, рассмотренной выше (табл. 2) с некоторыми дополнениями. Здесь в качестве вероятностей $P(d_i)$ наряду с данными, полученными из различных источников, могут быть использованы значения, полученные в табл. 3. И эти значения предпочтительны, поскольку получены они применительно не к одному симптому, а к выбранной их совокупности.

По существу в предлагаемом подходе наблюдаемая совокупность симптомов рассматривается как единый комплексный объект, где каждый из симптомов системно связан с остальными, что полностью соответствует объективной реальности. И это снимает с рассмотрения проблему их независимости.

Таблица 4

К расчету условных вероятностей

Сравниваемые гипотезы и их априорные вероятности $P(d_i)$			Наблюдаемые симптомы			
			s_1	s_2	…	s_n
Гипотезы	Гипотеза d_1	$P(d_1)$	$P(s_1\|d_1)$	$P(s_2\|d_1)$	…	$P(s_n\|d_1)$
	Гипотеза d_2	$P(d_2)$	$P(s_1\|d_2)$	$P(s_2\|d_2)$	…	$P(s_n\|d_2)$
	…	…	…	…	…	…
	Гипотеза d_n	$P(d_n)$	$P(s_1\|d_m)$	$P(s_2\|d_m)$	…	$P(s_n\|d_m)$
			$B_j = \sum_{i=1}^{n} P(d_i)\,P(s_j \mid d_i)$			

В нижней части таблицы под каждым j-м столбцом записываются вычисленные суммы произведений вероятностей $P(d_i)$ на условные вероятности $P(s_j|d_i)$. Для завершения расчетов заполняется еще одна матрица (табл. 5). В ячейки этой матрицы записываются условные вероятности $P(d_i|s_j)$, вычисляемые по правилу Байеса. Для этого, произведения вероятностей $P(d_i)P(s_j|d_i)$ для каждой ij-й ячейки, делят на суммы этих произведений, вычисленных для каждого j-го столбца и записанных под этим столбцом в табл. 4. В итоге получаем матрицу (табл. 5), все элементы которой (за исключением крайнего правого столбца) получены в полном соответствии с математическими формулами и правилами, используемыми в известных публикациях.

Что касается крайнего правого столбца, где должны заноситься результирующие оценки по каждой из гипотез, то выбор способа интеграции условных вероятностей, проставленных по каждой строке, остается пока не очевидным. Обычно, в известных публикациях предлагается подход, состоящий в том, чтобы считать наблюдаемые симптомы статистически независимыми тогда вероятность одновременного наблюдения вычисляется как произведения вероятностей. Однако такой подход не выдерживает никакой критики по нескольким причинам.

Во-первых, такой подход приемлем, если только причинно-следственные связи между наблюдаемыми случайными величинами неизвестны и природа их установлена быть не может. В данном же случае (выбор гипотез о заболевании) с полной достоверностью известно, что, если не все, то, по крайней мере, появление определенной группы симптомов обусловлен одним из заболеваний (из анализируемой совокупности).

Таблица 5

К расчету условных вероятностей

Сравниваемые гипотезы и их априорные вероятности $P(d_i)$			Наблюдаемые симптомы				Интегральная оценка значимости Q_i
			s_1	s_2	…	s_n	
Гипотезы	Гипотеза d_1	$P(d_1)$	$P(s_1\|d_1)/B_1$	$P(s_2\|d_1)/B_2$	…	$P(s_n\|d_1)/B_j$	Q_i
	Гипотеза d_2	$P(d_2)$	$P(s_1\|d_2)/B_1$	$P(s_2\|d_2)/B_2$	…	$P(s_n\|d_2)/B_j$	Q_i
	…	…	…	…	…	…	Q_i
	Гипотеза d_n	$P(d_n)$	$P(s_1\|d_m)/B_1$	$P(s_2\|d_m)/B_2$	…	$P(s_n\|d_m)/B_j$	Q_i
			$B_j = \sum_{i=1}^{n} P(d_i)P(s_j \mid d_i)$				

На самом деле, системной связи между каждым заболеванием и сопутствующими симптомами соответствует логическая схема «ИЛИ» (без исключения), подразумевающая, что каждому заболеванию сопутствует появление одного, двух, n симптомов в любой совокупности. Такое предположение полностью согласуется с медицинской практикой, зато требует проведения наиболее сложных и громоздких вычислений. При проведении этих вычислений в конечный результат будет входить как сумма всех вероятностей, так и сумма их произведений в различных сочетаниях.

В этой ситуации разумным представляется такой подход, когда для получения предварительной оценки апостериорного значения $\overline{P}^*(d \mid s_\Sigma)$ учитывается только сумма вероятностей, а остальными членами пока пренебрегаем. Такой подход может быть обоснован тем, что слагаемые, являющиеся произведениями вероятностей, явно меньше по величине, чем просто вероятности, так что их вклад в окончательный результат не столь значителен. Обсуждение обоснованности такого подхода планируется провести во второй части настоящей работы. Там же будут предложены несколько новых моделей, использование которых может уменьшить сложность вычислений при использовании Байесовского подхода и, одновременно, повысить обоснованность и правомочность используемых процедур.

Список литературы

1. Бейли Н. Математика в биологии и медицине. – М.: Мир, 1970. – 326 с.

2. Бескровный И.М. Системный анализ и информационные технологии в социальной сфере и здравоохранении // Системный анализ, управление и обработка информации в здравоохранении. – URL: http://econf.rae.ru/article/5287.

3. Бескровный И.М. Анализ и выбор диагностических гипотез на основе модифицированных моделей принятия решений. – URL: http://econf.rae.ru/article/5015.

4. Гудман С.Н. На пути к доказательной биостатистике. Часть 2: байесовский критерий. – URL: http://www.mediasphera.ru/mjmp/2002/2/r2-02-1.htm.

5. Де Гроот М. Оптимальные статистические решения. – М.: Мир, 1974. – 492 с.