

Метод монотонных функций и его применение к выбору типа операции больных раком предстательной железы по дооперационным данным¹

В.Г. Гитис*, А.Б. Дерендяев*, К.Н. Петров*, С.А. Пирогов*, Е.Ф. Юрков*, Н.С.
Сергеева**, Б. Я. Алексеев**, А.Д. Каприн**

*Институт проблем передачи информации, Российская академия наук, Москва, Россия

**Московский научно-исследовательский онкологический институт имени П.А. Герцена – филиал
ФГБУ “Национальный медицинский исследовательский радиологический центр” Министерства
здравоохранения Российской Федерации, Москва, 125284, Россия

Поступила в редколлегию 17.02.2021

Аннотация—Предлагается новый метод машинного обучения для решения задач бинарной классификации с упорядоченными классами. Метод использует модель, в которой решающее правило ищется в классе монотонных функций на пространстве признаков. Метод применен к решению задачи классификации больных агрессивной и индолентной формами рака предстательной железы по долечевым данным.

КЛЮЧЕВЫЕ СЛОВА: машинное обучение, метод монотонных функций, отношение правдоподобия, рак предстательной железы

1. ВВЕДЕНИЕ

В современной медицине широко применяются методы машинного обучения. Эти методы выявляют из массива медицинских данных знания, которые позволяют устанавливать диагноз заболевания, определять тяжесть состояния больного, выбирать способ лечения и прогнозировать его состояние после лечения.

В данной работе мы предлагаем новый метод машинного обучения и рассматриваем его применение для выбора одного из двух типов операции по поводу рака предстательной железы (РПЖ), основанного на дооперационных данных. Тип и объем хирургического вмешательства принципиально отличаются для агрессивной и индолентной формы РПЖ [1–4]. Агрессивная форма заболевания относится к более тяжелой форме заболевания, которое характеризуется наличием более распространенного опухолевого процесса.

Информация о упорядоченности форм заболевания по их тяжести может быть использована при классификации. В работах [5, 6] для решения подобных задач была предложена модель, в которой для восстановления зависимости тяжести заболевания от функциональных показателей больного использовались монотонные функции. Известно, что для функциональных показателей существует понятие нормы. Норма определяется, как интервал значений показателя, характерный для здорового человека. Модель вводит условие монотонности, согласно которому уменьшение отклонения любого из показателей от нормы, при прочих равных условиях, уменьшает или не изменяет (но не увеличивает) тяжесть заболевания.

В нашем случае мы применяем условие монотонности для классификации больных на два упорядоченных по тяжести заболевания класса. Метод обучения классификации мы назвали

¹ Исследование выполнено в ИППИ РАН частично за счет гранта РФФИ 20-07-00445.

методом монотонных функций. Этот метод является развитием наших работ по машинному обучению одноклассовой классификации, применяемых для решения задач прогноза землетрясений и поиска аномальных природных зон, таких, как зоны возможных очагов сильных землетрясений или зоны месторождений полезных ископаемых [7, 8]. В разделе 2 мы рассматриваем основы метода монотонных функций. В разделе 3 приведены результаты применения метода к задаче классификации больных с агрессивным и индолентным РПЖ.

2. МЕТОД МОНОТОННЫХ ФУНКЦИЙ

Пусть имеется два класса объектов ω_1 и ω_2 . Нам известна ограниченная выборка объектов. Объекты $q = 1, 2, \dots, Q_1$ относятся к классу ω_1 , а объекты $\bar{q} = 1, 2, \dots, Q_2$ относятся к классу ω_2 . Каждый объект описывается набором свойств (набором признаков). Объектам соответствуют точки i -мерного пространства признаков $\mathbf{x}^{(q)} = \{x_i^{(q)}\}$ и $\mathbf{x}^{(\bar{q})} = \{x_i^{(\bar{q})}\}$, $i = 1, \dots, I$. Задача состоит в том, чтобы по выборке найти правило классификации.

В нашем случае объектами классов являются больные, признаками объектов являются дооперационные показатели больных, классами — две формы заболевания, которые характеризуют его тяжесть. Форма заболевания определяется после операции. Задача состоит в том, чтобы по дооперационным данным найти правило классификации больных по форме заболевания.

Мы рассматриваем следующую модель данных. Пусть форма заболевания у больных класса ω_1 менее тяжелая, чем у больных класса ω_2 . Для каждого признака известен интервал значений $[a_i, b_i]$, называемый нормой данного показателя. Больные с показателями $\mathbf{x} = \{x_i\}$, отклонения которых от нормы покомпонентно меньше или равны отклонениям от нормы показателей точки $x^{(q)} \in \omega_1$ также относятся к классу ω_1 (условие монотонности).

Для простоты изложения без ограничения общности далее будем считать, что все точки $\mathbf{x}$ пространства признаков принимают значения либо из интервалов норм показателей, $x_i \in [a_i, b_i]$, либо большие правой границы этих интервалов, $x_i > b_i$. Тогда согласно условию монотонности больной, которому соответствует точка $\mathbf{x} = \{x_i\}$, $i = 1, 2, \dots, I$, относится к классу ω_1 , если $x_i \leq x_i^{(q)}$, $i = 1, 2, \dots, I$, где $\mathbf{x}^{(q)} = \{x_i^{(q)}\} \in \omega_1$. Очевидно, что условие монотонности соответствует идеальной ситуации, в которой представлены все необходимые для классификации показатели объектов и значения показателей измерены достаточно точно. В реальной ситуации из-за неполного описания дифференциальных свойств объектов и неточности измерений показателей неравенства $x_i \leq x_i^{(q)}$, $i = 1, 2, \dots, I$, для некоторых пар объектов не выполняются. Это является одной из причин ошибочной классификации.

Рассмотрим алгоритм обучения классификации.

С каждой точкой учебной выборки $\mathbf{x}^{(q)} \in \omega_1$ свяжем область пространства признаков

$$W^{(q)} = \left\{ \mathbf{x} : x_i \leq x_i^{(q)}, i = 1, 2, \dots, I \right\}.$$

Будем называть это множество ортантом $W^{(q)}$ с вершиной в точке $\mathbf{x}^{(q)}$. Поскольку условие монотонности не выполняется для всех объектов, то ортант с вершиной в точке $\mathbf{x}^{(q)}$ может содержать $Q_1^{(q)}$ точек класса ω_1 и $Q_2^{(q)}$ точек класса ω_2 . Каждому ортанту $W^{(q)}$ с вершиной $\mathbf{x}^{(q)}$ соответствуют три параметра: $P_1^{(q)} = Q_1^{(q)}/Q_1$ — вероятность попадания в ортант точек класса ω_1 , $P_2^{(q)} = Q_2^{(q)}/Q_2$ — вероятность попадания в ортант точек класса ω_2 и $L^{(q)} = P_1^{(q)}/(P_2^{(q)} + 1/Q_1Q_2)$ — отношение правдоподобия, где слагаемое $1/Q_1Q_2$ корректирует значение $L^{(q)}$ при $P_2^{(q)} = 0$.

Алгоритм состоит из следующих шагов.

1. Вычислить для всех ортантов класса ω_1 оценки отношения правдоподобия и получить выборку $\{L^{(q)}\}$, $q = 1, 2, \dots, Q_1$.
2. Перенумеровать ортанты $W^{(q)}$ в соответствии с уменьшением оценок отношения правдоподобия $L^{(1)} \geq L^{(2)} \geq \dots \geq L^{(Q_1)}$.
3. Присвоить точкам множества $W^{(1)}$ значения функций $U(\omega_1)^{(1)}$ и $V(\omega_1)^{(1)}$, где $U(\omega_1)^{(1)} = P_1^{(1)}$ – вероятность точек выборки класса ω_1 в ортанте $W^{(1)}$, $V(\omega_1)^{(1)} = P_2^{(1)}$ – вероятность точек выборки класса ω_2 в ортанте $W^{(1)}$. Присвоить точкам из множества $W^{(2)} \setminus W^{(1)}$ значения функций $U(\omega_1)^{(2)}$ и $V(\omega_1)^{(2)}$, где $U(\omega_1)^{(2)}$ – вероятность, которая определяется как отношение количества точек класса ω_1 из множества $W^{(1)} \cup W^{(2)}$ к количеству всех точек Q_1 учебной выборки класса ω_1 , $V(\omega_1)^{(2)}$ – вероятность, равная отношению количества точек класса ω_2 из множества $W^{(1)} \cup W^{(2)}$ к количеству всех точек Q_2 учебной выборки класса ω_2 . Присвоить множеству точек $W^{(3)} \setminus (W^{(1)} \cup W^{(2)})$ значения вероятностей $U(\omega_1)^{(3)}$ и $V(\omega_1)^{(3)}$, которые определяются по множеству $W^{(1)} \cup W^{(2)} \cup W^{(3)}$. Далее аналогично вычислять значения $U(\omega_1)^{(n)}$ и $V(\omega_1)^{(n)}$ на множествах точек пространства признаков в соответствии указанной в пункте 2 нумерацией ортантов. Присвоить оставшимся точкам пространства признаков значения $U(\omega_1) = V(\omega_1) = 1$.

Таким образом, на пространстве признаков по учебной выборке строится система вложенных множеств, которая определяет две функции $U(\mathbf{x})$ (вероятность успешной классификации объектов класса ω_1) и $V(\mathbf{x})$ (вероятность ошибочной классификации объектов класса ω_2), связанные между собой монотонно зависимостью $U(V)$. По значениям функций $U(\mathbf{x})$ и $V(\mathbf{x})$ для любых заданных значений порогов α и β можно вычислить решающее правило классификации: $\mathbf{x} \in \omega_1$, если $V(\mathbf{x}) \leq \alpha$ и $\mathbf{x} \in \omega_2$, если $V(\mathbf{x}) > \alpha$ или $\mathbf{x} \in \omega_1$, если $U(\mathbf{x}) \leq \beta$ и $\mathbf{x} \in \omega_2$, если $U(\mathbf{x}) > \beta$. Полученные решающие правила можно представить в виде логических функций от признаков. Такие решения дают объяснение правила классификации в терминах предметной области.

Алгоритм вычисления логического правила классификации состоит в следующем.

1. Отобрать точки $\mathbf{x}^{(q)}$ класса ω_1 , для которых вероятность ошибочной классификации объектов класса ω_2 как объектов класса ω_1 , $V(\mathbf{x}) \leq \alpha$ (для отбора точек можно использовать аналогичное условие $U(\mathbf{x}) \leq \beta$).
2. Построить матрицу принадлежности (M_{pq}) . Строки матрицы обозначают ортанты $W^{(P)}$, с вершинами в отобранных на предыдущем шаге точках $\mathbf{x}^{(P)}$, а столбцы матрицы обозначают точки класса $\mathbf{x}^{(q)}$. Элемент матрицы $m_{pq} = 1$, если точка $\mathbf{x}^{(q)}$ класса ω_1 принадлежит ортанту с индексом P и $m_{pq} = 0$ в противном случае. Принадлежность точки пространства признаков ортанту $W^{(P)}$ определяется конъюнкцией: $\mathbf{x} \in W^{(P)}$, если $x_i \leq x_i^{(P)}$ для всех $i = 1, 2, \dots, I$. Совокупность строк матрицы соответствует дизъюнкции конъюнкций.
3. Определить подмножество из минимального числа ортантов, которые содержат все точки класса, отобранные в пункте 2. Для этого надо в матрице принадлежности (M_{pq}) выбрать подматрицу из минимального числа строк, в каждом столбце которой оставалась хотя бы одна единица. Выбор строк матрицы (M_{pq}) выполняется с помощью метода минимизации дизъюнктивных нормальных форм булевых функций [9]. В результате минимизации получаем логическое правило в виде дизъюнкции конъюнкций. Далее с помощью лингвистических переменных и заранее заготовленных шаблонов полученное логическое правило можно представить в виде достаточно простого текстового выражения.

Число строк матрицы, оставшихся после минимизации, может оказаться слишком большим для наглядной интерпретации решающего правила в виде дизъюнкции конъюнкций. Алгоритм

коррекции аппроксимирует предыдущее решение и создает логическое правило из заранее заданного специалистом числа конъюнкций K , которое успешно классифицирует все отобранные в пункте 3.1 точки класса ω_1 .

Алгоритм коррекции состоит в следующем.

- Построить для всех пар ортантов с вершинами в точках $\mathbf{x}^{(q)}$ и $\mathbf{x}^{(r)}$ новые ортанты $W^{(s)}$ с вершинами $\mathbf{x}^{(s)} = \{x_i^{(s)}\}$, где $x_i^{(s)} = \max(x_i^{(q)}, x_i^{(r)})$. Очевидно, что ортант $W^{(s)}$ покрывает все точки $W^{(q)} \cup W^{(r)}$, которые покрывают ортанты $W^{(q)}$ и $W^{(r)}$.
- Отобрать из новых ортантов, ортант $W^{(s)}$ с максимальным значением отношения правдоподобия $L^{(s)}$ и удалить два ортанта $W^{(q)}$ и $W^{(r)}$, из объединения которых ортант $W^{(s)}$ был получен. При этом число оставшихся ортантов уменьшается на единицу.
- Повторять процедуру повторяется до тех пор, пока число оставшихся ортантов не окажется равным заранее заданному числу конъюнкций K .

Из-за увеличения размеров ортантов после коррекции может увеличиваться вероятность ложной тревоги. Если этот результат неприемлем, то число ортантов приходится увеличить. Пороговые величины α и β и число конъюнкций K в логической функции подбираются с помощью нескольких итераций.

3. ОПРЕДЕЛЕНИЕ ТИПА ОПЕРАЦИИ БОЛЬНЫХ РПЖ ПО ДОЛЕЧЕБНЫМ ДАННЫМ

Долечебные данные используются для принятия клинических решений по выбору типа лечения и оценке его близких и отдаленных клинических эффектов. Примером является цикл работ по индексу <Cancer of the Prostate Risk Assessment> (CAPRA), в которых предложены методы использования долечебных данных больных раком предстательной железы (РПЖ) [10–12]. Индекс CAPRA разработан для облегчения классификации риска заболевания при принятии по долечебным данным клинических оценок вероятности того, что данная опухоль будет рецидивировать после лечения, прогрессировать и представлять угрозу для жизни. В работах [1, 3] по долечебным данным решалась задача классификации больных с агрессивным и индолентным РПЖ. Ошибки в выборе типа операции приводят к ухудшению результатов лечения больных в случае недооценки формы опухолевого процесса, либо к ухудшению качества жизни пациентов в случае завышения формы РПЖ, и, вследствие этого, неоправданного расширения объема хирургического вмешательства. Для решения задачи применен метод логистической регрессии. Больные классифицировались по пяти признакам. По данным перекрёстной проверки алгоритм позволяет выявить группу, содержащую 55% больных с агрессивным РПЖ при отсутствии в ней больных с индолентным статусом РПЖ.

Для классификации больных с агрессивным и с индолентным РПЖ по дооперационным показателям мы используем данные, подготовленные специалистами Московского научно-исследовательского онкологического института (МНИОИ) имени П. А. Герцена. Выборка данных взята из историй болезней 341 больных, прооперированных с диагнозом рак предстательной железы (РПЖ): 124 с индолентным РПЖ (класс ω_1) и 217 больных с агрессивным РПЖ (класс ω_2). Форма РПЖ заболевания определялся по послеоперационным данным.

Дооперационное состояние больного описывается следующими долечебными показателями.

Данные гистологического анализа (игольная биопсия). Группа Глисона (GrGl) — биоптата по шкале Глисона для двух характерных участков [13]. Показатель принимает следующие значения:

GrGl=1, если сумма баллов 6 или менее.

GrGl=2, если сумма баллов 7=3+4 (порядок слагаемых важен).

$GrGl=3$, если сумма баллов $7=4+3$ (порядок слагаемых важен).

$GrGl=4$, если сумма баллов 8 или более.

Данные иммунохимического анализа крови (хемилуминесцентный анализатор Access 2, Beckman Coulter, США, калибровка Hybritech) [14].

1. ПСА (общПСА), простатический специфический антиген.
2. свПСА, свободный ПСА.
3. про2ПСА ($[-2]$ проПСА), изоформа-предшественник ПСА.
4. Отношение ПСА / свПСА.
5. $ИЗП=ПСА \times (про\ ПСА) / (свПСА)$ — индекс здоровья простаты.

Клинические данные.

6. Клиническая стадия (GrT):
 - (а) если опухолевый процесс локальный ($T1$ и $T2$).
 - (б) если опухолевый процесс местнораспространённый ($T3$).
7. Объём простаты [$см^3$] — по данным ультразвукового обследования.

Антропометрические данные.

8. Возраст [$г$].
9. Индекс массы тела [$кг/(Рост, м)^2$].

Форма заболевания после операции определялась по двум послеоперационным показателям: клиническая стадия (GrT) и группа Глисона ($GrGl$). Если в результате гистологического исследования установлено, что степень распространённости первичной опухоли в предстательной железе не выше $T2$ (опухолевый процесс локальный и метастазы как в регионарных лимфатических узлах, так и отдалённые отсутствуют) и при этом группа Глисона не более 6 баллов, то статус РПЖ считался индолентным. В противном случае статус РПЖ считался агрессивным.

Наилучшие результаты классификации получены для следующих трех признаков: GrT — клиническая стадия, $GrGl$ — Группа Глисона и ПСА. На Рис. 1 показаны графики зависимостей вероятности успешной классификации больных с индолентной формой заболевания $U(x)$ от вероятности ошибочной классификации больных с агрессивной формой заболевания $V(x)$ (вероятность ложной тревоги). График 1 показывает зависимость $U(V)$ при тестировании по учебной выборке, а график 2 показывает зависимость $U(V)$, полученную с помощью k -блочной кросс-валидации при $k = 10$. Высоты столбиков на графике 2 показывают величины среднеквадратичных отклонений. Видно, что графики 1 и 2 мало отличаются.

Следует отметить, что результаты классификации мало отличаются при замене показателя ПСА на ИЗП или на отношение ПСА к свободному ПСА. Добавление к выбранному набору показателей других признаков значительно увеличивает разброс между результатами тестирования по учебной выборке и результатами тестирования с помощью кросс-валидации. При этом результат тестирования правила классификации по учебной выборке заметно улучшается, а результат тестирования с помощью кросс-валидации немного ухудшается. Поэтому такие варианты решения задачи не рассматриваются.

Из Рис. 1 можно видеть, что результат классификации по учебной выборке незначительно отличается от результата тестирования с помощью кросс-валидации. Поэтому в качестве объяснения решения можно использовать логические правила, полученные по учебной выборке. Зависимость $U(V)$ определяет порог, который определяет точку баланса ошибок классификации больных с индолентным и агрессивным РПЖ [15, 16]. Для этого порога решению по

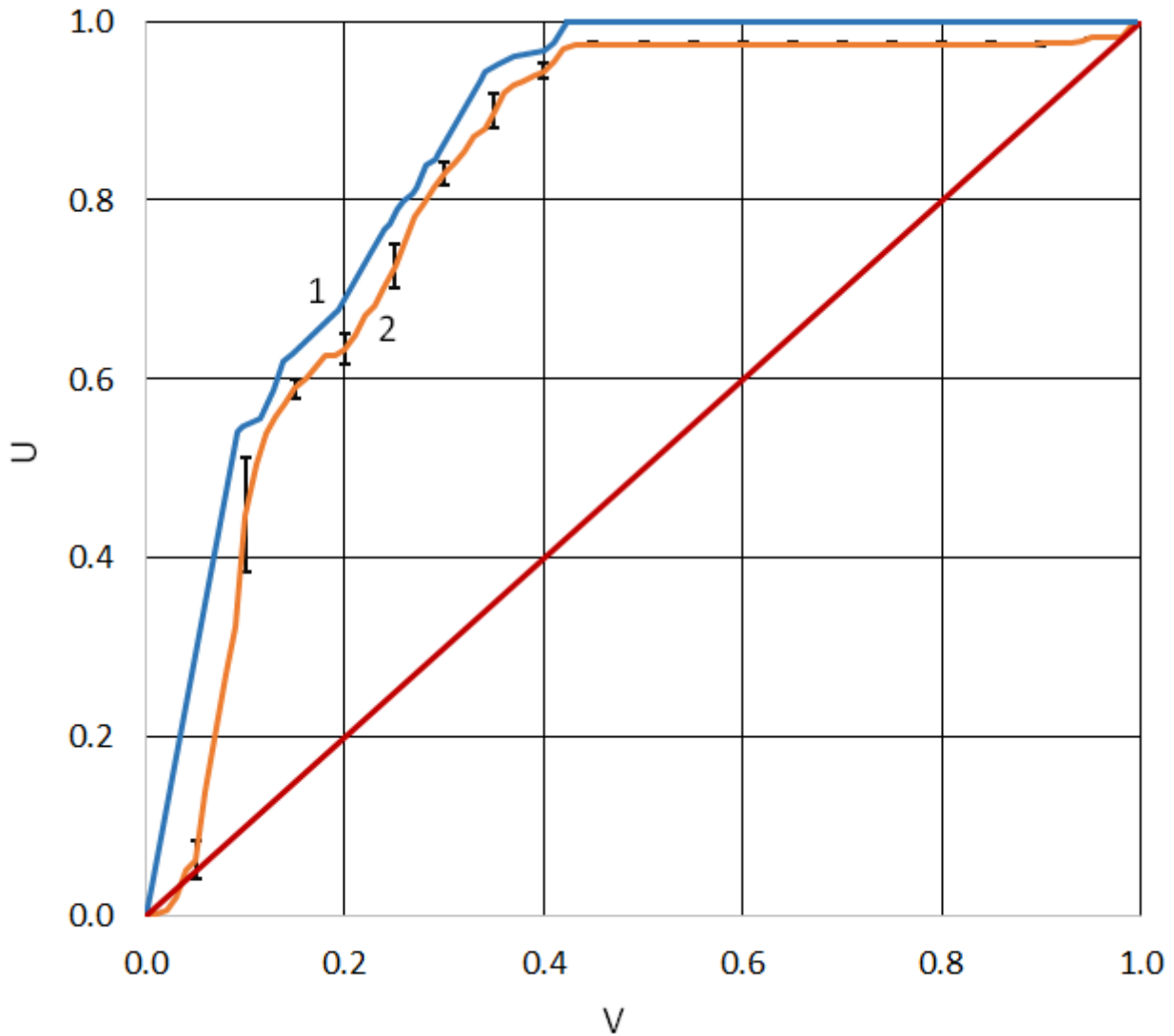


Рис. 1. Графики $U(V)$ зависимостей вероятности успешной классификации больных с индолентной формой РПЖ (U) от вероятности ошибочной классификации больных с агрессивной формой (V). 1 — зависимость $U(V)$, полученная при тестировании по учебной выборке; 2 — зависимость $U(V)$, полученная с помощью 10-блочной кросс-валидации; вертикальные отрезки определяют величины среднеквадратичных отклонений.

учебной выборке (график 1) соответствует вероятность успешной классификации больных с индолентным РПЖ $U(\mathbf{x}) = 0.765$, вероятность ошибочной классификации больных с агрессивным РПЖ $V(\mathbf{x}) = 0.235$.

ЕСЛИ ВЫПОЛНЯЕТСЯ УСЛОВИЕ:

($PSA \leq 9.72$) **И** ($GrGl=1$)

ИЛИ ($PSA \leq 8.24$) **И** ($GrGl \leq 2$) **И** ($GrT=1$)

ИЛИ ($PSA \leq 11.39$) **И** ($GrGl=1$) **И** ($GrT=1$)

ТО у больного индолентный РПЖ с вероятностью $U(\mathbf{x}) = 0.765$ при вероятности ошибочной классификации больного с агрессивным РПЖ, $V(\mathbf{x}) = 0.235$.

Если данное условие не выполняется, то у больного агрессивная форма РПЖ с вероятностью $1 - V(\mathbf{x}) = 0.765$ при вероятности ошибочной классификации больного с индолентной формой РПЖ, $1 - U(\mathbf{x}) = 0.235$.

Ниже мы приводим несколько примеров найденного правила классификации для порогов $\alpha = 0.10$, $\alpha = 0.15$, $\beta = 0.90$ и $\beta = 0.85$.

Порогу $\alpha = 0.10$ соответствует правило:

ЕСЛИ ВЫПОЛНЯЕТСЯ УСЛОВИЕ:

($PSA \leq 8.21$) **И** ($GrGl=1$)

ИЛИ ($PSA \leq 8.40$) **И** ($GrGl=1$) **И** ($GrT=1$)

ТО у больного индолентный РПЖ с вероятностью $U(\omega_1) = 0.548$ при вероятности ошибочной классификации больного с агрессивным РПЖ, $V(\omega_1) = 0.097$.

Порогу $\alpha = 0.15$ соответствует правило:

ЕСЛИ ВЫПОЛНЯЕТСЯ УСЛОВИЕ:

($PSA \leq 9.72$) **И** ($GrGl=1$)

ИЛИ ($PSA \leq 10.06$) **И** ($GrGl=1$) **И** ($GrT=1$)

ТО у больного индолентный РПЖ с вероятностью $U(\omega_1) = 0.629$ при вероятности ошибочной классификации больного с агрессивным РПЖ, $V(\omega_1) = 0.147$.

Порогу $\beta = 0.90$ соответствует правило:

ЕСЛИ НЕ ВЫПОЛНЯЕТСЯ УСЛОВИЕ:

($PSA \leq 9.72$) **И** ($GrGl=1$)

ИЛИ ($PSA \leq 8.24$) **И** ($GrGl \leq 2$) **И** ($GrT=1$)

ИЛИ ($PSA \leq 19.09$) **И** ($GrGl=1$) **И** ($GrT=1$)

ТО у больного агрессивный РПЖ с вероятностью $1 - V(\omega_1) = 0.659$ при вероятности ошибочной классификации больного с индолентным РПЖ, $1 - U(\omega_1) = 0.057$.

Порогу $\beta = 0.85$ соответствует правило:

ЕСЛИ НЕ ВЫПОЛНЯЕТСЯ УСЛОВИЕ:

($PSA \leq 9.72$) **И** ($GrGl=1$)

ИЛИ ($PSA \leq 8.24$) **И** ($GrGl \leq 2$) **И** ($GrT=1$)

ИЛИ ($PSA \leq 13.96$) **И** ($GrGl=1$) **И** ($GrT=1$)

ТО у больного агрессивный РПЖ с вероятностью $1 - V(\omega_1) = 0.719$ при вероятности ошибочной классификации больного с индолентным РПЖ, $1 - U(\omega_1) = 0.161$.

4. ЗАКЛЮЧЕНИЕ

Ряд дисциплин изучает сложные процессы, исследуя связи с более простыми их свойствами. При этом, одна из задач состоит в том, чтобы по значениям этих свойств (признакам) устанавливать определенные состояния процесса. Для этого специалисты анализируют связи процесса с каждым показателем в отдельности и определяют те значения показателей, при которых процесс не переходит в критические состояния. Например, в медицине для здоровых

людей установлены границы нормы для показателей функциональной диагностики. Можно предполагать, что, чем меньше выход показателя за пределы нормальных значений, тем, при прочих равных условиях меньше или равно (но не больше) отклонение процесса от своего нормального состояния. Это означает, что отклонения процесса от нормы могут быть представлены монотонными функциями от значений признаков, характеризующих выход показателей процесса за пределы интервала нормальных значений.

Метод монотонных функций разработан для решения задач бинарной классификации с упорядоченными классами. Метод решения использует предположение о том, что решающее правило классификации связано с порядковым номером класса монотонной неубывающей функцией от показателей состояния больного (от признаков классификации). В упрощенной формулировке метод классифицирует принадлежность объекта первому классу, если все значения признаков объекта меньше или равны значениям признаков любого объекта учебной выборки этого класса. Если значение хотя бы одного признака объекта больше соответствующего значения объекта учебной выборки первого класса, то принимается решение о принадлежности такого объекта второму классу. Условие монотонности допускает сравнение больных по тяжести заболевания. Действительно, если все ключевые показатели у наблюдаемого больного ближе к норме показателей здорового человека, чем у контрольного больного с известной тяжестью заболевания, то вполне естественно считать состояние данного больного более легким. С другой стороны, если хотя бы один из ключевых показателей наблюдаемого больного дальше от нормы, чем у контрольного больного, то возможно, что тяжесть заболевания данного больного выше.

Мы проанализировали данные, взятые из 341 историй болезней прооперированных больных с индолентными и агрессивными формами РПЖ. В результате анализа дооперационных и послеоперационных данных найдена зависимость формы заболевания РПЖ от дооперационных данных. Эта зависимость используется для классификации формы заболевания. Для этого требуется указать алгоритму либо пороговое значение вероятности успешной классификации больных с индолентной формой заболевания, либо пороговое значение вероятности ошибочной классификации больных с агрессивной формой заболевания. Приведенные примеры показывают, что вычисляемые логические правила являются достаточно простыми и легко интерпретируются в терминах дооперационных показателей формы заболевания.

СПИСОК ЛИТЕРАТУРЫ

1. Sergeeva N. S. et al. APHIG: a new multiparameter index for prostate cancer //Cancer Urology. – 2016. – Т. 12. – №. 4. – С. 94–103.
2. Сергеева Н. С. и др. Валидация пороговых решающих правил и калькулятора для алгоритма ВИЗГ, предназначенного для уточнения стадии рака предстательной железы до начала лечения //Онкоурология. – 2020. – №. 1.
3. Юрков Е. Ф. и др. Прогноз агрессивного статуса рака предстательной железы по дооперационным данным //Информационные процессы. – 2016. – Т. 16. – №. 4. – С. 301–311.
4. Юрков Е. Ф. и др. Модель диагностики с учётом врачебных предпочтений. Прогноз клинического статуса рака предстательной железы //Информационные процессы. – 2018. – Т. 18. – №. 4. – С. 243–260.
5. Веницкая Р.С., Гитис В.Г., Розина Н.Н., Ардашникова С.Н. Алгоритмический метод оценки легочной недостаточности у детей с хронической бронхолегочной патологией // Функциональные методы исследования в педиатрии, М.: 1976. С. 81–83.
6. Турбович И.Т., Веницкая Р.С., Гитис В.Г., Нагорнов В.С., Сунгурян Н.Н., Ерамян С.Г., Юрков Е.Ф. Анализ связи между тяжестью состояния больного и его физиологическими показателями // Распознавание образов, М.: Наука, 1977. 22–26.

7. Gitis V. G., Derendyaev A. B. Machine learning methods for seismic hazards forecast // *Geosciences*. – 2019. – Т. 9. – №. 7. – С. 308.
8. Gitis V. G., Derendyaev A. B., Petrov K. N. A Method for Identification of Anomalous Geological Zones // *Journal of Communications Technology and Electronics*. – 2020. – Т. 65. – №. 12. – С. 1531–1541.
9. Закревский А. Д. Оптимизация покрытий множеств // *Логический язык для представления алгоритмов синтеза релейных устройств*. – М.: Наука. – 1966. – С. 136–147.
10. Cooperberg M. R. et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy // *The Journal of urology*. – 2005. – Т. 173. – №. 6. – С. 1938–1942.
11. Cooperberg M. R., Hilton J. F., Carroll P. R. The CAPRA-S score: a straightforward tool for improved prediction of outcomes after radical prostatectomy // *Cancer*. – 2011. – Т. 117. – №. 22. – С. 5039–5046.
12. Lorent M. et al. Meta-analysis of predictive models to assess the clinical validity and utility for patient-centered medical decision making: application to the CAncer of the Prostate Risk Assessment (CAPRA) // *BMC medical informatics and decision making*. – 2019. – Т. 19. – №. 1. – С. 1–12.
13. Волченко Н. Н. и др. Сравнительный анализ степени дифференцировки рака предстательной железы на биопсийном и операционном материалах // *Онкология. Журнал им. П.А. Герцена*. – 2016. – Т. 5. – №. 1. – С. 18–22.
14. Stephan C. et al. Toward metrological traceability in the determination of prostate-specific antigen (PSA): calibrating Beckman Coulter Hybritech Access PSA assays to WHO standards compared with the traditional Hybritech standards // *Clinical Chemistry and Laboratory Medicine (CCLM)*. – 2008. – Т. 46. – №. 5. – С. 623–629.
15. Metz C. E. Basic principles of ROC analysis // *Seminars in nuclear medicine*. – WB Saunders, 1978. – Т. 8. – №. 4. – С. 283–298.
16. Kim D. S. et al. Detecting random responders with infrequency scales using an error-balancing threshold // *Behavior research methods*. – 2018. – Т. 50. – №. 5. – С. 1960–1970.

Method of monotonic functions and its application to the choice of the type of surgery in patients with prostate cancer according to preoperative data

V.G. Gitis, A.B. Derendyaev, K.N. Petrov, E.F. Yurkov, S.A. Pirogov, N.S. Sergeeva,
B.Ya. Alekseev, A.D. Kaprin

We propose a new machine learning method for solving binary classification problems with ordered classes. The method uses a model in which the decision rule is sought in the class of monotone functions on the feature space. The method is used to solve the problem of classifying patients with indolent and aggressive forms of prostate cancer according to pre-treatment data.

KEYWORDS: machine learning, monotonic function method, likelihood ratio, prostate cancer.