

Л.С. Файнзильберг, Т.Н. Жук

## Гарантированная оценка эффективности диагностических тестов на основе усиленного ROC-анализа

Предложен подход к оценке полезности диагностических методов (тестов) с использованием теории статистических решений. Доказаны формальные математические условия, при которых ожидаемые средние потери теста гарантированно меньше, чем потери классификатора, оперирующего только априорными распределениями классов. Эти условия позволили усилить традиционный ROC-анализ путем выделения соответствующей области ROC-кривой, в которой тест заведомо полезен.

An approach to the estimation of the usefulness of diagnostic methods (tests) with the use of the statistical decisions theory is suggested. The formal mathematical terms are proved under which the expected average costs of the test are assured less than costs of a classifier which operates only by a priori distributions of classes. These terms allowed the strengthening of a traditional ROC-analysis by a selection of the proper area of ROC-curve in which a test is a priori useful.

Запропоновано підхід до оцінки корисності діагностичних методів (тестів) з застосуванням теорії статистичних рішень. Доведено формальні математичні умови, за яких очікувані середні втрати тесту гарантовано менші за втрати класифікатора, що оперує лише априорними розподілами класів. Ці умови дозволили посилити традиційний ROC-аналіз шляхом виділення відповідної області ROC-кривої, в якій тест заздалегідь корисний.

**Введение.** Эффективность алгоритмов оценки текущего состояния объектов – одна из основных характеристик компьютерных систем, обеспечивающих решение задач медицинской и технической диагностики [1–3]. При решении задач в статистической постановке традиционно считается, что диагностический алгоритм (тест) эффективен, если он основан на измерении совокупности диагностических признаков, имеющих различные условные распределения вероятностей при разных состояниях объекта. Согласно определению Бен-Бассата [4] такие признаки называются релевантными. В работах [5, 6] было показано, что если состояния объекта неравновероятны, то релевантные признаки могут оказаться бесполезными для решения задач диагностики таких состояний.

Удобным средством оценки эффективности диагностического теста является метод, основанный на анализе так называемой операционной характеристической кривой (*ROC – Receiver Operating Characteristic curve*), широко известный в научных кругах, особенно за рубежом [7–9]. Традиционный ROC-анализ предусматривает сравнение операционных характеристик теста – чувствительности и специ-

фичности [10–18]. Обычно в качестве интегральной характеристики для оценки эффективности теста используется площадь под ROC-кривой [19–21].

В работах [9, 22] предложено развитие ROC-анализа для случая, когда выбор решения осуществляется из более чем двух вариантов (диагнозов). Усовершенствование ROC-анализа для выбора теста с точки зрения ожидаемых потерь от принятия ошибочного решения представлено в [23, 24], а в [25] показано практическое применение этого метода. Известен также метод оценки эффективности теста, основанный на анализе кривых потерь (*cost curves*), который в работах [26, 27] рассматривается как альтернатива ROC-анализу.

Задача настоящего исследования – провести дальнейшее усовершенствование ROC-анализа, для обеспечения выбора диагностического теста, гарантированно уменьшающего априорный риск ошибочной диагностики.

### Традиционный ROC-анализ

Первоначально ROC-анализ использовался для отображения соотношения между долей (процентом) обнаружения цели с помощью радаров и долей (процентом) ложной тревоги

[10, 11]. Позднее ROC-анализ нашел применение в диагностических системах [12] и в машинном обучении [13].

Поясним особенности этого метода на примере решения задачи медицинской диагностики (задачи скрининга), которая сводится к выявлению больных (класс  $V_1$ ) в большой группе обследуемых, где есть и здоровые люди (класс  $V_2$ ). При этом подразумевается, что класс  $V_2$  включает в себя не абсолютно здоровых людей, а тех, у кого не наблюдается рассматриваемого заболевания.

Пусть для выявления больных предполагается использовать диагностический тест. В данном случае нас не будет интересовать то, как построен алгоритм диагностики, т.е. будем рассматривать тест как «черный ящик», который на основании доступной информации принимает одно из двух решений (диагнозов): положительное  $\delta = 1$  («Болен») или отрицательное  $\delta = 2$  («Здоров»).

Для оценки диагностической эффективности теста проводится его «экзамен» на некоторой репрезентативной группе пациентов с заранее известными диагнозами. Результаты такого экзамена можно представить в виде  $2 \times 2$  матрицы (таблица).

Результаты тестирования репрезентативной группы

Истинный диагноз	Результат тестирования	
	Решение – «Болен» $\delta = 1$	Решение – «Здоров» $\delta = 2$
Класс $V_1$ – «Болен»	$TP$	$FN$
Класс $V_2$ – «Здоров»	$FP$	$TN$

В таблице приняты следующие обозначения:

$TP$  (*True Positive*) – число правильных диагнозов «Болен» (истинноположительный результат);

$TN$  (*True Negative*) – число правильных диагнозов «Здоров» (истинноотрицательный результат);

$FP$  (*False Positive*) – число здоровых, ошибочно признанных больными (ложноположительный результат или ошибка ложной тревоги);

$FN$  (*False Negative*) – число больных, ошибочно признанных здоровыми (ложноотрицательный результат или ошибка пропуски цели).

По данным таблицы легко оценить принятые в медицинской диагностике операционные характеристики теста – чувствительность и специфичность [1].

*Чувствительность* (*Sensitivity*)  $S_E$  определяется долей (процентом) предварительно верифицированных больных, которые тестом признаны больными (класс  $V_1$ )

$$S_E = \frac{TP}{TP + FN}, \quad (1)$$

а *специфичность* (*Specificity*)  $S_P$  определяется долей (процентом) здоровых, которые тестами отнесены к группе здоровых (класс  $V_2$ )

$$S_P = \frac{TN}{TN + FP}. \quad (2)$$

Имея эти характеристики, можно представить результаты экзамена в двумерном ROC-пространстве, в котором по оси ординат откладываются значения  $S_E$ , а по оси абсцисс – значения  $1 - S_P$  (рис. 1). Таким образом, диагностический тест (бинарный классификатор) с фиксированными операционными характеристиками (1) и (2) отображается точкой в ROC-пространстве (рис. 1).

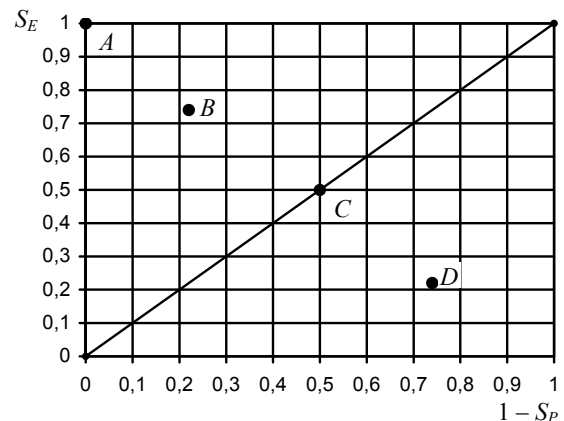


Рис. 1. Бинарные тесты в ROC-пространстве

ROC-пространство дает наглядное графическое представление о диагностической ценности теста и позволяет сравнивать эффективность различных тестов.

*Идеальный* тест (тест  $A$  на рис. 1) расположен в точке с координатами  $(0, 1)$  ROC-пространства. Такой тест всегда принимает истин-

ноположительный результат без ошибок ложной тревоги, а значит, относит всех больных к классу больных, не причисляя к ним здоровых. Отсюда следует, что диагностически ценные тесты находятся в левом верхнем углу  $ROC$ -пространства: чем ближе точка, определяемая операционными характеристиками  $S_E$  и  $S_P$ , к точке  $A$ , тем эффективнее тест.

Тесты, «расположенные» в левом нижнем углу  $ROC$ -пространства, – «консервативны», поскольку при малом проценте ошибок ложной тревоги имеют низкую чувствительность. Понятно, что если консервативный тест, имеющий высокое значение специфичности, принял решение «Болен», то к такому решению следует относиться с высокой степенью доверия. В то же время решение «Здоров» может быть и ошибочным из-за низкой чувствительности.

Тесты, «расположенные» в правом верхнем углу, «либеральны», так как при больших значениях истинноположительных результатов дают также большой процент ошибок ложной тревоги. Поскольку такие тесты имеют высокое значение чувствительности, их решение «Здоров», скорее всего, правильно, в то время как противоположное решение «Болен» может быть ошибочным из-за низкой специфичности.

Тесты, «расположенные» на диагонали  $S_E = 1 - S_P$   $ROC$ -пространства (например, тест  $C$  на рис. 1), не дают дополнительную диагностическую информацию о классах, а потому бесполезны. Такие тесты эквивалентны классификаторам, использующим стратегию «случайного угадывания» классов.

Тесты, «расположенные» ниже диагонали (например, тест  $D$  на рис. 1), еще хуже тех, что используют стратегию случайного угадывания. Однако такой тест легко «переводится» в точку  $B$ , симметричную относительно диагонали, заменив стратегию принятия решений на противоположную [14, 15]. В этом случае значение истинноположительного результата классификации становится значением ошибки ложной тревоги и наоборот.

При решении задач медицинской диагностики часто используют пороговые решающие правила [16], например, правило вида

принимает решение «Болен», если  $x > x_0$ ,  
принимает решение «Здоров», если  $x \leq x_0$ , (3)  
где  $x$  – некоторый измеряемый параметр (диагностический признак),  $x_0$  – пороговое значение.

В таком случае появляется возможность оценить чувствительность и специфичность теста при различных значениях порога  $x_0$ , что соответствует последовательности точек бинарного классификатора в  $ROC$ -пространстве (рис. 2).

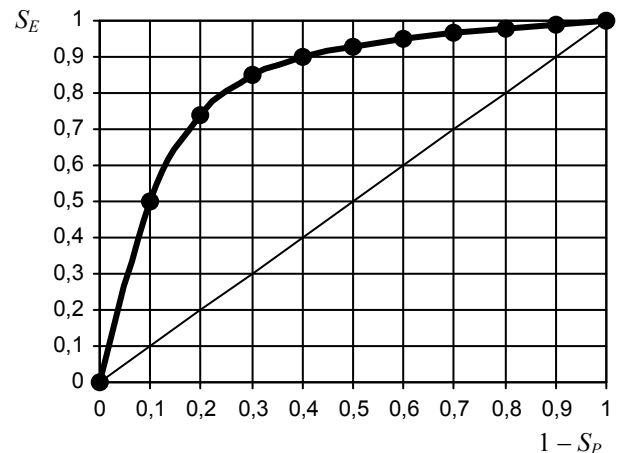


Рис. 2.  $ROC$ -кривая теста, основанного на пороговом решающем правиле – отсортировать наблюдаемые значения признака  $x$  в порядке убывания;

Из правила (3) следует, что если при фиксированном пороговом значении  $x_0$  пациент отнесен к классу «Здоров», то для любого более низкого значения признака  $x$  пациент также будет отнесен к этому классу. Отсюда вытекает следующий простой алгоритм построения экспериментальной  $ROC$ -кривой [14]:

- отсортировать наблюдаемые значения признака  $x$  в порядке убывания;
- последовательно уменьшать (с некоторым шагом) порог  $x_0$ , перемещаясь вниз по списку отсортированных значений  $x$ , и по экзаменационной выборке наблюдений вычислять соответствующие пары значений  $S_E$  и  $S_P$ ;
- отобразить полученные таким образом последовательности пар значений  $S_E$  и  $S_P$  в  $ROC$ -пространстве.

Обозначим  $p(x/V_1)$  и  $p(x/V_2)$  – условные распределения вероятностей признака  $x$  в клас-

сах  $V_1$  и  $V_2$ . Множество  $X_1 = \{x : p(x/V_1) \neq 0\}$  будем называть носителем распределения диагностического признака  $x$  для класса  $V_1$ , а множество  $X_2 = \{x : p(x/V_2) \neq 0\}$  – носителем распределения  $x$  для класса  $V_2$ .

Принципиально возможны четыре ситуации, характеризующие топологию носителей  $X_1$ ,  $X_2$  (рис. 3) и соответствующие этим ситуациям формы  $ROC$ -кривых (рис. 4).

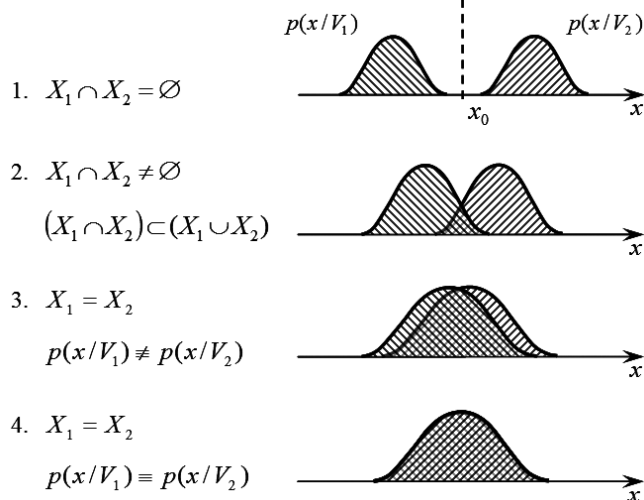


Рис. 3. Топология носителей классов

*Ситуация 1* представляет собой случай, когда множества  $X_1$  и  $X_2$  не пересекаются. В этом случае может быть установлен порог  $x_0$ , полностью разделяющий значения диагностического признака  $x$  для здоровых и больных пациентов, а значит, правило (3) определит идеальный тест. График соответствующей  $ROC$ -кривой (рис. 4, кривая 1) проходит от точки с координатами (0,0) в точку с координатами (0,1) и далее в точку с координатами (1,1).

В *ситуации 2* множества  $X_1$  и  $X_2$  частично пересекаются, и соответствующая  $ROC$ -кривая имеет форму кривой 2 (рис. 4).

В *ситуации 3*, когда множества  $X_1$  и  $X_2$  совпадают, но условные распределения  $p(x/V_1)$  и  $p(x/V_2)$  различны, соответствующая  $ROC$ -кривая имеет форму 3 (рис. 4). Именно такая форма чаще всего присуща большинству реальных диагностических тестов.

И, наконец, в *ситуации 4*, когда совпадают не только множества  $X_1$  и  $X_2$ , но и условные

распределения  $p(x/V_1)$  и  $p(x/V_2)$ , тест становится бесполезным и соответствующая  $ROC$ -кривая лежит на диагонали (рис. 4, кривая 4).

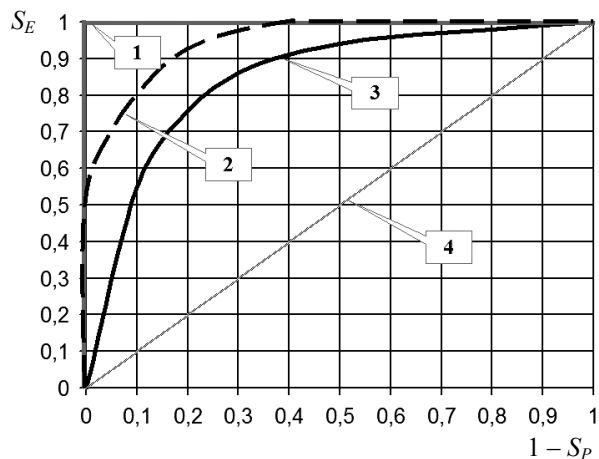


Рис. 4. Взаимосвязь  $ROC$ -кривых с топологией носителей классов

Заметим, что для операционных характеристик тестов, лежащих выше диагонали  $ROC$ -пространства, выполняется условие

$$S_E > 1 - S_P, \quad (4)$$

тогда как для бесполезных тестов справедливо соотношение  $S_E = 1 - S_P$ .

Для сравнения эффективности различных диагностических тестов удобно использовать некоторую интегральную характеристику. В качестве такой характеристики часто используют площадь  $AUC$  (Area Under Curve), вычисленную под  $ROC$ -кривой.

Очевидно, что величина  $AUC$  определяет среднюю чувствительность теста  $\bar{S}_E$  при возможных значениях специфичности  $0 \leq S_P \leq 1$  или среднюю специфичность теста  $\bar{S}_P$  при возможных значениях чувствительности  $0 \leq S_E \leq 1$ , причем  $AUC = 1$  для идеального теста и  $AUC = 0,5$  для бесполезного теста, операционные характеристики которого не удовлетворяют условию (4). Поэтому считается, что чем больше  $AUC$  приближается к единице, тем эффективнее тест [19].

Площадь под  $ROC$ -кривой может быть вычислена с помощью правила трапеций [19],  $U$ -критерия Манна-Уитни [20] или коэффициента Джини [21].

ROC-анализ используют для нахождения оптимального значения порога  $x_0$ , или *точки отсечения* (*Cut-off value*). Понятно, что оптимальное значение  $x_0$  зависит от предъявляемых к тесту требований, т.е. от выбранного критерия оптимальности [16]. Например, в работе [17] оптимальной предлагается считать точку на ROC-кривой, которая имеет наибольшее значение чувствительности при наименьшем значении доли ложноположительных результатов. Эта точка может быть найдена с помощью индекса Йодена ( $J$  – *Youden index*) как максимальное отличие ординат ROC-кривой и диагонали (рис. 5) [18]:

$$J = \max(S_E + S_P - 1). \quad (5)$$

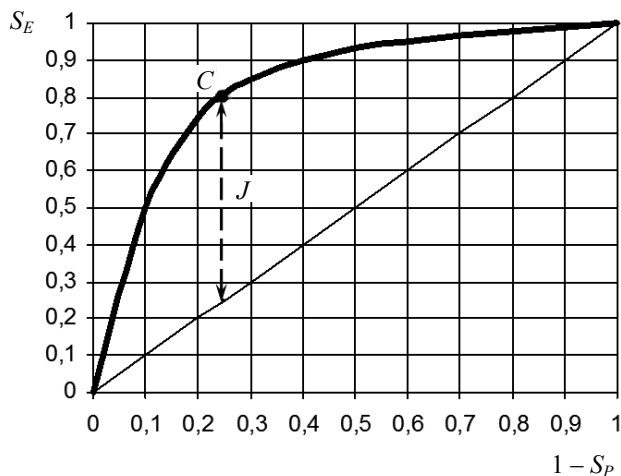


Рис. 5. Оптимальная точка отсечения  $C$  (*cut-off value*)

Итак, рассмотрены некоторые особенности традиционного ROC-анализа для двух классов. В общем случае число возможных диагнозов  $M \geq 2$ . Очевидно, что для каждого истинного диагноза  $V_1, \dots, V_M$  диагностический тест может принять одно из  $M$  решений, но только одно из них будет правильным. Отсюда следует, что в общем случае существует  $N_M = M^2 - M$  возможных ошибок диагностического теста.

В рассмотренном случае ( $M = 2$ ) число ошибок  $N_2 = 2$ , при числе классов  $M = 3$  число возможных ошибок  $N_3 = 6$  и т.д. Следовательно, уже при  $M = 3$  ROC-анализ становится достаточно громоздким и не столь наглядным: вместо двумерной ROC-кривой требуется проводить анализ шестимерной ROC-поверхности.

Но можно пойти и другим путем, последовательно решая задачу распознавания одного класса  $V_i \in \{V_1, \dots, V_M\}$  и всех остальных. При таком подходе, вместо  $N_M$ -мерной ROC-поверхности достаточно построить и проанализировать  $M$  обычных двумерных ROC-кривых, отображающих взаимосвязь истинноположительных и ложноположительных результатов, полученных при распознавании каждого класса  $V_i \in \{V_1, \dots, V_M\}$  и объединенного класса  $\bar{V}_i = \bigcup_{j \neq i} V_j, \quad i, j = 1, \dots, M.$

### ROC-анализ с позиции теории статистических решений

Рассмотрим теперь некоторые усовершенствования традиционного ROC-анализа, позволяющие оценивать эффективность диагностического теста с точки зрения ожидаемых потерь (среднего риска).

Пусть  $P(V_1)$  – вероятность появления больного в группе обследуемых (преваленс заболевания). Будем характеризовать возможные ошибки диагностического теста величинами  $L_{12}, L_{21}$ , определяющими плату (потери) за одну ошибку пропуска цели (невывявление больного) и ложной тревоги (отнесение здорового пациента к группе больных).

Тогда ожидаемые потери (средний риск) от неправильной диагностики определит соотношение [23, 29]

$$R = P(V_1) \cdot (1 - S_E) \cdot L_{12} + [1 - P(V_1)] \cdot (1 - S_P) \cdot L_{21}. \quad (6)$$

Из (6) непосредственно следует, что два теста, имеющие операционные характеристики  $S_E^{(1)}, S_P^{(1)}$  и  $S_E^{(2)}, S_P^{(2)}$ , будут иметь одинаковые потери, если выполняется условие [24]

$$\frac{S_E^{(2)} - S_E^{(1)}}{(1 - S_P^{(2)}) - (1 - S_P^{(1)})} = m, \quad (7)$$

где

$$m = \frac{L_{21} [1 - P(V_1)]}{L_{12} P(V_1)}. \quad (8)$$

Семейство параллельных прямых  $S_E = m(1 - S_P) + \text{const}$  с положительным угловым коэф-

фициентом  $m$ , зависящим от  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$ , образует в  $ROC$ -пространстве *изолинии эффективности* (*Iso-performance lines*) [24]. Все тесты, соответствующие точкам на изолинии, имеют одни и те же ожидаемые потери, причем, чем больше значение пересечения изолинии с осью ординат  $S_E$ , тем меньше потери. Поэтому, если изолинию эффективности провести как касательную к  $ROC$ -кривой, то в точке (или на отрезке) касания изолинии и  $ROC$ -кривой потери будут минимальными [16].

Для выбора оптимального теста, минимизирующего ожидаемые потери, в работах [23, 24] предложен так называемый *метод выпуклой ROC-оболочки* (*ROC convex hull method*), суть которого состоит в следующем:

- определяются операционные характеристики каждого  $i$ -го бинарного теста в  $ROC$ -пространстве и отображаются точки с координатами  $S_E[i]$  и  $1 - S_P[i]$ , а для каждого  $j$ -го теста, основанного на пороговом решающем правиле (3), последовательно изменяется порог  $x_0^{(j)}$  и отображается множество пар  $S_E^{(j)}[i]$ ,  $1 - S_P^{(j)}[i]$  точек  $j$ -й  $ROC$ -кривой;

- строится выпуклая оболочка множества точек, что для  $n$  точек может быть выполнено с помощью алгоритма *QuickHull* за время  $O[n \log(n)]$  [28];

- по формуле (8) вычисляется угловой коэффициент  $m$  наклона изолинии эффективности и отображается в  $ROC$ -пространстве семейство изолиний эффективности, соответствующих заданным значениям  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$ ;

- точка (или их совокупность) касания изолинии эффективности к построенной выпуклой оболочке определит операционные характеристики  $S_E^{opt}$  и  $S_P^{opt}$  наиболее эффективного теста, поскольку именно такой тест ближе всего к «идеальному» тесту с координатами  $1 - S_P = 0$  и  $S_E = 1$ .

Для иллюстрации приведенного алгоритма на рис. 6 представлены примеры экспериментальных  $ROC$ -кривых трех различных тестов и соответствующая выпуклая оболочка (отмечена пунктирной линией). Различные заданные

значения  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$  определяют наклоны трех изолиний эффективности с угловыми коэффициентами  $m_1$ ,  $m_2$  и  $m_3$ .

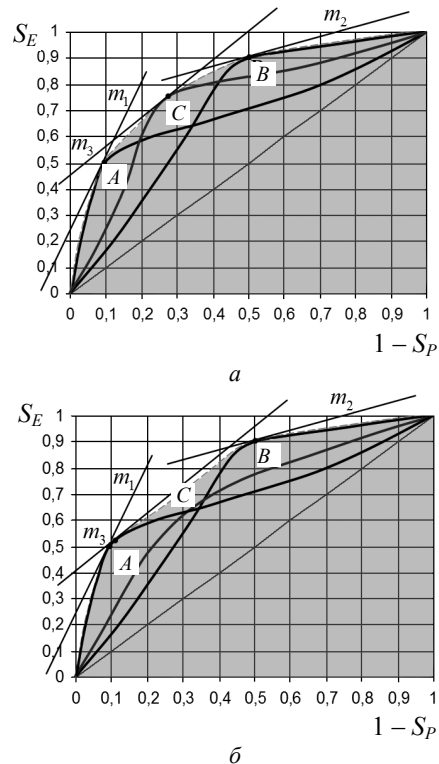


Рис. 6. Иллюстрация метода выпуклой  $ROC$ -оболочки

На рис. 6, *а*  $ROC$ -кривые всех трех тестов «касаются» выпуклой оболочке. Поэтому при условии  $m = m_1$  потенциально оптимальным будет тест  $A$ , при  $m = m_2$  – тест  $B$ , а при  $m = m_3$  – тест  $C$ .

На рис. 6, *б*  $ROC$ -кривая теста  $C$  не «касается» выпуклой оболочке. Поэтому в такой ситуации для условия  $m = m_3$  потенциально оптимальным с точки зрения ожидаемых потерь будет уже не тест  $C$ , как в предыдущем случае, а тест  $A$ .

Еще один пример показан на рис. 7. При условии  $m = m_1$  потенциально оптимальным будет тест  $A$ , при  $m = m_2$  – тест  $B$ , а в случае  $m = m_3$  оба теста  $A$  и  $B$  не обеспечивают заданных ожидаемых потерь, поскольку не существует точек касания соответствующей изолинии эффективности с выпуклой оболочкой.

Понятно, что после того, как для некоторой совокупности тестов построена выпуклая обо-

лочка, исследование эффективности «нового» теста предполагает перепостроение выпуклой оболочки. Однако из добавляемых тестов потенциально оптимальными могут быть лишь те, которые «расширяют» выпуклую оболочку.

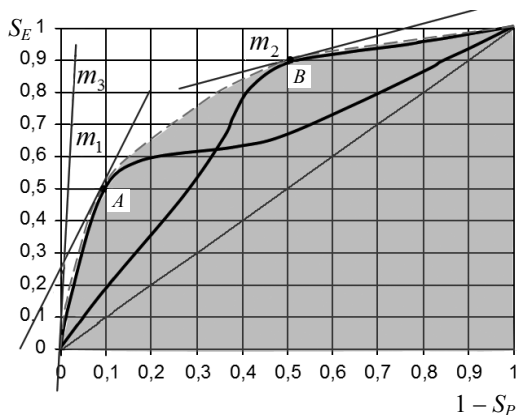


Рис. 7. Иллюстрация метода выпуклой ROC-оболочки

Например, добавим к выпуклой оболочке, построенной по двум ROC-кривым, три бинарных теста  $C$ ,  $D$  и  $E$  (рис. 8). Очевидно, что при условии  $m \in [m_1, m_2]$  оптимальными могут быть только тесты, находящиеся на ломаной  $ACB$ . Следовательно, тест  $C$  может быть оптимальным, поскольку он расширяет выпуклую ROC-оболочку, в то время как тесты  $D$  и  $E$  не являются потенциально оптимальными.

#### Достаточные условия полезности диагностического теста

Условие (6) позволяет по заданным значениям  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$  вычислить ожидаемый средний риск ошибочной диагностики (апостериорный риск) при использовании теста с фиксированными операционными характеристиками  $S_E$  и  $S_p$ . Возникает естественный вопрос: всегда ли такой тест полезен?

Для того чтобы ответить на этот вопрос, прежде всего нужно сформулировать определение «полезность» диагностического теста. По мнению авторов, таким определением может быть условие того, что тест обеспечивает средние потери меньше тех, которые наблюдались бы без привлечения диагностического теста [29]. Другими словами, понятие полезности разумно связать с таким определением.

**Определение 1.** При заданных значениях  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$  диагностический тест полезен, если выполняется *строгое* неравенство

$$R < R_0, \quad (9),$$

т.е. апостериорный риск  $R$  строго меньше априорного риска  $R_0$ , реализующего стратегию принятия оптимальных решений лишь на основании значений  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$ .

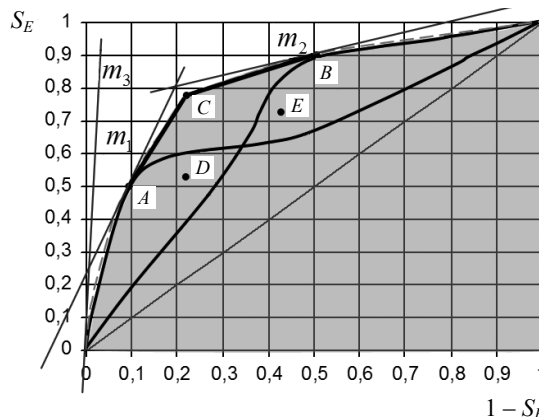


Рис. 8. Расширение выпуклой ROC-оболочки

Здесь нужны пояснения. Понятно, что если алгоритм диагностики построен формальным путем и реализует байесовскую стратегию принятия решений [30], обеспечивающую минимизацию апостериорного риска  $R$ , то заведомо выполняется условие (9) или условие  $R = R_0$ . Иными словами, формально построенный алгоритм диагностики не может быть «вредным». Однако на практике часто отсутствует полная информация об условных распределениях вероятностей признаков, необходимая для реализации байесовской стратегии, либо тест построен не формальным путем, а на основе «здорового» смысла и интуиции конструктора диагностической системы.

В этих существенных для практического применения случаях при фиксированных  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$  и известных  $S_E$  и  $S_p$  вместо условия (9) может выполняться на первый взгляд «парадоксальное» условие  $R > R_0$ . Поскольку численные примеры довольно часто более убедительны, чем формальные выкладки, продемонстрируем такую ситуацию на следующем примере.

**Модельный пример.** Пусть для выявления больных в группе из 10 тыс. чел. предполага-

ется использовать диагностический тест, высокие операционные характеристики которого  $S_E = 0,9$  и  $S_P = 0,9$  были определены на репрезентативной выборке наблюдений с известными диагнозами. Пусть известен prevalence заболевания  $P(V_1) = 0,02$ , а платы за ошибки пропуска цели и ложной тревоги приняты равными  $L_{12} = L_{21} = 1$ .

На рис. 9 представлены ожидаемые результаты тестирования. Если не использовать диагностический тест, то оптимальная «стратегия», минимизирующая априорный риск  $R_0$ , должна быть такой: всех обследуемых признавать здоровыми (в противном случае потери будут еще большими!). В этом случае средние (на одного обследуемого) априорные потери, связанные с тем, что 200 больных не будут выявлены, составят

$$R_0 = (L_{12} \cdot 200) / 10000 = (1 \cdot 200) / 10000 = 0,02. (10)$$

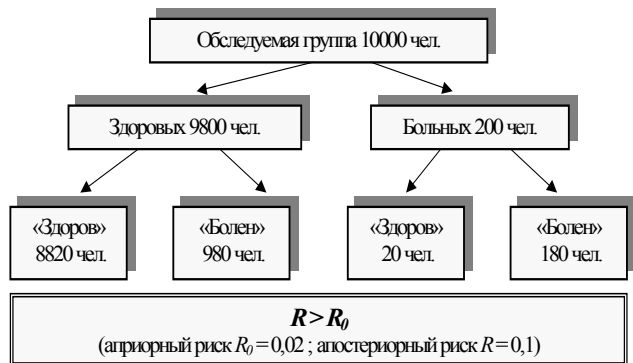


Рис. 9. Ожидаемые результаты тестирования

При использовании диагностического теста с чувствительностью  $S_E = 0,9$  и специфичностью  $S_P = 0,9$  средние потери, связанные с ложноположительными результатами диагностики 980 здоровых пациентов и невыявлением 20 больных, составят

$$R = (L_{21} \cdot 980 + L_{12} \cdot 20) / 10000 = (1 \cdot 980 + 1 \cdot 20) / 10000 = 0,1. (11)$$

Следовательно,  $R \gg R_0$  и такой тест нельзя признать полезным, поскольку при заданных платах за ошибки пропуска цели и ложной тревоги он только увеличивает ожидаемые потери.

Несмотря на кажущуюся «парадоксальность» того, что  $R \gg R_0$ , полученный результат име-

ет вполне обоснованное пояснение, приведенное в работе [29]: тест не является оптимальным по байесовской стратегии. При указанных  $P(V_1)$ ,  $L_{12}$ ,  $L_{21}$ ,  $S_E$  и  $S_P$  решения «Болен» являются необоснованными с точки зрения минимизации апостериорного риска, а значит, для любого обследуемого всегда должен приниматься диагноз «Здоров», совпадающий с априорными решениями.

Перейдем теперь к формальной стороне вопроса и покажем, что условие (4) является необходимым, но не достаточным условием, гарантирующим выполнение строгого неравенства (9).

**Утверждение 1.** Тест гарантированно полезен по Определению 1, если чувствительность  $S_E$  и специфичность  $S_P$  связаны соотношением

$$S_E > m(1 - S_P), \text{ при } m \geq 1 (12)$$

или

$$S_E > 1 - m + m(1 - S_P), \text{ при } m < 1, (13)$$

где величина  $m$ , зависящая от  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$ , определяется соотношением (8).

**Доказательство.** Если не использовать диагностический тест и принимать решения только на основании значений  $P(V_1)$ ,  $L_{12}$  и  $L_{21}$ , то априорная «стратегия» сводится к одному из двух вариантов: либо всех обследуемых отнести к классу больных  $V_1$ , и тогда априорный риск равен

$$R_0^+ = L_{21}[1 - P(V_1)], (14)$$

либо принимать решение о том, что каждый обследуемый здоров, и тогда априорный риск равен

$$R_0^- = L_{12}P(V_1). (15)$$

Из (14) и (15) с учетом (8) следует, что  $R_0^- < R_0^+$ , при  $m > 1$  и  $R_0^- > R_0^+$ , при  $m < 1$  и, естественно,  $R_0^- = R_0^+$ , когда  $m = 1$ .

Следовательно, минимальный априорный риск  $R_0$ , фигурирующий в правой части неравенства (9), можно записать так:

$$R_0 = \begin{cases} L_{12}P(V_1), & \text{при } m \geq 1, \\ L_{21}[1 - P(V_1)], & \text{при } m < 1. \end{cases} (16)$$



Подстановка выражений (6) и (16) в неравенство (9) после простых преобразований приводит к соотношениям (12) и (13). Утверждение доказано.

Условия (12) и (13) позволяют обоснованно подходить к оценке пригодности нового диагностического теста с известными операционными характеристиками  $S_E$  и  $S_p$  для диагностики заболевания с преваленсом  $P(V_1)$  и заданным соотношением потерь  $\omega = L_{12}L_{21}^{-1}$  от ошибок пропуска цели и ложной тревоги. Для такой оценки (в отличие от метода выпуклой ROC-оболочки), вообще говоря, не требуется привлекать визуальный анализ результатов в ROC-пространстве. Например, формальная подстановка данных рассмотренного модельного примера в условие (12) показывает, что тест заведомо бесполезен.

В то же время на основе Утверждения 1 можно усилить традиционный ROC-анализ, ограничив соответствующую его область ROC-кривой.

Рис. 10 иллюстрирует предлагаемый подход для случаев  $m \geq 1$  (рис. 10,а) и  $m < 1$  (см. рис. 2,б). Серым цветом показаны области гарантированной полезности теста. В соответствии с (12) прямая, ограничивающая область полезности теста для случаев  $m \geq 1$ , проходит через точку с координатами (0,0) и точку с координатами  $(1/m, 1)$ , а для случая  $m < 1$  – через точку с координатами (1,1) и точку с координатами  $(0, 1-m)$ .

Отрезок  $OA$  (см. рис. 10, а) соответствует случаю, когда  $P(V_1) = 0,15$  и  $\omega = L_{12}L_{21}^{-1} = 1$ . Видно, что в данном случае тест бесполезен, поскольку ROC-кривая расположена ниже отрезка  $OA$ . Но если допустить, что потери от пропуска цели в четыре раза выше потерь от ложной тревоги, т.е. допустить, что  $L_{12}L_{21}^{-1} = 4$ , то соответствующая прямая  $OB$  уже пересекает ROC-кривую. Следовательно, тест, имеющий чувствительность  $S_E = 62,5\%$  и специфичность  $S_p = 80\%$ , окажется гарантированно полезным с учетом строгого неравенства (9), поскольку точка  $C$ , соответствующая таким операцион-

ным характеристикам, расположена на допустимой части ROC-кривой.

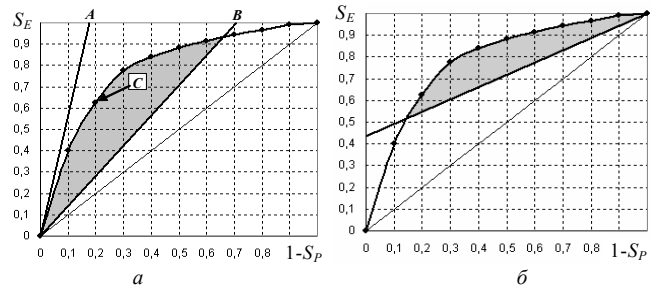


Рис. 10. Иллюстрация идеи усиленного ROC-анализа для случаев:  $a - m \geq 1$ ,  $b - m < 1$

Заметим также, что условия (12) и (13) позволяют решить и обратную задачу: определить допустимый интервал соотношения потерь  $\omega = L_{12}L_{21}^{-1}$ , при котором тест с чувствительностью  $S_E$  и специфичностью  $S_p$  можно считать полезным для скрининга заболевания с известным преваленсом  $P(V_1)$ .

Этот интервал определяется формулой

$$\frac{1 - P(V_1)}{P(V_1)} \frac{1 - S_p}{S_E} \leq \omega \leq \frac{1 - P(V_1)}{P(V_1)} \frac{S_p}{1 - S_E}. \quad (17)$$

Например, тест, имеющий чувствительность  $S_E = 62,5\%$  и специфичность  $S_p = 80\%$ , можно рекомендовать для выявления заболеваний с преваленсом  $P(V_1) = 0,15$ , если допустимо соотношение потерь от ложноотрицательных и ложноположительных ошибок в пределах  $2,66 \leq \omega \leq 17,7$ .

Примечательно, что полученные результаты можно сопоставить с изолиниями эффективности, упомянутых при обсуждении метода выпуклой ROC-оболочки. Действительно, из выражения (6) следует, что уравнение изолинии с ожидаемыми потерями  $R = R^*$  имеет вид

$$S_E = m(1 - S_p) + 1 - \frac{R^*}{L_{12}P(V_1)}. \quad (18)$$

Из сравнения (18) с (12) и (13) заключаем, что изолинии эффективности (18) совпадет с прямыми, ограничивающими области полезных тестов, если выполняется условие

$$R^* = \begin{cases} L_{12}P(V_1), & \text{при } m \geq 1, \\ L_{21}[1 - P(V_1)], & \text{при } m < 1, \end{cases} \quad (19)$$

правая часть которого совпадает с правой частью (16).

Таким образом, область полезных тестов ограничивается одной изолинией эффективности, соответствующей уровню *априорного* риска  $R_0$ , которая *непрерывно* проходит через точку с координатами (0,0) при  $m \geq 1$  или точку с координатами (1,1) при  $m < 1$ , причем, если такая изолиния

- пересекает ROC-кривую, то  $R < R_0$ ,
- касается ROC-кривой, то  $R = R_0$ ,
- не пересекает ROC-кривую, то  $R > R_0$ .

### Практическое применение результатов

Усиленный ROC-анализ придает формальной процедуре оценки эффективности диагностического теста наглядную интерпретацию в виде когнитивного графического образа. Предложенный подход реализован в виде инструментального вычислительного модуля (рис. 11), который позволяет:

- автоматически строить экспериментальную ROC-кривую любого из диагностических признаков, хранимых в базе данных, с возможностью «фильтрации» записей по определенным критериям;
- отобразить допустимую область ROC-пространства, удовлетворяющую условиям (12) и (13);
- вычислить допустимый интервал соотношения потерь по соотношению (17).

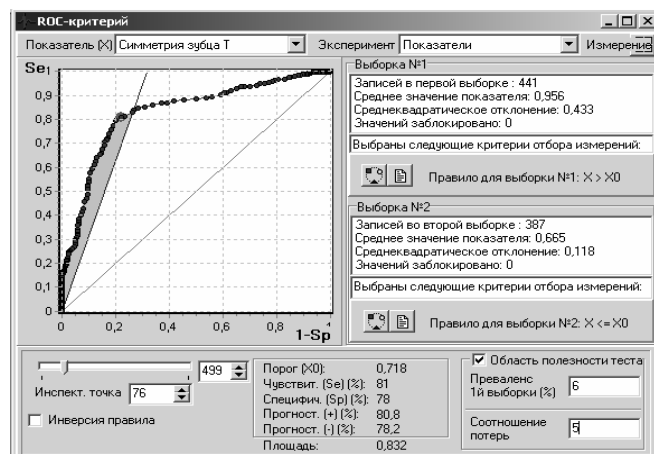


Рис. 11. Интерфейс вычислительного модуля, реализующего усиленный ROC-анализ

Модуль включен в состав информационной технологии **ФАЗАГРАФ<sup>TM</sup>** и позволил обосновать выбор эффективного диагностического теста, обеспечивающего выявление начальных признаков ишемической болезни по фазовому портрету ЭКГ [31].

**Заключение.** Таким образом доказано, что условия (12) и (13) позволяют обоснованно подходить к оценке пригодности нового диагностического теста с известной чувствительностью и специфичностью для диагностики заболевания с преваляенсом  $P(V_1)$  и заданным соотношением потерь  $\omega = L_{12}L_{21}^{-1}$  от ошибок пропуска цели и ложной тревоги. Полученные условия позволили усилить традиционный ROC-анализ путем соответствующего ограничения области ROC-кривой. Предложенный подход нашел практическое применение при решении актуальной задачи выявления начальных признаков ишемической болезни сердца.

1. Власов В.В. Эффективность диагностических исследований. – М.: Медицина, 1988. – 256 с.
2. Ластед Л. Введение в проблему принятия решений в медицине. – М.: Мир, 1971. – 282 с.
3. Бургер И.А. Техническая диагностика. – М.: Машиностроение, 1978. – 240 с.
4. Ben-Bassat M. Irrelevant features in pattern recognition // IEEE Trans. Comp. – 1978. – C-27, N 8. – P. 749–766.
5. Fainzilberg L.S. Why Relevant Features May Be Unuseful in Statistical Recognition of Two Classes // Proc. of the 13th Intern. Conf. on Pat. Recog. (ICPR'96). – Viena (Austria). – 1996. – P. 730–734.
6. Файнзилберг Л.С. Оценка полезности признаков при решении задач диагностики в статистической постановке // Математические машины и систем. – 1998. – № 1. – С. 57–64.
7. Metz C.E. Fundamental ROC analysis // Progress in Medical Physics and Psychophysics Handbook of Medical Imaging. – SPIE Press, Bellingham WA. – 2000. – 1. – P. 754–769.
8. Fawcett T. Using Rule Sets to Maximize ROC Performance // Proc. of the IEEE Intern. Conf. on Data Mining (ICDM–2001) // IEEE Com. Soci., Los Alamitos, CA. – 2001. – P. 131–138.
9. Ferri C., Hernández-Orallo J., Salido M.A. Volume Under the ROC Surface for Multiclass Problems. Exact Computation and Evaluation of Approximations // Techn. Rep. DSIC. 2003, Univ. Politèc. València. – P. 36–43.

10. *Green D.M., Swets J.A.* Signal detection theory and psychophysics. – New York: John Wiley and Sons Inc., 1966. – 328 p.
11. *Egan J.P.* Signal detection theory and ROC analysis. – New York: Acad. Press, 1975. – 386 p.
12. *Swets J.A.* Measuring the accuracy of diagnostic systems // *Science*. – 1988. – N 240. – P. 1285–1292.
13. *Spackman K.A.* Signal detection theory: Valuable tools for evaluating inductive learning // *Proc. of the Sixth Intern. Workshop on Machine Learning*. – San Mateo, CA. – 1989. – P. 160–163.
14. *Fawcett T.* ROC Graphs: Notes and Practical Considerations for Researchers – Kluwer Acad. Publ., 2004. – 38 p.
15. *Flach P., Wu S.* Repairing concavities in ROC curves. // *Proc. UK Workshop on Comp. Intel.* – 2003. – P. 38–44.
16. *Zweig M.H., Campbell G.* Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine // *Clinical Chemistry*. – 1993. – **39**, N 4. – P. 561–577.
17. *Youden W.J.* Index for rating diagnostic tests // *Cancer*. – 1950. – N 3. – P. 32–35.
18. *Optimal* Cut-point and Its Corresponding Youden Index to Discriminate Individuals Using Pooled Blood Samples / E.F. Schisterman, N.J. Perkins, A. Liu et al. // *Epidemiology*. – 2005. – **16**. – P. 73–81.
19. *Van den Hout W.B.* The area under an ROC curve with limited information // *Medical Decision Making*. – 2003. – N 23. – P. 160–166.
20. *Hanley J.A., McNeil B.J.* The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve // *Radiology*. – 1982. – N 143. – P. 29–36.
21. *Hand D.J., Till R.J.* A simple generalization of the area under the ROC curve to multiple class classification problems // *Machine Learning*. – 2001. – N 45(2). – P. 171–186.
22. *Srinivasan A.* Note on the Location of Optimal Classifiers in N-dimensional ROC Space // *Techn. Rep. PRG-TR-2-99*, Oxford Univ. Comp. Labor., 1999. – 260 p.
23. *Provost F., Fawcett T.* Analysis and visualization of classifier performance: Comparison under imprecise class and cost distribution // *Proc. of The Third Intern. Conf. on Knowledge Discovery and Data Mining (KDD-97)*. – AAAI Press, 1997. – P. 43–48.
24. *Provost F., Fawcett T.* Robust Classification for Imprecise Environments // *Machine Learning*. – 2001. – **42**, N 3. – P. 203–231.
25. *Bettinger R.* Cost-Sensitive Classifier Selection Using the ROC Convex Hull Method // *Comp. Scie. and Statistics*. – 2003. – N 35. – P. 36–42.
26. *Drummond C., Holte R.* Explicitly representing expected cost: an alternative to ROC representation // *Proc. of Knowledge Discovery and Data Mining*. – 2000. – P. 198–207.
27. *Drummond C., Holte R.* What ROC curves can't do (and cost curves can) // *ROCAI*. – 2004. – P. 19–26.
28. *Barber C., Dobkin D., Huhdanpaa H.* The quickhull algorithm for convex hull. // *Techn. Rep. GCG53*, University of Minnesota. – 1993. – 185 p.
29. *Файнзильберг Л.С.* К вопросу о полезности диагностических методов в задачах скрининга // *УСиМ*. – 2002. – № 6. – С. 10–17.
30. *Шлезингер М., Главач В.* Десять лекций по статистическому и структурному распознаванию. – К.: Наук. думка, 2004. – 545 с.
31. *Чайковский И.А., Файнзильберг Л.С.* Медицинские аспекты применения устройства ФАЗАГРАФ в клинической практике и в домашних условиях. К.: МНУЦИТиС НАН и МОН Украины, 2004. – 74 с.

Поступила 09.06.2009

Тел. для справок: (044) 411-6904, 254-3659 (Киев)

© Л.С. Файнзильберг, Т.Н. Жук, 2009

## Внимание !

**Оформление подписки для желающих  
опубликовать статьи в нашем журнале обязательно.  
В розничную продажу журнал не поступает.  
Подписной индекс 71008**