

УДК 004.93
<https://doi.org/10.37661/1816-0301-2021-18-1-61-71>

Оригинальная статья
Original Paper

Об оценке результатов классификации несбалансированных данных по матрице ошибок

В. В. Старовойтов[✉], Ю. И. Голуб

*Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь
[✉]E-mail: valerys@newman.bas-net.by*

Аннотация. При применении классификаторов в реальных приложениях часто число элементов одного класса больше числа элементов другого, т. е. имеет место дисбаланс данных. В статье исследуются оценки результатов классификации данных такого типа. Рассматривается, какой из переводов термина confusion matrix более точен, как предпочтительнее представлять данные в такой матрице и какими функциями лучше оценивать результаты классификации по ней.

На реальных данных демонстрируется, что с помощью популярной функции точности ассигасу не всегда корректно оцениваются ошибки классификации несбалансированных данных. Нельзя также сравнивать значения функции ассигасу, вычисленные по матрицам с абсолютными количественными и нормализованными по классам результатами классификации. При дисбалансе данных точность, вычисленная по матрице ошибок с нормализованными значениями, как правило, будет иметь меньшие значения, поскольку она определяется по иной формуле. Такой же вывод сделан относительно большинства функций, используемых в литературе для нахождения оценок результатов классификации.

Показывается, что для представления матриц ошибок лучше использовать абсолютные значения распределения объектов по классам вместо относительных, так как они описывают количество протестированных данных каждого класса и их дисбаланс.

При построении классификаторов рекомендуется оценивать ошибки функциями, не зависящими от дисбаланса данных, что позволяет надеяться на получение более корректных результатов классификации реальных данных.

Ключевые слова: классификация объектов, несбалансированные данные, матрица ошибок, функции точности классификации, парадокс точности, нейронная сеть, диагностика заболеваний

Благодарности. Работа частично выполнена в рамках проекта БРФФИ Ф20РА-014.

Для цитирования. Старовойтов, В. В. Об оценке результатов классификации несбалансированных данных по матрице ошибок / В. В. Старовойтов, Ю. И. Голуб // Информатика. – 2021. – Т. 18, № 1. – С. 61–71. <https://doi.org/10.37661/1816-0301-2021-18-1-61-71>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 27.12.2020
Подписана в печать | Accepted 04.01.2021
Опубликована | Published 26.03.2021

About the confusion-matrix-based assessment of the results of imbalanced data classification

Valery V. Starovoitov[✉], Yuliya I. Golub

*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus*

[✉]E-mail: valerys@newman.bas-net.by

Abstract. When applying classifiers in real-world applications, data imbalance often occurs when the number of elements of one class is greater than another. The article examines the estimates of the classification results for this type of data. The paper provides answers to three questions: which term is a more accurate translation of the phrase confusion matrix, how it is preferable to represent data in this matrix, and what functions are better to evaluate the results of classification by such a matrix.

The paper demonstrates on real data that the popular accuracy function cannot correctly estimate the classification errors for imbalanced data. It is also impossible to compare the values of this function, calculated by matrices with the absolute quantitative results of classification and normalized by classes. If the data is imbalanced, the accuracy calculated from the confusion matrix with normalized values will usually have lower values, since it is calculated by a different formula. The same conclusion is made for most of the classification accuracy functions used in the literature for estimation of classification results.

It is shown that to represent confusion matrices it is better to use absolute values of object distribution by classes instead of relative ones, since they give an idea of the amount of data tested for each class and their imbalance.

When constructing classifiers, it is recommended to evaluate errors by functions that do not depend on the data imbalance, which allows us to hope for more correct classification results for real data.

Keywords: classification, imbalanced data, confusion matrix, classification accuracy functions, accuracy paradox, neural network, disease diagnosis

Acknowledgements. This work was partially performed within the framework of the BRFFR project F20RA-014.

For citation. Starovoitov V. V., Golub Yu. I. About the confusion-matrix-based assessment of the results of imbalanced data classification. *Informatics*, 2021, vol. 18, no. 1, pp. 61–71 (in Russian). <https://doi.org/10.37661/1816-0301-2021-18-1-61-71>

Conflict of interest. The author declare of no conflict of interest.

Введение. Классификация – это процедура разделения некоторого множества объектов на заранее определенные классы. Так, скрининг в здравоохранении представляет собой обследование группы клинически бессимптомных лиц с целью выявления случаев некоторого заболевания, т. е. разделения людей на здоровых и больных. Флюорография служит для выявления скрыто протекающих заболеваний легких и является примером бинарной классификации. При диагностике многих заболеваний определяют их стадии согласно ряду признаков. Можно выполнить разделение группы пациентов с выявленным заболеванием на классы согласно стадии заболевания по образцу многоклассовой классификации. Пропорции классов неизвестны, но, как правило, имеют разные размеры. Этим и определяется дисбаланс данных [1].

Результаты деления некоторого множества данных на N классов часто описываются в виде квадратной матрицы $C = [c_{ij}]$, $i, j \in [1, 2, \dots, N]$, где i – номер строки, j – номер столбца, N – число предварительно определенных классов. По строкам обычно записываются результаты, определенные экспертами и называемые истинными, по столбцам – предсказанные классификаторами. Тогда c_{ij} есть число объектов i -го класса, отнесенных к j -му классу. Количество объектов m -го класса определено экспертами и равно $n_m = \sum_{j=1}^N c_{mj}$. Коэффициент дисбаланса классов данных (отношение числа объектов большего класса i к числу объектов меньшего класса k) определяется выражением $IR = \max_{m,k} (n_m/n_k)$ [1].

В англоязычной литературе матрицу, подобную $C = [c_{ij}]$, называют confusion matrix. Одно из первых упоминаний такой матрицы при описании результатов классификации приведено

в статье [2]. Известны разные трактовки данного выражения в русскоязычной литературе (табл. 1). На взгляд авторов, более точным и распространенным вариантом перевода является словосочетание «матрица ошибок».

Таблица 1
 Частота использования вариантов перевода термина confusion matrix

Термин	Число ссылок при поиске в google.com	Число ссылок на сайте scholar.google.com
Confusion matrix	1 350 000	121 000
Matrix error	111 000	6270
Матрица ошибок	508 000	8130
Матрица неточностей	1580	105
Матрица путаницы	1720	9

Чтобы представить данные в матрице ошибок для последующей оценки результатов классификации, обычно используются два варианта значений: абсолютные количественные и нормализованные по классам. В первом варианте элементы матрицы ошибок – это числа объектов, отнесенных к каждому классу. Во втором варианте сначала формируется матрица абсолютных значений, затем она нормализуется путем деления каждого элемента матрицы на сумму элементов строки, в которой он находится, т. е. на число элементов класса согласно экспертным оценкам. После этого сумма элементов каждой строки матрицы ошибок становится равной единице.

Для определения лучшего способа оценки результатов классификации несбалансированных данных, представленных матрицей ошибок, оценим на примерах инвариантность оценок результатов классификации, вычисленных по обоим вариантам представления матриц ошибок.

Одной из самых популярных функций оценок результатов классификации является функция точности ассигасу, которая вычисляется по данным матрицы ошибок. При использовании этой функции часто возникают некорректные заключения о качестве работы классификатора, когда не учитывается представление данных в матрице (абсолютное или нормализованное). Кроме того, ассигасу часто используется в процессе обучения искусственных нейронных сетей разных типов, а некорректные оценки результатов классификации могут привести к выбору не самого лучшего варианта классификатора.

В табл. 2 приведены основные формулы оценки результатов многоклассовой классификации, используемые в статье.

Таблица 2

Основные функции оценок многоклассовой классификации

Обозначение	Математическое выражение
ACC – функция точности ассигасу	$ACC = \frac{1}{N} (c_{11} + c_{22} + \dots + c_{NN})$
ACCBal – сбалансированная по классам функция точности [3]	$ACCBal = \frac{1}{N} \left(\frac{c_{11}}{n_1} + \frac{c_{22}}{n_2} + \dots + \frac{c_{NN}}{n_N} \right)$
SinACC – сбалансированная по классам функция точности на базе синусов [4]	$SinACC = 1 - \frac{1}{N} \sum_{i=1}^N \frac{\sqrt{\sum_{j \neq i}^N c_{ij}^2}}{\sqrt{\sum_{j=1}^N c_{ij}^2}}$
AU1U – среднее сумм функций чувствительности и специфичности, вычисленных для всех вариантов бинарных матриц $[c_{ii}, c_{ki}, c_{kk}, c_{ik}]$, построенных из матрицы ошибок C [5]	$AU1U = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{k>i}^N \left(\frac{c_{ii}}{c_{ii} + c_{ki}} + \frac{c_{kk}}{c_{kk} + c_{ik}} \right)$

Далее в табл. 3 и 4 используются следующие наименования оценочных функций, формулы которых можно найти в работах [5–7]:

Карра – каппа-функция Коэна;

AveragePrecision – среднее арифметическое функций Precision, вычисленных для отдельных классов;

GeomMeanPrecision – среднее геометрическое функций Precision, вычисленных для отдельных классов;

GeomMeanSensitivity – среднее геометрическое функций Sensitivity, вычисленных для отдельных классов;

CosineCoef – косинусный коэффициент;

VM – среднее геометрическое функций Sensitivity и Precision, вычисленных для отдельных классов;

Fmicro – среднее арифметическое гармонических средних Sensitivity и Precision для отдельных классов;

Fmacro – гармоническое среднее Sensitivity и Precision для отдельных классов;

Jmacro – среднее арифметическое индексов Юдена для отдельных классов;

sInd – среднее арифметическое индексов сходства в ROC-пространстве на базе специфичности и чувствительности отдельных классов;

normMCC – нормализованный коэффициент Коэна с диапазоном значений [0, 1];

AUNU – средняя площадь под ROC-кривыми N бинарных матриц ошибок, построенных по принципу «один против всех» из матрицы C [6];

AUNP – то же с учетом весов, пропорциональных числу элементов каждого класса к сумме всех данных.

Если классификация бинарная ($N = 2$), то $AU1U = ACCBal$ по определению. Кроме того, $AU1U = AUNU = AUNP$, а для нормализованных матриц ошибок все эти функции совпадают с ACC.

Анализ оценок результатов классификации заболеваний кожи. В работе [8] исследована задача классификации семи классов заболеваний кожи с помощью глубокой сверточной нейронной сети. Экспериментировали на наборе данных, который использовался при проведении международного конкурса компьютерных алгоритмов ISIC-2018 Challenge. В нем собраны цветные цифровые изображения пораженных участков кожи по семи приведенным ниже классам заболеваний:

1. *Actinic keratoses*. Актинический кератоз и интраэпителиальная карцинома, распространенные неинвазивные варианты плоскоклеточного рака, которые можно лечить без хирургического вмешательства.

2. *Basal cell carcinoma*. Базально-клеточная карцинома – распространенный вариант эпителиального рака кожи, который редко метастазирует, но при отсутствии лечения разрастается.

3. *Benign keratosis*. Доброкачественный кератоз, себорейный кератоз, солнечный лентигио и лишайник плоскоклеточный как кератоз.

4. *Dermatofibroma*. Дерматофиброма – доброкачественное поражение кожи как воспалительная реакция на травму.

5. *Melanocytic nevi*. Меланоцитарные невусы – доброкачественные новообразования меланоцитов.

6. *Melanoma*. Меланома – смертельная форма рака кожи, которая на ранней стадии часто не диагностируется или ошибочно диагностируется как доброкачественное поражение кожи.

7. *Vascular lesions*. Сосудистые поражения кожи.

Раннее и точное выявление меланомы крайне важно, поскольку от этого зависит жизнь пациентов. На рис. 1, *a* показана матрица ошибок C из статьи [8], где в шестом классе описаны результаты определения меланомы: в шестом столбце – экспертами, в шестой строке – нейронной сетью. Отметим, что во всех матрицах ошибок, представленных в настоящей статье, истинные результаты классификации записаны в столбцах.

1	100	1	10	2	3	8	0
2	11	181	6	3	13	2	0
3	11	2	293	1	17	38	0
4	3	1	0	24	4	1	0
5	13	12	60	6	2627	156	3
6	2	6	33	3	20	242	0
7	0	0	0	0	2	2	64
	1	2	3	4	5	6	7

а)

1	0.71	0	0.02	0.05	0	0.02	0
2	0.08	0.89	0.01	0.08	0	0	0
3	0.08	0.01	0.73	0.03	0.01	0.08	0
4	0.02	0	0	0.62	0	0	0
5	0.09	0.06	0.15	0.15	0.98	0.35	0.04
6	0.01	0.03	0.08	0.08	0.01	0.54	0
7	0	0	0	0	0	0	0.96
	1	2	3	4	5	6	7

б)

Рис. 1. Примеры представления матрицы ошибок: а) абсолютные значения; б) нормализованные (округленные до сотых) значения

Показатель точности (ассигасу, АСС) для данной матрицы ошибок составляет $АСС = 0,8864$, или 88,64 %, а для ее нормализованного варианта (рис. 1, б) $АСС = 0,77465$, или 77,465 %. Разные оценки точности получились потому, что тестируемые данные не сбалансированы по классам, а функция АСС этого не учитывает. Самым большим классом в исходной матрице является пятый, самым маленьким – четвертый, коэффициент дисбаланса $IR = 68,87$. Для более объективной оценки точности классификации следует применять функцию сбалансированной точности (balanced accuracy, АССBal) либо другой показатель, например функцию точности на базе синусов SinАСС [4].

Если в матрице ошибок изменить результаты классификации шестого класса, увеличив их пропорционально в десять и сто раз (рис. 2), значения функции АСС станут равными 0,71194 и 0,56777 соответственно. Однако после нормализации измененных значений все матрицы ошибок примут одинаковый вид, как на рис. 1, б. При этом значение функции АСС для трех вариантов матриц ошибок будет одинаковым – 0,77465. Это легко доказать:

$$АССBal = \frac{1}{N} \left(\frac{c_{11}}{n_1} + \frac{c_{22}}{n_2} + \dots + \frac{c_{kk}}{n_k} + \dots + \frac{c_{NN}}{n_N} \right) = \frac{1}{N} \sum_i^N c_{ii}^{norm} = АСС,$$

где c_{ii}^{norm} – значения диагональных элементов нормализованной матрицы ошибок.

Таким образом, при наличии несбалансированных классов данных при нормализации значений матрицы ошибок теряется информация о дисбалансе классов. Популярная функция точности АСС трансформируется в функцию сбалансированной точности АССBal. Вследствие этого сравнивать значения функции АСС, вычисленные для нормализованных и ненормализованных матриц, некорректно.

1	100	1	10	2	3	80	0
2	11	181	6	3	13	20	0
3	11	2	293	1	17	380	0
4	3	1	0	24	4	10	0
5	13	12	60	6	2627	1560	3
6	2	6	33	3	20	2420	0
7	0	0	0	0	2	20	64
	1	2	3	4	5	6	7

а)

1	100	1	10	2	3	800	0
2	11	181	6	3	13	200	0
3	11	2	293	1	17	3800	0
4	3	1	0	24	4	100	0
5	13	12	60	6	2627	15600	3
6	2	6	33	3	20	24200	0
7	0	0	0	0	2	200	64
	1	2	3	4	5	6	7

б)

Рис. 2. Количество объектов шестого класса, увеличенное в десять (а) и сто (б) раз. Ошибки классификации распределены в тех же пропорциях

Если требуется выполнить разделение на два класса (меланому и другие заболевания), то данные будут сбалансированы по-иному: $IR = 7,8775$. Как видно на рис. 3, 79 % от общего числа изображений меланомы диагностированы верно на фоне других заболеваний кожи. Точность диагностики по матрице абсолютных значений ошибок функцией ассигасу равна 0,9320, при

этом точность по нормализованным значениям ошибок, вычисленная другими функциями, равна $ACC_{Bal} = AUNU = AUNP = AUIU = 0,8673$.

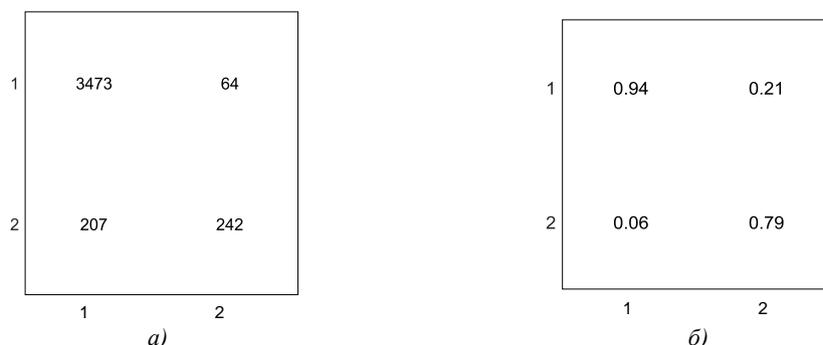


Рис. 3. Заболевания кожи, разделенные на два класса: меланома и другие заболевания.
Матрицы ошибок с абсолютными (а) и нормализованными значениями (б)

Рассмотрим оценки, вычисляемые по матрицам ошибок при мультиклассовой классификации (табл. 3). Курсивом выделены значения функций, которые изменяются при разном дисбалансе классов. Только четыре функции в табл. 3 инвариантны к дисбалансу классов: ACC_{Bal} , $SinACC$, $GeomMeanSensitivity$ и $AUIU$.

Таблица 3

Оценки, вычисленные по матрицам, представленным на рис. 1 и 2

Функция	С абсолютными значениями, IR = 68,87	С нормализованными значениями, IR = 1	С увеличенным в сто раз шестым классом, IR = 1151,3
ACC	0,8864	<i>0,7746</i>	<i>0,5677</i>
ACC_{Bal}	<i>0,7746</i>	<i>0,7746</i>	<i>0,7746</i>
$SinACC$	<i>0,7966</i>	<i>0,7966</i>	<i>0,7966</i>
Kappa	0,7681	<i>0,7371</i>	<i>0,1591</i>
AveragePrecision	0,8324	<i>0,8133</i>	<i>0,3116</i>
GeomMeanPrecision	0,8296	<i>0,7992</i>	<i>0,2152</i>
GeomMeanSensitivity	0,7581	0,7581	0,7581
CosineCoef	0,8030	<i>0,7937</i>	<i>0,4913</i>
VM	0,8010	<i>0,7849</i>	<i>0,4363</i>
FMicro	0,8025	<i>0,7935</i>	<i>0,4445</i>
FMacro	0,7985	<i>0,7762</i>	<i>0,3603</i>
JMacro	0,7392	<i>0,7371</i>	<i>0,7077</i>
sInd	0,8232	<i>0,8276</i>	<i>0,8069</i>
normMCC	0,8862	<i>0,8726</i>	<i>0,6422</i>
AUNU	0,8696	<i>0,8685</i>	<i>0,8538</i>
AUNP	0,8759	<i>0,8685</i>	<i>0,7656</i>
AUIU	0,9527	0,9527	0,9527

Анализ оценок результатов классификации землетрясений. Проанализируем матрицы ошибок, описанные в статье [9]. Целью работы было исследование возможностей искусственных нейронных сетей по предсказанию мощности землетрясений. Рассматривались три варианта деления наблюдаемых сейсмических сигналов в зависимости от предсказанной и зарегистрированной магнитуды M в эпицентре. Все данные были суммированы по месяцам с июля 2012 г. по июнь 2013 г. По данным каждого месяца строились матрицы ошибок класси-

фикации землетрясений в зависимости от магнитуды на два класса ($M \leq 2,0$; $M > 2,0$), три класса ($M \leq 1,5$; $1,5 < M < 3,0$; $M \geq 3,0$) и четыре класса ($M < 0,5$; $0,5 \leq M < 1,5$; $1,5 \leq M < 3,0$; $M \geq 3,0$) соответственно.

Рассмотрим результаты классификации данных о землетрясениях за июнь 2013 г. на три класса (рис. 4, а). Авторы работы [9] утверждают, что точность прогноза, представленного этой матрицей, очень хорошая, особенно для слабых землетрясений, которых подавляющее большинство. В данном примере коэффициент дисбаланса классов $IR = 3066,5$. Формально результат классификации хорош, так как значение точности $ACC = 0,9932$. Однако в матрице ошибок только 37,98 % объектов второго и 30,00 % объектов третьего класса классифицированы верно. При этом сбалансированная точность $ACC_{Bal} = 0,6966$, т. е. в 1,4258 раза меньше, чем точность при абсолютном представлении данных в матрице ошибок.

В результаты прогноза для второго класса в матрицу ошибок были внесены изменения (рис 4, б). В новой матрице всего 27 % объектов второго класса классифицированы верно. При этом $ACC = 0,9908$, т. е. показатель точности при оценке абсолютных данных уменьшился всего на 0,24 %. Однако для нормализованного варианта матрицы ошибок сбалансированная точность $ACC_{Bal} = 0,5214$, т. е. она уменьшилась на 33,6 % относительно исходного нормализованного значения функции ACC_{Bal} .

Далее в матрице ошибок изменим результаты классификации третьего класса. Он описан в третьем столбце матрицы на рис. 4, в и соответствует самым мощным землетрясениям. Абсолютная погрешность по сравнению с исходной матрицей невелика: всего пять лишних объектов третьего класса ошибочно отнесены к соседнему второму классу землетрясений средней мощности. Этой матрице соответствуют следующие оценки результатов классификации: абсолютная точность $ACC = 0,9931$; сбалансированная точность $ACC_{Bal} = 0,6113$, т. е. на 37,54 % ниже.

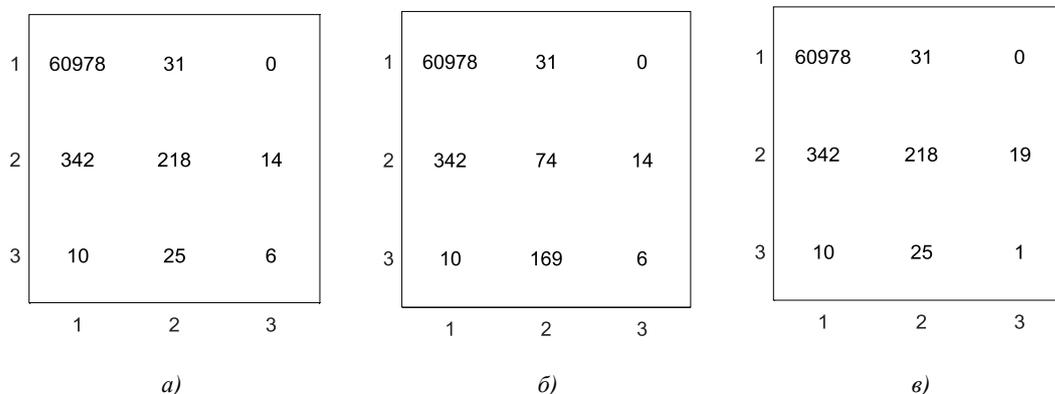


Рис. 4. Матрица ошибок предсказания землетрясений за июнь 2013г. (а), измененные результаты классификации второго (б) и третьего (в) классов

Приведем еще два примера оценок классификации реальных результатов предсказания землетрясений из работы [9]. На рис. 5, а показаны ошибки классификации по измерениям, полученным за январь 2013 г. (всего 63 675 объектов), а на рис. 5, б – за декабрь 2012 г. (всего 63 677 объектов, на два больше). Вместе с тем IR первой матрицы на рис. 5 равен 3961,8, а второй – 5279,8. При этом значения функции точности ACC практически одинаковы: 0,9956 и 0,99536, но значения сбалансированной точности ACC_{Bal} намного меньше: 0,6423 и 0,6012. Вычислим точность классификации по отдельным классам: для первой матрицы 0,9962, 0,8474 и 0,0833; для второй – 0,9972, 0,6815 и 0,1250. Последние числа в каждой тройке объясняют, почему снизились усредненные значения сбалансированной точности, но ничего не говорят о дисбалансе классов. Матрицы с оценками ошибок изображены на рис. 6.

1	63209	44	1
2	169	184	13
3	11	42	2
	1	2	3

a)

1	63119	39	1
2	228	261	10
3	10	8	1
	1	2	3

б)

Рис. 5. Матрицы ошибок классификации данных за январь 2013 г. (а) и декабрь 2012 г. (б)

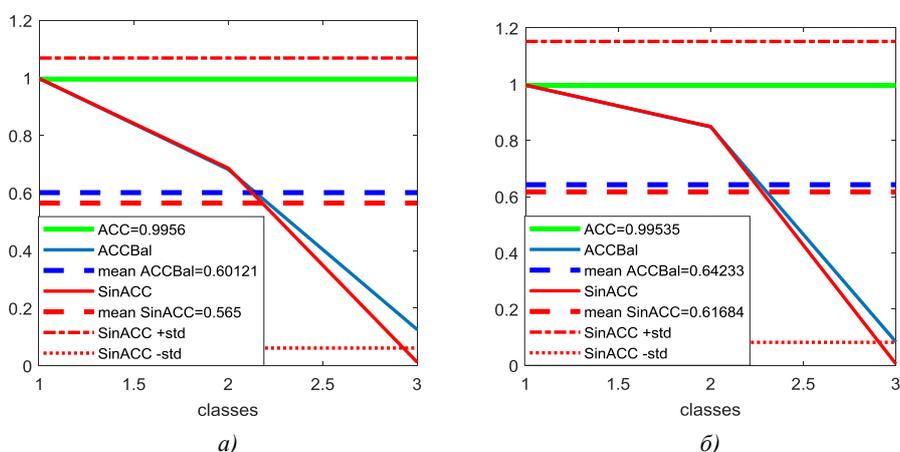


Рис. 6. Графики и оценки классификации (общие и по классам), вычисленные по матрицам на рис. 5 за январь 2013 г. (а) и декабрь 2012 г. (б)

На рис. 7 показаны матрицы ошибок предсказания землетрясений при разделении одних и тех же данных на два, три и четыре класса землетрясений согласно магнитудам (см. данные за декабрь 2012 г. в статье [9]). Коэффициенты дисбаланса данных в этих матрицах равны 470,7, 5279,8 и 5256,9. При делении данных на четыре класса (рис. 7, в) второй и четвертый классы были распознаны недостаточно: всего 11,68 и 8,33 % образцов в этих классах распознаны верно. Тем не менее для всех четырех классов стандартная функция точности дает значение 0,9953, а нормализованная – всего 0,5119. При делении данных на два класса (рис. 7, а) значения этих функций равны соответственно 0,9934 и 0,9302, т. е. достаточно высокие. Таким образом, при создании классификатора нужно учитывать дисбаланс при делении данных на классы и результаты классификации отдельных классов.

1	63141	18
2	401	117
	1	2

a)

1	63119	39	1
2	228	261	10
3	10	8	1
	1	2	3

б)

1	63082	4	11	1
2	1	32	28	0
3	0	228	261	10
4	0	10	8	1
	1	2	3	4

в)

Рис. 7. Матрицы ошибок классификации данных за декабрь 2012 г. при их разделении на два (а), три (б) и четыре (в) класса землетрясений

Приведенные примеры матриц ошибок (рис. 2–5, 7) подтверждают так называемый парадокс точности (ассигасу paradox, впервые описанный в статье [10]), который отмечает некорректность применения классической функции точности к несбалансированным данным. Если абсолютно все данные будут отнесены неким примитивным классификатором к самому большому классу, точность этого классификатора будет равна отношению реального числа объектов наибольшего класса к общему объему данных. Например, для данных, представленных матрицей ошибок на рис. 4, а, точность будет равна 98,95 %. Однако такой классификатор будет абсолютно бесполезен, так как он не предскажет ни одного мощного землетрясения, а в медицинских приложениях – ни одного заболевания.

В табл. 4 приведены значения других оценок, используемых при мультиклассовой классификации данных и вычисленных по матрицам, представленным на рис. 4, а и б, а также их нормализованным вариантам. Курсивом выделены значения оценок, изменившихся после нормализации матрицы ошибок; жирным курсивом – значения оценок, увеличившихся после существенного уменьшения числа правильно классифицированных данных второго класса; жирным шрифтом – названия функций, значения которых не меняются после нормализации матрицы ошибок.

Все месячные данные и результаты их классификации, приведенные в работе [9], очень схожи. Показатели точности разделения сейсмических сигналов этих данных на два, три и четыре класса выше 99 %, поэтому авторы статьи [9] заявляют о разработке очень хорошего классификатора сигналов для предсказания землетрясений. Проведенный анализ показывает, что это не совсем корректно.

Таблица 4

Оценки результатов классификации, вычисленные по матрицам ошибок предсказания землетрясений

Функция	Матрица на рис. 4, а с абсолютными значениями, IR = 3066,5	С нормализованными значениями, IR = 1	Матрица на рис. 4, б с измененными результатами второго класса, IR = 3066,5	С нормализованными значениями, IR = 1
ACC	0,9932	<i>0,5085</i>	0,9908	0,4249
ACCBal	0,5085	0,5085	0,4249	0,4249
SinACC	0,3933	0,3933	0,3473	0,3473
Kappa	0,5329	<i>0,2628</i>	0,3689	0,1374
AveragePrecision	0,6966	<i>0,5947</i>	0,5339	0,3576
GeomMeanPrecision	0,6191	<i>0,5632</i>	0,2792	0,3232
GeomMeanSensitivity	0,3816	0,3816	0,2662	0,2662
CosineCoef	0,5952	<i>0,5499</i>	0,4763	0,3898
VM	0,5854	<i>0,4910</i>	0,4476	0,3711
FMicro	0,5879	<i>0,5483</i>	0,4732	0,3883
FMacro	0,5692	<i>0,4453</i>	0,4226	0,3523
JMacro	0,3174	<i>0,2628</i>	0,2287	0,1374
sInd	0,5177	<i>0,5369</i>	0,4555	0,4783
normEMCC	0,7851	<i>0,6499</i>	0,6985	0,5764
AUNU	0,6587	<i>0,6314</i>	0,6143	0,5687
AUNP	0,7132	<i>0,6314</i>	0,7056	0,5687
AUIU	0,6495	0,6495	0,5122	0,5122

Заключение. Популярная функция точности ассигасу не может корректно оценивать ошибки классификации несбалансированных классов данных. Нельзя также сравнивать значения данной функции, вычисленные по матрицам с абсолютными количественными и нормализованными по классам результатами классификации. В этом случае оценки могут существенно

различаться при дисбалансе классов данных, поскольку функция точности, вычисленная по матрице ошибок с нормализованными значениями, является сбалансированной и имеет другую формулу.

Такой же вывод сделан относительно большинства функций, используемых для вычисления оценок результатов классификации данных и представленных в статье. Исключением являются только функции сбалансированной точности ACCBal, функции SinACC, GeomMeanSensitivity и AU1U. Любые вычисляющие геометрическое среднее функции, такие как GeomMeanSensitivity и GeomMeanPrecision, равняются нулю, если хотя бы один из классов будет полностью неверно классифицирован, т. е. не будет верно распознанных объектов этого класса. Все это имеет место при использовании простейших классификаторов. Поэтому функции геометрического усреднения не следует применять для оценок результатов классификации, особенно при большом числе классов и (или) малом количестве объектов одного из классов.

При использовании любого классификатора для прогнозирования в реальных ситуациях никогда не известен истинный баланс данных. В реальных приложениях дисбаланс имеет место всегда: здоровых больше, чем больных; слабых землетрясений больше, чем сильных, и т. п. Это подтверждают оценки матриц ошибок на рис. 4, которые приведены в табл. 4. Таким образом, при обучении и настройке классификаторов лучше оценивать результаты функциями, не зависящими от дисбаланса данных. Независимость от дисбаланса позволяет надеяться на получение более корректных результатов классификации реальных данных.

Для представления матриц ошибок лучше использовать абсолютные значения вместо относительных, так как они дают представление о количестве протестированных данных каждого класса и их дисбалансе.

Список использованных источников

1. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches / M. Galar [et. al.] // IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). – 2012. – Vol. 42, no. 4. – P. 463–484. <https://doi.org/10.1109/tsmcc.2011.2161285>
2. Miller, G. A. An analysis of perceptual confusions among some English consonants / G. A. Miller, P. E. Nicely // The J. of the Acoustical Society of America. – 1955. – Vol. 27, no. 2. – P. 338–352. <https://doi.org/10.1121/1.1907526>
3. The balanced accuracy and its posterior distribution / K. H. Brodersen [et. al.] // Intern. Conf. on Pattern Recognition. – Istanbul, 2010. – P. 3121–3124.
4. Starovoitov, V. New function for estimating imbalanced data classification results / V. Starovoitov, Yu. Golub // Pattern Recognition and Image Analysis. – 2020. – Vol. 30, no. 3. – P. 295–302.
5. Hand, J. D. A simple generalisation of the area under the ROC curve for multiple class classification problems / J. D. Hand, J. R. Till // Machine Learning. – 2001. – Vol. 45, no. 2. – P. 171–186. <https://doi.org/10.1023/A:1010920819831>
6. Fawcett, T. Using rule sets to maximize ROC performance / T. Fawcett // Proc. IEEE Intern. Conf. on Data Mining. – 2001. – P. 131–138. <https://doi.org/10.1109/ICDM.2001.989510>
7. Ballabio, D. Multivariate comparison of classification performance measures / D. Ballabio, F. Grisoni, R. Todeschini // Chemometrics and Intelligent Laboratory Systems. – 2018. – Vol. 174. – P. 33–44. <https://doi.org/10.1016/j.chemolab.2017.12.004>.
8. Классификация повреждений кожи по данным дерматоскопии с использованием методов глубокого обучения / Е. Ю. Щетинин [и др.] // Математическая биология и биоинформатика. – 2020. – Т. 15, № 2. – С. 180–194.
9. Buscema, P. M. Artificial Adaptive Systems to predict the magnitude of earthquakes / P. M. Buscema, G. Massini, G. Maurelli // Bollettino di Geofisica Teorica ed Applicata. – 2015. – Vol. 56, no. 2. – P. 227–256.
10. Valverde-Albacete, F.J. 100 % classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox / F. J. Valverde-Albacete, C. Pelaez-Moreno // PloS One. – 2014. – Vol. 9, no. 1. – P. e84217. <https://doi.org/10.1371/journal.pone.0084217>

References

1. Galar M., Fernandez A., Barrenechea E., Bustince H., Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012, vol. 42, no. 4, pp. 463–484. <https://doi.org/10.1109/tsmcc.2011.2161285>
2. Miller G. A., Nicely P. E. An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 1955, vol. 27, no. 2, pp. 338–352. <https://doi.org/10.1121/1.1907526>
3. Brodersen K. H., Ong C. S., Stephan K. E., Buhmann J. M. The balanced accuracy and its posterior distribution. *International Conference on Pattern Recognition*, 2010, Istanbul, pp. 3121–3124.
4. Starovoitov V., Golub Yu. New function for estimating imbalanced data classification results. *Pattern Recognition and Image Analysis*, 2020, vol. 30, no. 3, pp. 295–302.
5. Hand J. D., Till J. R. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, 2001, vol. 45, no. 2, pp. 171–186. <https://doi.org/10.1023/A:1010920819831>
6. Fawcett T. Using rule sets to maximize ROC performance. *Proceedings IEEE International Conference on Data Mining*, 2001, pp. 131–138. <https://doi.org/10.1109/ICDM.2001.989510>
7. Ballabio D., Grisoni F., Todeschini R. Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems*, 2018, vol. 174, pp. 33–44. <https://doi.org/10.1016/j.chemolab.2017.12.004>
8. Shhetinin E. Ju., Sevast'janov L. A., Demidova A. V., Kuljabov D. S. Klassifikacija povrezhdenij kozhi po dannym dermatoskopii s ispol'zovaniem metodov glubokogo obuchenija [Dermatoscopy classification of skin lesions using deep learning methods]. *Matematicheskaja biologija i bioinformatika [Mathematical biology and bioinformatics]*, 2020, vol. 15, no. 2, pp. 180–194.
9. Buscema P. M., Massini G., Maurelli G., Artificial Adaptive Systems to predict the magnitude of earthquakes. *Bollettino di Geofisica Teorica ed Applicata*, 2015, vol. 56, no. 2, pp. 227–256.
10. Valverde-Albacete F. J., Pelaez-Moreno C. 100 % classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PloS One*, 2014, vol. 9, no. 1, p. e84217. <https://doi.org/10.1371/journal.pone.0084217>

Информация об авторах

Старовойтов Валерий Васильевич, доктор технических наук, профессор, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: valerys@newman.bas-net.by

Голуб Юлия Игоревна, кандидат технических наук, доцент, старший научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: 6423506@gmail.com

Information about the authors

Valery V. Starovoitov, Dr. Sci. (Eng.), Professor, Chief Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus.
E-mail: valerys@newman.bas-net.by

Yuliya I. Golub, Cand. Sci. (Eng.), Associate Professor, Senior Researcher, the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus.
E-mail: 6423506@gmail.com