

УДК 519.2:534

ТЕОРЕТИКО-ВОЗМОЖНОСТНАЯ МОДЕЛЬ КОМПЬЮТЕРНОЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ

В. А. Газарян, Н. В. Иваницкая, Ю. П. Пытьев, А. К. Шаховская*)

(кафедра компьютерных методов физики)

E-mail: pytyev@phys.msu.su

Теоретико-возможностный метод используется для решения задач медицинской диагностики. Создан возможностный аналог алгоритма «Кора». Возможностный метод диагностики сравнивается с используемыми ранее алгоритмами.

1. Оптимальное решение задачи диагностики в результате минимизации возможности потерь

Функциональные нарушения в системе пищеварения активно изучаются на протяжении ряда последних лет [1]. Проводимые совместно с врачами клиники НИИ питания РАМН исследования позволили значительно продвинуться по пути создания нового комплексного метода компьютерной оценки состояния больных с функциональными нарушениями. На основании опросников пациентов клиники были найдены количественные и качественные характеристики симптомов расстройств ЖКТ, выявлены их связи с изменением психоэмоциональной сферы. По обучающей выборке больных — пациентов НИИ питания с верифицированным диагнозом были определены наиболее часто встречающиеся сочетания признаков (симптомов) синдрома раздраженной толстой кишки (СРТК) с помощью алгоритма типа «Кора» [2], что позволило создать комплексный тест оценки самочувствия больного и описать симптоматику СРТК в общем виде. Однако, как было показано в [3], детерминистские и статистические методы недостаточно эффективны при неформализованном характере данных и ограниченном размере обучающей выборки, содержащей больных с верифицированным диагнозом. Для уточнения диагностических критериев был привлечен теоретико-возможностный метод, что привело к созданию алгоритма, ранжирующего группы симптомов по их возможностям при определенных заболеваниях [3]. Вследствие этого можно определить условную возможность заболевания при наличии тех или иных наборов симптомов. Выявленная теоретико-возможностная симптоматика позволила вплотную подойти к конечной цели — диагностике заболеваний при нечетких данных на основании решающего правила, оптимального с теоретико-возможностной точки зрения, т. е. минимизирующего возможность ошибки классификации.

Сопоставим каждому субъекту (пациенту) n -

мерный нечеткий вектор признаков $\chi = (\chi^1, \chi^2, \dots, \chi^n)$, принимающий значение $x = (x^1, x^2, \dots, x^n)$, где x^j — значение j -го признака (симптома). Пусть \varkappa — нечеткий элемент, значением которого является номер класса k , $k = 1, \dots, M$. Пусть $\varphi^{\chi, \varkappa}(x, k)$ — распределение возможностей нечеткого вектора (χ, \varkappa) , $l(k, q)$ — возможность потерь при отнесении субъекта класса k к классу q , $q = 1, \dots, M$. Будем считать, что $l(k, q) = 0$ при $k = q$ и $l(k, q) = 1$ при $k \neq q$. Далее не будем делать различий между субъектом и присущим ему вектором значений признаков $x = (x^1, x^2, \dots, x^n)$.

Оптимальное решение задачи классификации достигается при минимизации возможности (необходимости) потерь [4]. Согласно теореме о P -оптимизации, субъект χ следует отнести к классу q^* :

$$P(q^*, x) = \min_q P(q, x),$$

$$P(q, x) = \sup_k \min(\varphi^{\chi, \varkappa}(x, k), l(k, q)),$$

где $P(q, x)$ — возможность ошибки при отнесении субъекта $\chi = x$ к классу $\varkappa = q$.

Аналогично строится N -оптимальная оценка q^* класса \varkappa , минимизирующая необходимость ошибки классификации нечеткого вектора χ .

Поскольку $\varphi^{\chi, \varkappa}(x, k) = \min(\phi^{\chi|\varkappa}(x|k), \phi^{\varkappa}(k)) = \phi^{\chi|\varkappa}(x|k)$ в предположении равенства априорных возможностей $\phi^{\varkappa}(k) = 1$, $k = 1, \dots, M$, где $\phi^{\chi|\varkappa}(x|k)$ — условное распределение возможностей пациенту класса k обладать признаками (симптомами) x , то решающее правило можно представить в следующем виде:

$$P(q^*, x) = \min_q P(q, x), \quad (1)$$

$$P(q, x) = \sup_k \min(\phi^{\chi|\varkappa}(x|k), l(k, q)). \quad (2)$$

Условное распределение $\phi^{\chi|\varkappa}(x|k)$ получим в результате стохастического моделирования условных

*) ГУ НИИ питания РАМН.

возможностей определенных наборов признаков при условии принадлежности субъекта выделенным классам.

2. Обучение на основании ранжированных по возможностям групп признаков. Возможностный аналог алгоритма классификации «Кора»

В работе [3] рассмотрен алгоритм поиска всех разбиений $T_1, T_2, \dots \in \{T\}$, множества значений признаков на группы признаков $w_i^{T_j}$, $i = 1, \dots, N_j$, $j \in \{1, 2, \dots\}$ (N_j — число найденных групп признаков, соответствующих разбиению T_j), ранжированных по значениям их условных возможностей $p_i^{T_j} = p(w_i^{T_j}|k)$ при условии принадлежности субъекта, обладающего набором признаков $w_i^{T_j}$, к классу k , $k = 1, \dots, M$:

$$p_1^{T_j} > p_2^{T_j} > \dots > p_N^{T_j}, \quad T_j \in \{T\}.$$

Наборы признаков $w_1^T, w_2^T, \dots, T \in \{T\}$ удовлетворяют по определению условию максимальной разделенности (\gg). Было показано, что так же упорядочены апостериорные возможности $p(k|w_i^T)$ заболевания k при наличии симптомов w_i^T [4]. Наиболее существенными признаками заболевания k являются наборы w_1^T, w_2^T, \dots разбиений T_1, T_2, \dots , имеющие максимальную возможность $p_1^T = p(w_1^T|k) = 1$.

По обучающей выборке находятся множества $\{w\}_k$, $k = 1, \dots, M$, всех (\gg) наборов признаков, имеющих ненулевую возможность. Предъявляется субъект $x = (x^1, x^2, \dots, x^n)$, где x^i — значение i -го признака (симптома), который следует отнести к одному из M классов. Выявляются все наборы $(w_s)_k$, $k = 1, \dots, M$, из множества $\{w\}_k$, присущие субъекту. Пусть найдено S_k таких наборов в классе k : $s = 1, \dots, S_k$. В каждом классе k определяется набор w_{s_k} , имеющий максимальную апостериорную возможность:

$$p(k|w_{s_k}) = \max_s p(k|w_s), \quad s = 1, \dots, S_k.$$

Далее в (2) значение $p(x|k)$ используется в качестве $\phi^{x|k}(x|k)$, причем вместо вектора значений признаков $x = (x^1, x^2, \dots, x^n)$, где x^i — значение i -го признака субъекта x , берется наиболее существенный в классе k набор w_{s_k} .

Субъект x относится к тому классу q^* , которому соответствует минимальная возможность ошибки (1) (в котором есть набор $w_{s_{q^*}}$, имеющий минимальную возможность ошибки):

$$P(q^*, w_{s_{q^*}}) = \min_q P(q, w_{s_q}),$$

что в случае двоичности потерь $l(k, q)$ соответствует максимальной возможности $p(q^*|x)$ субъекту x принадлежать классу q^* . Тогда условная возмож-

ность $P(q^*, w_{s_{q^*}})$ ошибки классификации субъекта x с $w_{s_{q^*}}$ при отнесении его к классу q^* также будет минимальна.

3. Результаты вычислительного эксперимента

Классификация синдрома раздраженной толстой кишки на две группы

Алгоритм возможностной классификации был вначале применен к двум группам СРТК, подробно описанным в работе [2], в которых ранее ранжировались условные возможности замедленной (ускоренной) моторики при условии определенных наборов симптомов у пациента [3]. При классификации по гастроэнтерологическим признакам и предпосылкам болей и при анализе факторов, предшествующих заболеванию [2], в 75% случаев было принято верное решение. Следовательно, можно говорить о том, что даже в рамках одного заболевания СРТК надо устранять несколько различные наборы предпосылок при ускоренной и замедленной моторике. Несмотря на невысокую точность классификации, вызванную тесным переплетением симптомов разных групп одного заболевания, результаты предварительной диагностики имеют большое значение для врача. Так как органических нарушений при СРТК, как правило, нет, картина заболевания отражена прежде всего в тестах. Теперь кропотливую работу по всестороннему анализу опросников берет на себя компьютерный метод оценки состояния больного. Врач, пользуясь полученными результатами, ставит окончательный диагноз и назначает лечение.

Диагностика острого аппендицита

Не вызывает сомнения чрезвычайная важность своевременного обнаружения острого аппендицита для здоровья и жизни больного. При подозрении на острый аппендицит очень важно опираться не на отдельные симптомы желудочно-кишечных расстройств, а выделить такие группы признаков, которые действительно характеризуют данное заболевание и позволяют отличить больного человека от здорового. Острый аппендицит, как класс заболеваний, в свою очередь делится на три группы: гангренозный, флегмонозный и катаральный. Врачи отмечают, что их признаки сильно пересекаются, и привлечение компьютерных методов классификации к анализу симптоматики и диагностике трех форм острого аппендицита также является актуальным.

Обучающая выборка — результаты обследования 103 человек из четырех классов: три группы острого аппендицита: гангренозный (28 человек), флегмонозный (25), катаральный (26) и группа больных с неподтвержденным диагнозом (24). В качестве признаков врачами выбрано 8 различных симптомов (таблица). Каждый признак может принимать

Типичные признаки острого аппендицита как класса

Интенсивность болей	Продолжительность болей	Частота пульса	Лейкоциты крови	Изменение языка	Симптом Щеткина	Симптом Ровзинга	Защитное мышечное напряжение
2				1			
2						2	
				1	2		
				1		2	
					2	2	
					2		2
						2	2
Различные для трех видов аппендицита признаки							
гангренозный аппендицит							
2	3						
2					2		
2	3					2	
2							2
	3					2	
2				1	2		
2				1		2	
				1			2
2				1			2
2					2	2	
2					2		2
2						2	2
флегмонозный аппендицит							
	4						2
		1			2		
			2		2		
катаральный аппендицит							
2		1					
		1		1			
		1				2	
			2	1			
Значения признаков, указанные в таблице, отражают степень выраженности симптома.							

от двух до четырех значений в ранговой шкале в зависимости от выраженности симптома. Образом каждого субъекта здесь также является n -мерный нечеткий вектор признаков $\chi = (\chi^1, \chi^2, \dots, \chi^n)$, $n = 8$, принимающий значение $x = (x^1, x^2, \dots, x^n)$, где x^j — значение j -го признака (симптома).

Ранее для диагностики острого аппендицита использовался метод дискриминантного анализа [5], требующий существенных ограничений. Во-первых, необходима линейная независимость признаков. Во-вторых, ковариационные матрицы классов должны быть одинаковы. В-третьих, необходимо, чтобы образы всех классов подчинялись нормальному закону распределения. Недостатком метода является также отсутствие математического критерия качества классификации. Результат целиком зависит от положения центроидов в пространстве дискриминантных функций и считается тем лучше, чем большее число образов имеет правильную классификацию. Невозможно выявить группы симптомов, характеризующие классы, в пространстве

дискриминантных функций. Врачам непонятно, на каком основании компьютер ставит диагнозы их пациентам, и подобный способ диагностики вызывает естественное недоверие.

Теоретико-возможностный метод классификации не требует ограничений на свойства образов и осуществляет классификацию больных острым аппендицитом на основании решающего правила, минимизирующего возможность ошибки классификации. Возможностный алгоритм прост в программной реализации, результаты представлены в привычном для врачей виде. В таблице наглядно демонстрируются характерные наборы симптомов четырех групп заболеваний. Как и при дискриминантном анализе [5], отмечено, что 4-я группа (неподтвержденный диагноз) резко отличается от первых трех. Но если ранее она образовывала отдельный кластер в пространстве дискриминантных функций, то теперь врачам представлены те сочетания симптомов, которые отличают данный класс от других. Семь типичных наборов признаков острого аппендицита

как класса в целом, выявленные возможным методом, являются диагностическим критерием данного класса. Возможность диагноза «острый аппендицит» равна единице, если у больного присутствуют данные наборы, присущие всем трем группам. И наконец, найдены характерные для каждой группы аппендицита комбинации симптомов.

4. Сравнительная характеристика возможностного метода распознавания в медицине и алгоритма «Кора»

Метод «Кора» использует аналог байесового решающего правила, минимизирующего вероятность ошибки классификации [2]. Предварительно в каждом классе находятся так называемые представительные наборы, состоящие из признаков, частота встречаемости которых в обучающей выборке превышает некоторый порог, задаваемый самим исследователем, что вносит произвол в выбор существенных характеристик заболевания (класса). Все признаки представляются в бинарном виде. В процессе бинаризации также используется пороговое значение, в пределах которого признаки считаются «неразличимыми» при классификации. При этом выбор обоих порогов обусловлен лишь здравым смыслом, математическое обоснование отсутствует. Объект сравнивается по всем признакам с каждым другим объектом обучающей выборки в отдельности, и такая «близость» не ассоциативна.

Близость признаков при возможностном подходе устанавливается самим алгоритмом индивидуально для каждого класса. Признаки, объединенные в группу w_i^T , т.е. имеющие одинаковую возмож-

ность $p_i^T = p(w_i^T|k)$ для каждого i , будут «неразличимы» с точки зрения их значимости в классе. Соответственно признаки принципиально «различны», если принадлежат группам, имеющим разную возможность в классе k . Процессы сравнения признаков и выбора характерных в классе в возможностном подходе объединены.

На этапе распознавания механизмы принятия решений схожи. «Кора» осуществляет выбор класса, используя представительный набор субъекта, «наилучший» с вероятностной точки зрения, при возможностном распознавании выбирается класс, в котором возможность наиболее значимой группы максимальна.

Литература

1. Шаховская А.К., Есаулов В.И., Лоранская Т.И. // Мат-лы III Российской гастроэнтерологической недели. М., 1997. С. 65.
2. Газарян В.А., Иваницкая Н.В., Пытьев Ю.П., Шаховская А.К. // Вестн. Моск. ун-та. Физ. Астрон. 2003. № 2. С. 12 (Moscow University Phys. Bull. 2003. N 2. P. 12).
3. Газарян В.А., Илюшин В.Л., Пытьев Ю.П., Шаховская А.К. // Вестн. Моск. ун-та. Физ. Астрон. 2005. № 4. С. 3 (Moscow University Phys. Bull. 2005. N 4. P. 1).
4. Пытьев Ю.П. Возможность. Элементы теории и применения. М., 2000.
5. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. М., 1989.

Поступила в редакцию
26.12.05