

УДК 616-07-036.8

УСИЛЕННЫЙ ROC-АНАЛИЗ КАК СРЕДСТВО ОЦЕНКИ ЭФФЕКТИВНОСТИ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ ДИАГНОСТИЧЕСКИХ РЕШЕНИЙ

Л.С. Файнзильберг*, Т.Н. Жук*, Н.А. Семергей*, О.А. Жуковская**

*МНУЦ ИТИС, **НТУУ «КПИ»

e-mail: fainzilberg@voliacable.com

Эффективность компьютерных систем поддержки принятия решений, ориентированных на медицинскую и техническую диагностику, прежде всего связана с эффективностью диагностических алгоритмов (тестов), на основе которых принимаются решения о текущем состоянии объекта. При решении задач диагностики в статистической постановке традиционно считается, что тест эффективен, если он основан на измерении совокупности диагностических признаков, имеющих различные условные распределения при разных состояниях объекта. Согласно определению Бен-Бассата [1] такие признаки называются релевантными. В то же время еще в работах [2, 3] было показано, что если состояния объекта неравновоятны, 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*). Первоначально **ROC**-анализ использовался в теории обнаружения сигналов для отображения соотношения между долей (процентом) обнаружения цели с помощью радаров и долей (процентом) ложной тревоги. Позже **ROC**-анализ нашел применение в диагностических системах и в машинном обучении.

Поясним детали этого метода на примере решения задачи медицинской диагностики (задачи скрининга), которая сводится к выявлению больных людей в большой группе обследуемых, где есть и больные (класс V_1) и здоровые (класс V_2). При этом подразумевается, что класс V_2 включает не абсолютно здоровых людей, а лиц, у которых нет рассматриваемого заболевания.

Пусть для выявления больных предполагается использовать диагностический тест (алгоритм, правило), в соответствии с которым для конкретного пациента из обследуемой группы принимается одно из двух решений (диагнозов): положительное решение $\delta = 1$ («Болен») или отрицательное решение $\delta = 2$ («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 2×2 матрицы (таблицы 1).

Таблица 1

Результаты тестирования репрезентативной группы

ИСТИННЫЙ ДИАГНОЗ	РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ	
	Решение – «Болен» ($\delta=1$)	Решение – «Здоров» ($\delta=2$)
Класс V_1 – «Болен»	<i>TP</i>	<i>FN</i>
Класс V_2 – «Здоров»	<i>FP</i>	<i>TN</i>

В таблице приняты следующие обозначения:

- TP (*True Positive*) – число правильных диагнозов «Болен»;
- TN (*True Negative*) – число правильных диагнозов «Здоров»;
- FP (*False Positive*) – число здоровых, ошибочно признанных больными;
- FN (*False Negative*) – число больных, ошибочно признанных здоровыми.

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

Чувствительность (*sensitivity*) S_E определяется долей (процентом) предварительно верифицированных больных, которые тест признал больными

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

а специфичность (*specificity*) S_P определяется долей (процентом) здоровых, которые тест отнес к группе здоровых

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

При решении задач медицинской диагностики часто используют пороговые решающие правила вида

принимает решение «Болен», если $x > x_0$,

принимает решение «Здоров», если $x \leq x_0$, (3)

где – некоторый измеряемый параметр (диагностический признак), а x_0 – пороговое значение. В этом случае появляется возможность оценить чувствительность и специфичность теста при различных значениях x_0 и отобразить последовательность полученных пар S_E и S_P в *ROC*-пространстве (рис. 1).

Для сравнения эффективности диагностических тестов традиционно используют площадь *AUC* (*Area Under Curve*), вычисленную под *ROC*-кривой, которая 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.

Таким образом, в соответствии с традиционным *ROC*-анализом считается, что диагностический тест эффективен, если *ROC*-кривая лежит выше диагонали, а значит выполняется условие

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

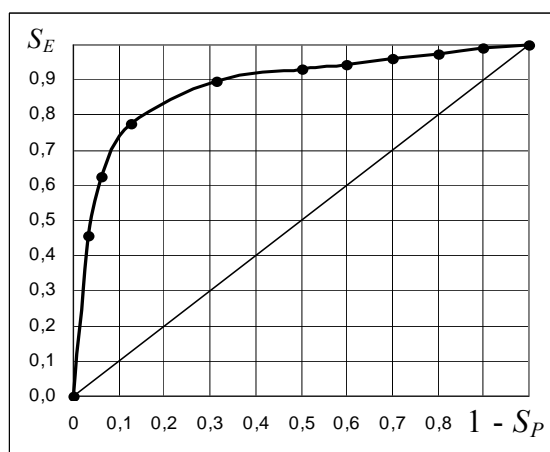


Рис.1 *ROC*-кривая теста, основанного на пороговом решающем правиле

Покажем, что условие (4) является необходимым, но не достаточным условием эффективности диагностического теста с точки зрения уменьшения априорного риска ошибочной диагностики.

В самом деле, пусть $P(V_1)$ – вероятность появления больного в группе обследуемых (преваленс заболевания). Будем характеризовать возможные ошибки диагностического теста величинами L_{12}, L_{21} , которые определяют потери от ошибок пропуска цели (не выявление больного) и ложной тревоги (отнесение здорового пациента к группе больных).

Тогда ожидаемые потери (средний риск) от неправильной диагностики определит соотношение

$$R = P(V_1) \cdot (1 - S_E) \cdot L_{12} + [1 - P(V_1)] \cdot (1 - S_P) \cdot L_{21}. \quad (5)$$

С другой стороны, если не использовать диагностический тест и принимать решения только на основании априорной вероятности $P(V_1)$ и величин L_{12}, L_{21} , то средний риск таких решений определит соотношение [4]

$$R_0 = \begin{cases} L_{12}P(V_1), & \text{если } P(V_1)(1 + \omega) < 1 \\ L_{21}[1 - P(V_1)], & \text{если } P(V_1)(1 + \omega) > 1 \end{cases} \quad (6)$$

где $\omega = L_{12}L_{21}^{-1}$ – безразмерная величина, характеризующая соотношение потерь от ошибок пропуска цели и ложной тревоги.

На основании соотношений (5) и (6) может быть доказано утверждение.

Утверждение. Тест гарантированно эффективен с точки зрения уменьшения априорного риска ошибочной диагностики, т.е. с точки зрения *строгого* неравенства

$$R < R_0, \quad (7)$$

когда чувствительность S_E и специфичность S_P теста связаны соотношением

$$S_E > \theta(1 - S_P), \quad \text{если } \theta \geq 1; \quad (8)$$

или

$$S_E > 1 - \theta + \theta(1 - S_P), \quad \text{если } \theta < 1; \quad (9)$$

где величина θ определяется по формуле

$$\theta = \frac{1 - P(V_1)}{\omega P(V_1)}. \quad (10)$$

Поскольку, как следует из утверждения, соотношение (4) не является достаточным условием, гарантирующим эффективность теста с точки зрения уменьшения риска ошибочной диагностики, предлагается усилить традиционный **ROC**-анализ, ограничив соответствующую область анализа **ROC**-кривой.

Рис.2 иллюстрирует предлагаемый подход для случаев $\theta \geq 1$ (рис. 2, а) и $\theta < 1$ (рис. 2, б). В соответствии с (8) для случаев $\theta \geq 1$ ограничивающая прямая проходит через

точку с координатами (0,0) и точку с координатами (1/θ, 1), а для случая θ < 1 – через точку с координатами (1,1) и точку с координатами (0, 1-θ).

Отрезок OA (рис. 2, а) соответствует случаю, когда преваленс заболевания равен 15 %, а ω = 1. Видно, что в данном случае тест неэффективен, 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 ω = 4, то соответствующая прямая OB уже пересекает ROC-кривую. Следовательно, тест, имеющий чувствительность C_E = 62,5% и специфичность C_P = 80 %, окажется гарантированно эффективным с точки зрения строгого неравенства (7), поскольку соответствующая таким операционным характеристикам точка C расположена на допустимой части ROC-кривой.

Усиленный ROC-анализ позволяет упростить процедуру оценки эффективности диагностического теста, придав ей наглядную интерпретацию в виде когнитивного графического образа.

Заметим также, что условия (8) и (9) позволяют решить и обратную задачу: определить допустимый интервал соотношения потерь ω, при котором тест с известными операционными характеристиками – чувствительностью S_E и специфичностью S_P можно считать полезным для скрининга заболевания с известным преваленсом P(V₁).

Этот интервал определяется формулой

$$\frac{1 - P(V_1)}{P(V_1)} \frac{1 - S_P}{S_E} \leq \omega \leq \frac{1 - P(V_1)}{P(V_1)} \frac{S_P}{1 - S_E}. \quad (11)$$

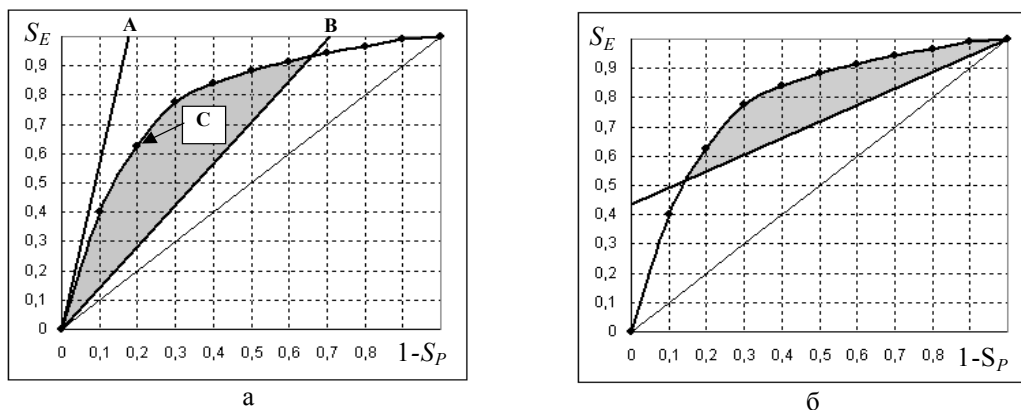


Рис. 2. Иллюстрация идеи усиленного ROC-анализ для случаев θ ≥ 1 (а) и θ < 1 (б)

Например, тест, имеющий чувствительность S_E = 62,5% и специфичность S_P = 80%, можно рекомендовать для выявления заболеваний с преваленсом P = 15%, если допустимо соотношение потерь от ложноотрицательных и ложноположительных ошибок в пределах 2,66 ≤ ω ≤ 17,7.

Предложенный подход реализован в виде инструментального вычислительного модуля (рис. 3) информационной технологии ФАЗАГРАФ и обеспечил гарантированную

оценку эффективности диагностического теста, реализованного в этой технологии для решения задачи скрининга ишемической болезни сердца по фазовому портрету ЭКГ [5].

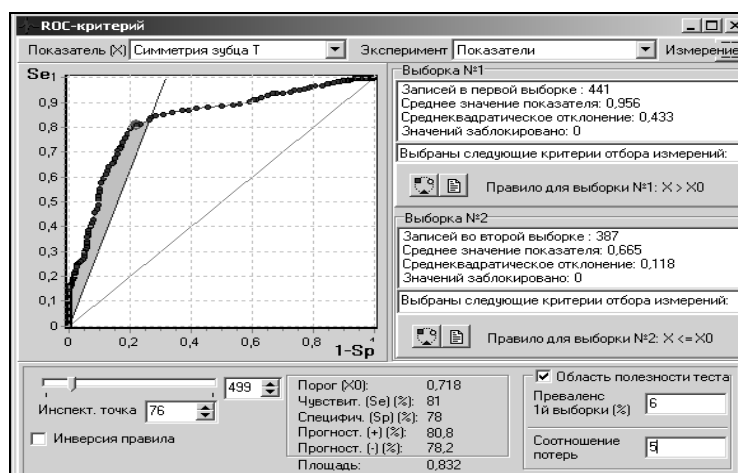


Рис. 3. Интерфейс вычислительного модуля построения ROC-кривой

Литература

1. Ben-Bassat M. Irrelevant features in pattern recognition // IEEE Trans. Comput. – 1978. – Vol. C-27. – No 8. – P. 749–766.
2. Fainzilberg L.S. Why Relevant Features May Be Unuseful in Statistical Recognition of Two Classes // Proc. of the 13th International Conference on Pattern Recognition (ICPR'96). – Viena (Austria). – 1996. – P. 730–734.
3. Файнзильберг Л.С. Оценка полезности признаков при решении задач диагностики в статистической постановке // Математические машины и системы. – 1998. – № 1. – С. 57-64.
4. Файнзильберг Л.С. К вопросу о полезности диагностических методов в задачах скрининга // Управляющие системы и машины. – 2002. – № 6. – С. 10-17.
5. Чайковский И.А., Файнзильберг Л.С. Медицинские аспекты применения устройства ФАЗАГРАФ в клинической практике и в домашних условиях.– Киев: МНУЦ ИТИС НАН Украины, 2009.– 74 с.