ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

Системный анализ, управление и обработка информации (05.13.01)

УДК 004.052.4
DOI: 10.24160/1993-6982-2022-1-98-110

Критерий оценки качества классификации за пределами обучающей выборки

А.О. Гурина, В.Л. Елисеев

Исследована распространённая проблема классификации на основе моделей машинного обучения. Ввиду непредсказуемости классификации объектов за пределами обучающей выборки классификаторы могут работать некорректно на новых данных, а также быть уязвимы к состязательным атакам. Сделано предположение о том, что при достаточно полной оценке качества классификатора этих проблем можно избежать. Проанализирована эффективность применения традиционного подхода к оценке качества классификации. Описаны недостатки традиционных показателей качества, не позволяющие оценить риск возникновения ошибок и степень подверженности модели машинного обучения состязательным атакам. Предложен новый критерий качества классификации, включающий четыре показателя: Excess, Deficit, Coating, Approx (EDCA). Вычисление показателей основано на соотношении размеров области пространства, занимаемого обучающей выборкой, и результатов классификации всех точек дискретизированного пространства признаков в рабочем диапазоне их значений. Выполнено экспериментальное исследование визуальной оценки и сравнения качества двух многоклассовых SVM классификаторов на характерных синтетических наборах данных с помощью традиционных и предлагаемых показателей качества. Продемонстрированы эффективность и преимущество введенных показателей по сравнению с традиционными. Подтверждена хорошая интерпретируемость значений показателей качества, а также субъективное соответствие метрик ожидаемым результатам сравнения двух SVM классификаторов. Есть основания полагать, что применение нового подхода к оценке качества позволит строить более надёжные классификаторы на основе машинного обучения.

Ключевые слова: показатели качества классификации, количественная оценка качества, машинное обучение, состязательные атаки, достоверность, точность, полнота.

Для цитирования: Гурина А.О., Елисеев В.Л. Критерий оценки качества классификации за пределами обучающей выборки // Вестник МЭИ. 2022. № 1. С. 98—110. DOI: 10.24160/1993-6982-2022-1-98-110.

The Classification Quality Assessment Criterion Outside a Training Set

A.O. Gurina, V.L. Eliseev

The article addresses a commonly encountered problem of classification based on machine learning models. Given that attempts to classify objects outside the training sample are prone to yield unpredictable results, the classifiers may operate incorrectly on new data and may also be vulnerable to adversarial attacks. It is conjectured that these problems can be avoided provided that a sufficiently complete assessment of the classifier quality is made. The effectiveness of applying the conventional approach to estimating the classification quality is analyzed. Disadvantages of the conventional quality indicators, which do not allow one to evaluate the risk of errors and degree of machine learning model susceptibility to adversarial attacks, are described. A new classification quality criterion is proposed, which includes four characteristics: Excess, Deficit, Coating, and Approx (EDCA). The characteristics are quantified based on the ratio between the size of the space occupied by the training sample and the results of the classification of all points of the discretized space of features in the working range of their values.

An experimental study for visual assessment and comparison of the quality of two multiclass SVM classifiers on characteristic synthetic data sets using the conventional and proposed quality indicators is carried out. The effectiveness and advantage of the newly introduced indicators in comparison with the conventional ones is demonstrated. Good interpretability of the quality indicator values, as well as the subjective consistency between the metrics and expected results from comparison of two SVM classifiers is confirmed. There is a reason to believe that application of the new approach to quality assessment will make it possible to construct more reliable classifiers based on machine learning.

Key words: classification quality indicators, quality quantification, machine learning, adversarial attacks, reliability, accuracy, completeness.

For citation: Gurina A.O., Eliseev V.L. The Classification Quality Assessment Criterion Outside a Training Set. Bulletin of MPEI. 2022;1:98—110. (in Russian). DOI: 10.24160/1993-6982-2022-1-98-110.

Введение

Методы машинного обучения используются в системах принятия решений различных прикладных областей: для биометрической идентификации, в промышленности — для управления производством и обнаружения угроз безопасности, в медицине — для диагностики заболеваний, в маркетинге — для персонализированной рекламы, в сфере информационной безопасности — для обнаружения аномалий и кибератак и т. п. Так, в 2020 г. 34% компаний Европы, США и Китая применяли машинное обучение, а по оценкам экспертов к 2024 г. спрос на машинное обучение вырастет ещё на 42% [1].

Увеличение спроса на решения на основе машинного обучения связано с растущим внедрением облачных сервисов, ростом объема неструктурированных данных и потребностью в глобальной автоматизации процессов. Однако, согласно исследованию IBM [2], повсеместному внедрению машинного обучения препятствует недоверие к технологии.

Действительно, множество известных прецедентов некорректной работы машинного обучения [3 — 5] не позволяют полностью полагаться на этот инструмент при решении задач с высокой ценой ошибки. Очевидно, что в некоторых системах ошибки алгоритмов могут нанести серьезный ущерб. Справедливо, что для работы с технологией, призванной заменить умственный и физический труд человека, должна быть уверенность, что классификатор на основе машинного обучения даст предсказуемый и корректный ответ на любых данных, подаваемых на вход.

В связи с этим на первый план выходят критерии для оценки качества полученной модели машинного обучения, без которых, в условиях многомерных данных, невозможно оценить эффективность и надежность построенной модели или сравнить между собой два различных алгоритма классификации.

Широко распространенная задача, решаемая алгоритмами машинного обучения, — классификация, или отнесение наблюдаемого объекта к тому или другому классу для принятия последующего решения (автоматического или с помощью человека). Для оценки качества построенной модели классификатора используют следующие показатели:

- confusion matrix;
- accuracy;
- precision;
- recall;
- f-score.

Традиционный метод оценки классификатора путем сравнения значений общепринятых показателей качества с эталонными обладает одним серьезным ограничением — такая оценка в полной мере справедлива только для проверяемых объектов ограниченного тестового набора. Даже если показатели качества близки к эталонным на тестовом наборе, это не означает, что классификатор будет работать корректно на новых данных. Таким образом, есть основания полагать, что традиционные критерии качества неполны и не позволяют оценить качество классификатора за пределами набора обучающих и тестовых данных. Именно это и приводит к неожиданным ошибкам высококачественных, по мнению разработчиков, классификаторов.

Представляется актуальным провести исследование достаточности традиционных показателей качества для создания предсказуемых классификаторов.

Отметим, что известный метод SVM, как и многие классификаторы, после обучения делит пространство признаков на открытые области классов, что делает возможным формирование состязательных примеров и некорректный результат классификации примеров за пределами целевых классов. Разработка показателей качества, оценивающих величину и положение областей, формируемых классификатором после обучения, относительно целевых, позволила бы оценивать и минимизировать риск неправильной классификации за пределами обучающей выборки, а также повысить доверие к работе методов машинного обучения в прикладных задачах.

В [6] предложен способ классификации на основе автокодировщиков, позволяющий получать замкнутую область деформированного множества для каждого класса, а также управлять его величиной для точной классификации и выявления аномалий. Данный способ совместно с предлагаемым критерием качества позволит строить классификаторы, близкие к идеальным даже в пространстве высокой размерности.

Обзор литературы

Для оценки качества моделей машинного обучения и сравнения различных алгоритмов используют стандартные показатели качества, рассчитываемые на основе составляющих матрицы ошибок (confusion matrix) [7]. Матрица ошибок для оценки качества алгоритма двухклассовой классификации (binary classification)

дана в табл. 1, в ячейках которой указано количество фактов корректной (True) и некорректной (False) классификаций для каждого из классов, традиционно называемых положительными (Positive) и отрицательными (Negative).

Основные показатели, учитываемые при оценке качества алгоритма классификации — рассчитываемые на основе элементов матрицы:

● ошибки первого α и второго β родов, оценивающие долю ложных срабатываний алгоритма классификации (False Positive, FP) и долю ложно отвергнутых примеров (False Negative, FN):

$$\alpha = \frac{FP}{FP + TN};$$

$$\beta = \frac{FN}{FN + TP};$$

● мощность критерия — характеристика способности критерия не упустить значимое событие: 1 – β.

Большинством разработчиков систем на основе машинного обучения используется следующая методология оценки качества моделей:

● рассчитывается матрица ошибок;

● на основе элементов матрицы ошибок вычисляются показатели качества:

– доля правильно классифицированных примеров:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN};$$

– часть положительно классифицированных примеров, действительно являющихся положительными:

$$\text{Precision} = \frac{TP}{TP + FP};$$

– доля примеров положительного класса, распознанных среди всех примеров положительного класса:

$$\text{Recall} = \frac{TP}{TP + FN};$$

● исходя из представленных показателей качества уточняются агрегированные критерии качества — среднее гармоническое Precision и Recall:

$$F_1 = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}};$$

● на основе сравнения F_1 и эталонного значения, равного 1, принимается решение о достижении требуемого уровня качества модели.

Помимо приведенных показателей некоторыми исследователями вводятся и другие, более эффективно характеризирующие те или иные аспекты алгоритмов машинного обучения [7, 8]. Также для оценки качества используется так называемые ROC-кривые [9].

Несмотря на принятую методологию, известны случаи некорректной работы введенных в эксплуатацию алгоритмов классификации, высоко оцениваемых с её помощью. Согласно [4] большой класс классификаторов имеет фундаментальный недостаток — непредсказуемость результата классификации для объектов за пределами обучающей выборки. Досадные ошибки такого рода наблюдаются в системах распознавания объектов [10], компьютерного зрения [11], обработки естественного языка [12] и других системах на основе машинного обучения.

Таблица 1

Матрица ошибок

Классификация	**Класс**	
	Positive	**Negative**
Positive	*TP*	*FP*
Negative	*FN*	*TN*

Более того, встает вопрос защиты алгоритмов классификации от состязательных атак злоумышленников, целенаправленно формирующих примеры, вызывающие ошибки классификаторов [5]. Известно, что наклейки, прикрепленные к стандартному дорожному знаку остановки, заставляют систему компьютерного зрения автономного транспортного средства ошибочно идентифицировать его как знак ограничения скорости [11]. В [13] показано, что системы машинного обучения, получающие входные данные с камер и других датчиков, также уязвимы для состязательных атак. Для обмана системы распознавания лиц [14] достаточно надеть специальные очки. Приведенные примеры неудач машинного обучения показывают, что даже простые алгоритмы машинного обучения ведут себя совсем не так, как предполагают разработчики.

Исследователями были предложены некоторые методы борьбы с ошибками классификаторов, среди которых — состязательное обучение, автоэнкодеры. Однако ни один способ не может полностью защитить системы машинного обучения от атак этого типа, на текущий момент нет какого-либо общепринятого решения.

В [15] обсуждалось, что состязательные примеры неизбежны, особенно при высокой размерности данных. В связи с этим создание способов предотвращения ошибок классификаторов становится важной областью исследований. Необходимо, чтобы модели машинного обучения давали адекватные результаты для всех возможных входных данных.

Следует отметить, что возможная причина таких неожиданных ошибок в том, что на этапе оценки качества полученной модели не выполняется проверка наличия вблизи области целевого класса примеров, не относящихся к классу, но для которых ответ классификатора — положительный. Для данной проверки не обязательно строить ещё одну нейронную сеть для состязательного обучения, достаточно пересмотреть способ оценки качества моделей машинного обучения.

В большинстве случаев модели машинного обучения работают достаточно точно, но только на небольшом количестве входных данных относительно всех возможных, с которыми они могут столкнуться при эксплуатации. Оценка качества классификатора в большинстве практических задач происходит на основе расчёта показателей для тестового набора примеров, зачастую существенно меньшего ограниченного счетного обучающего множества, что обусловлено труднодоступностью наборов данных для обучения и тестирования. На обособленной известной части примеров можно добиться высокой точности классификатора согласно традиционным показателям, но значит ли это, что классификатор будет обеспечивать такое же высокое качество и принимать корректные решения на новых данных, которых не было в тестовой или обучающей выборке?

Очевидно, что традиционный подход к оценке качества классификаторов не может гарантировать полного доверия к показателям качества и результатам применения классификатора в практических задачах, особенно с дрейфом данных, в силу ограниченности набора данных для тестирования и расчета показателей качества на значениях элементов матрицы ошибок. Одним из возможных решений проблемы надежности использования моделей машинного обучения стала разработка нового более объективного подхода к оценке качества моделей классификации.

В [16] описана задача оценки качества одноклассового классификатора и предложены показатели для оценки качества, оперирующие не количеством примеров в выборке, а областями, покрывающими точки обучающего множества. Такой подход устранял недостатки, свойственные общеизвестным критериям качества классификации, и позволил повысить достоверность оценки качества и доверие к решениям классификаторов. Рассмотрим обобщение введенных в [16] показателей для оценки алгоритмов бинарной и многоклассовой классификации (далее — классификации).

Методология

Для оценки алгоритмов классификации обобщим введенный ранее критерий качества одноклассовой классификации [16] на любое количество классов. Новый критерий качества оперирует дискретными оценками объемов, занимаемых данными в пространстве признаков, и включает четыре новых показателя: Excess, Deficit, Coating, Approx (EDCA). В случае оценки бинарного или многоклассового классификаторов показатели рассчитывают для каждого класса.

Под оценкой объема данных понимается сумма объемов атомарных ячеек дискретного разбиения ограниченной области в пространстве признаков, таких, что в каждой из ячеек есть хотя бы одна из точек множества данных. Обозначим объем счетного ограниченного множества данных X^* с разбиением в области Ω на атомарные ячейки размером h как $|X^*|_{\Omega,\, h}$. Если все множества данных рассматривать в одной и той же области разбиения и с одним и тем же размером атомарной ячейки, то индекс Ω, h можно опустить. Область разбиения обозначим областью сканирования.

Рассчитаем показатели на основе объемов, занимаемых обучающим $|X_T^*|$ и классифицированным $|X_D^*|$ множествами в пространстве признаков, по следующим формулам:

$$\text{Excess} = \frac{\left|X_D^* \setminus X_T^*\right|}{\left|X_T^*\right|}; \quad (1)$$

$$\text{Deficit} = \frac{\left|X_T^* \setminus X_D^*\right|}{\left|X_T^*\right|}; \quad (2)$$

$$\text{Coating} = \frac{\left|X_T^* \cap X_D^*\right|}{\left|X_T^*\right|}; \quad (3)$$

$$\text{Approx} = \frac{\left|X_T^*\right|}{\left|X_D^*\right|}.$$

Под указанными объемами понимается сумма объемов атомарных элементов разбиения пространства, затронутых имеющимися элементами обучающего $\hat{X}_T$ и классифицированного $\hat{X}_D$ множеств. Классифицированное множество $\hat{X}_D$ — множество сгенерированных точек пространства сканирования, отнесенных обученным классификатором к целевому классу, т. е. деформированное классификатором целевое множество.

Для расчета показателя Excess объем $|X_D^* \setminus X_T^*|$ определяют как сумму объемов атомарных элементов разбиения пространства (ячеек), в которые входят элементы $\hat{X}_D$, за исключением ячеек, в которые входят также элементы $\hat{X}_T$.

Для расчета Deficit объем $|X_T^* \setminus X_D^*|$ находят как сумму объемов ячеек, затронутых элементами множества $\hat{X}_T$, но не затронутых элементами множества $\hat{X}_D$.

С целью расчета показателя Coating объем $|X_T^* \cap X_D^*|$ вычисляют как сумму объемов ячеек, затронутых как элементами множества $\hat{X}_T$, так и $\hat{X}_D$.

Особенность предложенного критерия, которую необходимо учитывать, заключается во влиянии выбора размера атомарной ячейки разбиения на значения показателей.

Введенные показатели интерпретируются следующим образом.

Если классификатор неправильно классифицирует объекты за пределами целевого класса, показатель Excess (избыток (англ.)) принимает значение больше 0 (аналог α).

Если классификатор неправильно классифицирует объекты в пределах целевого класса, показатель Deficit (недостаток (англ.)) станет больше 0 (аналог β).

Если классификатор правильно классифицирует все объекты в пределах целевого класса, показатель Coating принмет значение 1 (аналог 1 – β).

Точность классификатора с точки зрения аппроксимации целевого множества — Approx, оценивается как отношение объема целевого множества к объему множества, отнесенного обученным классификатором к целевому классу.

Идеальный одноклассовый классификатор характеризуется следующими значениями показателей:

$$\text{Excess} = 0;\ \text{Deficit} = 0;\ \text{Coating} = 1;\ \text{Approx} = 1. \quad (4)$$

Поскольку предложенный критерий легко обобщить на любое число классов, то методология оценки бинарного или многоклассового классификаторов выглядит так:

- сканирование расширенной области обучающего пространства с некоторым шагом сетки h;
- получение ответов классификатора для примеров обучающего множества $\hat{X}_T$;
- расчет объема $|\hat{X}_T|$ для каждого класса;
- получение ответов классификатора для примеров множества сканирования;
- расчет объема $|\hat{X}_D|$ для каждого класса;
- расчет показателей Approx, Excess, Deficit, Coating для каждого класса;
- принятие решения относительно качества классификатора путем сравнения значений показателей качества для каждого класса со значениями, установленными в (4).

В отличие от традиционного подхода, качество классификатора оценивается не по агрегированным показателям качества классификации объектов всех классов, а по четырем показателям, отражающим точность классификации объектов каждого класса в отдельности. Данный подход представляется более точным и надежным.

Эксперименты

Для наглядного сравнения объективности введенных показателей с традиционными поставлен эксперимент по оценке качества методов SVM с разными параметрами по двум методологиям:

- традиционной: матрице ошибок, Recall, Precision и F-score для классификатора;
- предлагаемой: Excess, Deficit, Coating, Approx (EDCA) для каждого класса.

В качестве исходных данных взят сгенерированный набор двумерных данных и разделен на обучающее и тестовое множества. Обучающее множество состоит из трех классов, изображенных на рис. 1.

Для классификации объектов обучено три классификатора SVM с радиально-базисными функциями ядра:

- SVM1 с gamma = 0,0001 (рис. 2);
- SVM2 с gamma = 900 (рис. 3).

Параметр gamma в методе SVM определяет, какое влияние при построении «идеальной» разделяющей линии имеют далеко находящиеся элементы обучающего набора данных. Чем ниже gamma, тем больше элементов, достаточно далеких от разделяющей линии, принимают участие в процессе определения линии. При высоком значении gamma алгоритм опирается только на наиболее близкие к линии элементы. Следу-

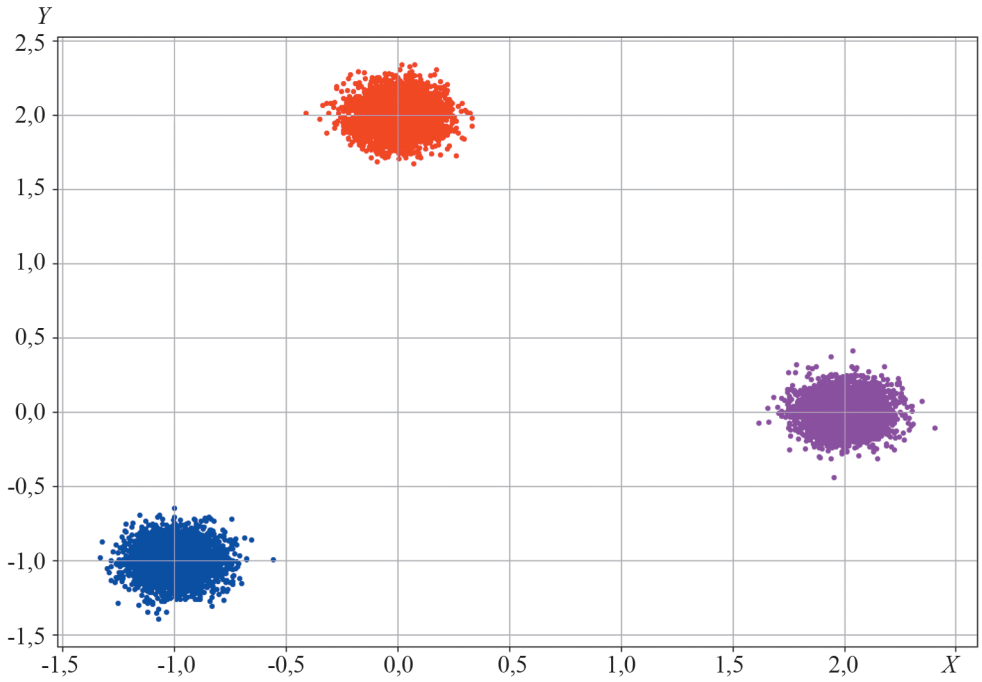


Рис. 1. Обучающие данные:
● — класс № 1; ● — класс № 2; ● — класс № 3

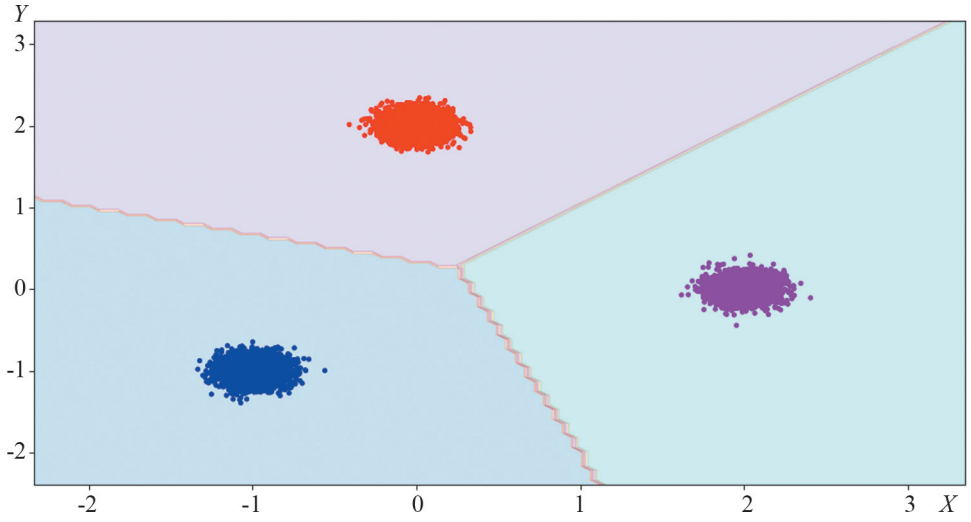


Рис. 2. Разделение пространства признаков обученным классификатором SVM1

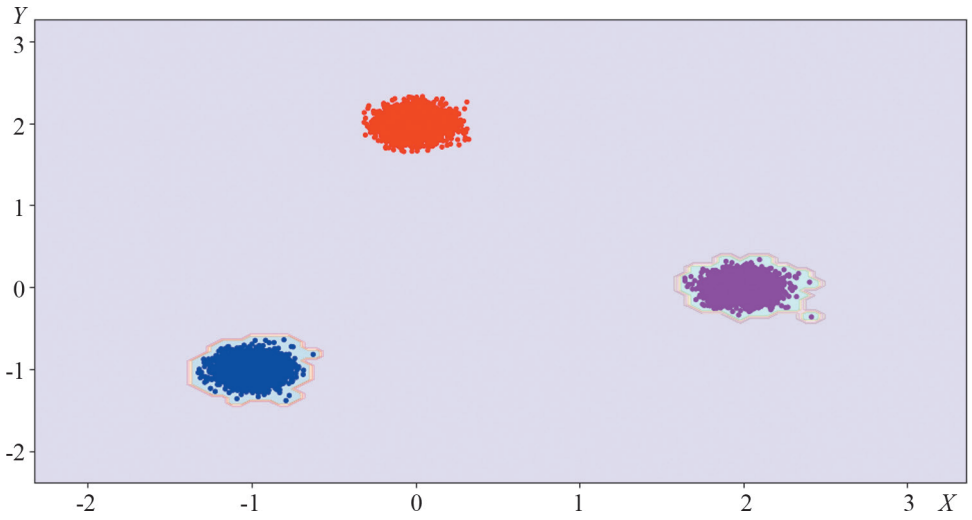


Рис. 3. Разделение пространства признаков обученным классификатором SVM2

ет отметить, что высокие значения gamma приводят к переобучению.

Как следует из рис. 2, при низком значении параметра gamma всё пространство признаков делится на три открытые области, каждая из которых — отдельный класс. Отметим, что при использовании такого классификатора объекты, не относящиеся ни к одному из классов, классифицируются ошибочно.

При высоком значении gamma для данных двух классов получаются две замкнутые области, при этом точки остального пространства отходят к третьему классу, что также в большинстве случаев является некорректным решением. Очевидно, что идет переобучение для двух классов, поэтому точки за пределами выстроенных границ классифицированы неверно.

Результаты оценки качества трех классификаторов на тестовом наборе получены согласно традиционной методологии и сведены в табл. 2.

Согласно традиционным показателям оценки качества, все классификаторы имеют одинаково высокую точность и служат для решения практической задачи.

Оценим качество трех классификаторов по новой методологии: оценка качества каждого классификатора определяется качеством классификации объектов каждого класса. Рассчитанные значения новых показателей качества даны в табл. 3.

Несмотря на то, что традиционные оценки качества показали, что оба классификатора обеспечивают высокую точность, по значениям показателей Excess и Approx видно, что точность классификаторов разная, как и точность классификации каждого класса. Качество первого классификатора — неудовлетворительно, поскольку значения Excess для всех классов велики, следловательно, классификатор неправильно классифицирует объекты за пределами целевого класса (ошибочная классификация точек, не относящихся к целевым классам), а значения показателя Approx, наоборот, близки к 0 для всех классов, что демонстрирует низкий уровень аппроксимации классификатором

Таблица 2

Результаты традиционного метода оценки качества

Классификаторы		Традиционные показатели оценки качества					
SVM	Gamma	F-score	Precision	Recall	Матрица ошибок		
SVM1	0,0001	1,00	1,00	1,00	1000	0	0
					0	1000	0
					0	0	1000
SVM2	900	0,99	0,99	0,99	999	0	1
					0	998	2
					0	0	1000

Таблица 3

Результаты нового метода оценки качества EDCA

Классификаторы		Новые показатели оценки качества				
SVM	Gamma	Номер класса	Excess	Deficit	Coating	Approx
SVM1	0,0001	1	8,50	0	1	0,10
		2	13,40	0	1	0,07
		3	11,60	0	1	0,07
SVM2	900	1	0	0	1	1
		2	0,12	0	1	0,89
		3	30,90	0	1	0,03

целевого множества. Кроме того, высокие значения показателя Approx для 1 и 2 классов подтверждают, что увеличение параметра обучения gamma приводит к высокой степени аппроксимации классификатором области обучающего множества, т. е., только по значениям показателя Approx можно сделать вывод о том, насколько границы класса, выстроенные классификатором, близки к фактическим границам области обучающего множества. Эта информация особенно ценна в многомерном пространстве признаков, когда привычная визуализация обучающего множества и границ затруднительна, а решение о достижении требуемого уровня качества принимается только на основе числовых показателей качества.

Таким образом, выводы, сделанные благодаря анализу значений введённых показателей качества, полностью соответствуют действительному качеству классификаторов, визуально оцениваемому по рис. 2, 3. Поскольку традиционные показатели качества не дают такой точной оценки качества классификаторов, то можно утверждать, что введенные показатели качества более информативны и объективны.

Подробно рассмотрим этапы расчета показателей качества.

Этап 1. Сканирование пространства и определение размера $|X_T^*|$ и $|X_D^*|$.

Поскольку размерность пространства признаков — 2, то для сканирования пространства построим сетку с одинаковым шагом h по двум осям в расширенной области значений признаков (рис. 4, 5).

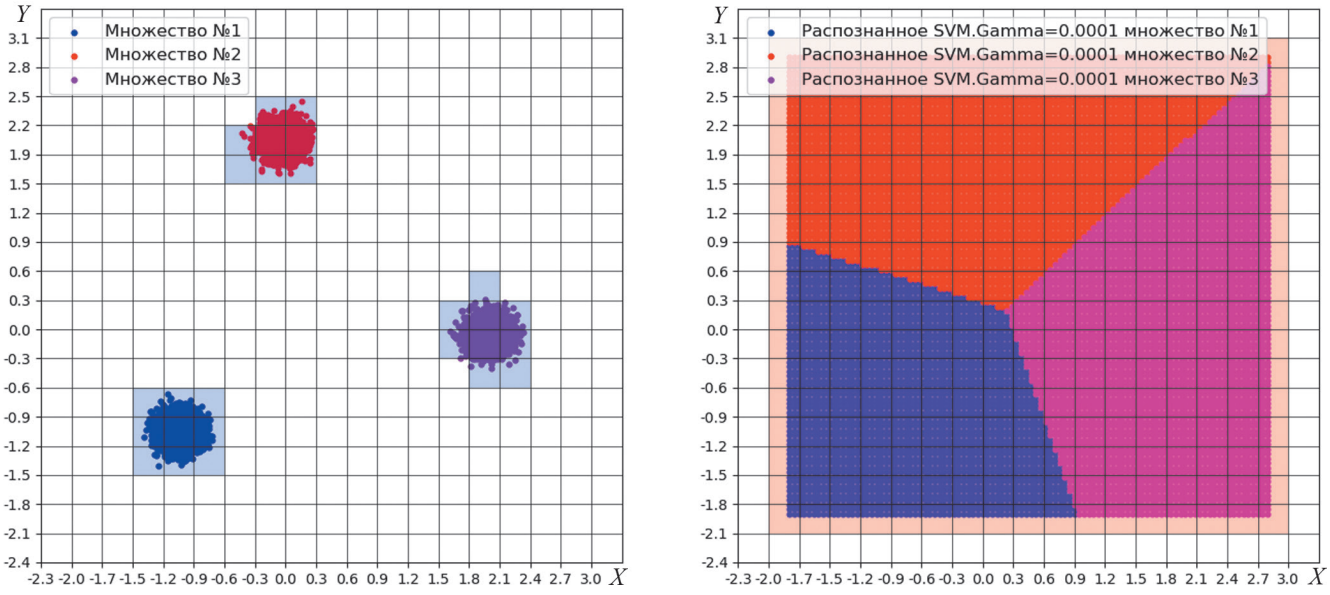
Из рис. 4, *а* следует, что обучающие примеры первого класса занимают 9 ячеек сетки, соответственно, объем области $|X_T^*|$ рассчитывается как площадь совокупности этих ячеек: $9h^2$. Примеров пространства сканирования, отнесенных классификатором SVM1 к первому классу, гораздо больше — они занимают 86 ячеек сетки (см. рис. 4, *б*), и размер области $|X_D^*|$ составляет $86h^2$.

Для сравнения, площадь классифицированного SVM2 множества $|X_D^*|$ для первого класса полностью совпадает с площадью обучающего $|X_T^*|$, а вот для третьего класса при $|X_T^*| = 9h^2$ $|X_D^*| = 272h^2$ (см. рис. 5). Такая разница в размерах областей целевого класса и данных, относимых классификатором к целевому классу, говорит об избыточном обобщении и риске ошибок классификатора.

Этап 2. Расчет и сравнение значений показателей качества EDCA

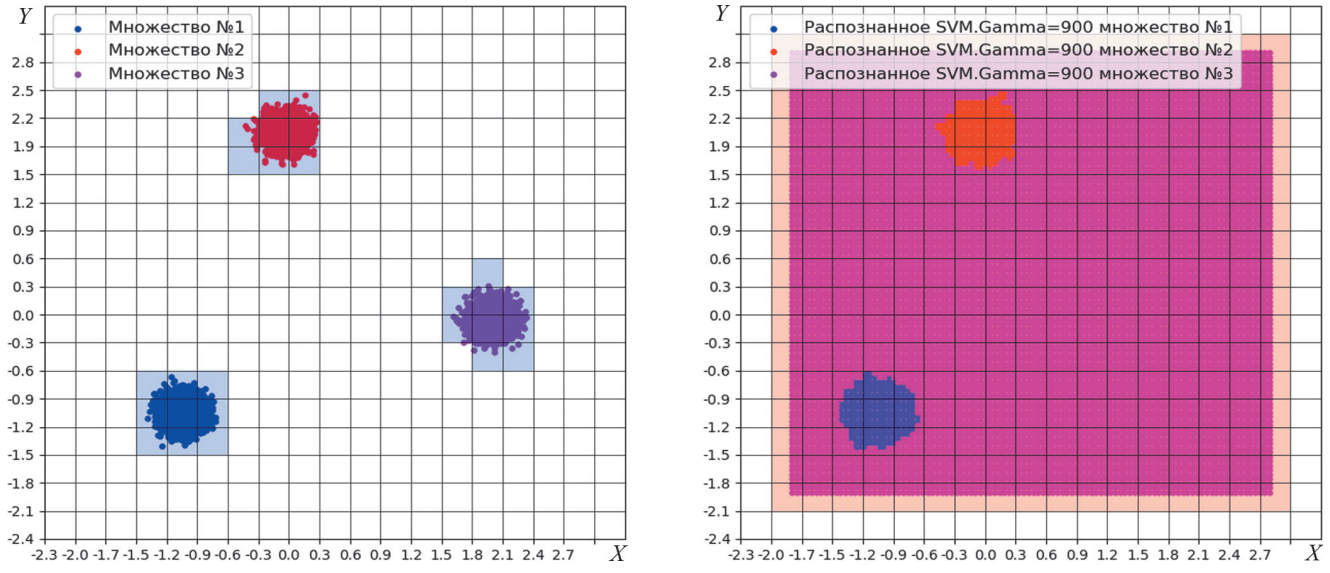
Для расчета величины показателя Excess, согласно (1), необходимо определить область, включающую X_D^* за исключением области X_T^*, а затем найти её отношение к области обучающего множества X_T^*. На рисунках 6 — 11 продемонстрированы области X_T^* и X_D^*, искомая область — заштрихована. На рисунке 6, *а* область $X_D^* \backslash X_T^*$ заштрихована.

Для расчета показателя Deficit необходим размер области, включающей X_T^* и не содержащей X_D^*. Поскольку область, построенная классификатором, полностью охватывает область обучающего множества, то такой области нет (см. рис. 6, *б*), а показатель, согласно (2), равен 0.



а *б*

Рис. 4. Дискретные области обучающего X_T^* и деформированного X_D^* множеств (SVM1)



а *б*

Рис. 5. Дискретные области обучающего X_T^* и деформированного X_D^* множеств (SVM2)

Области X_T^* и X_D^* пересекаются только в области X_T^*, и, согласно (3), величина Coating достигает максимального значения –1.

По рисунку 6, *а* можно оценить отношение размеров областей $|X_D^* \backslash X_T^*|$ и $|X_T^*|$ показателя Excess. Величина показывает, насколько область, построенная классификатором, больше области целевого класса, и насколько велик риск ошибок первого рода α. Чем больше значение показателя Excess, тем больше данных, для которых велик риск ложного срабатывания классификатора.

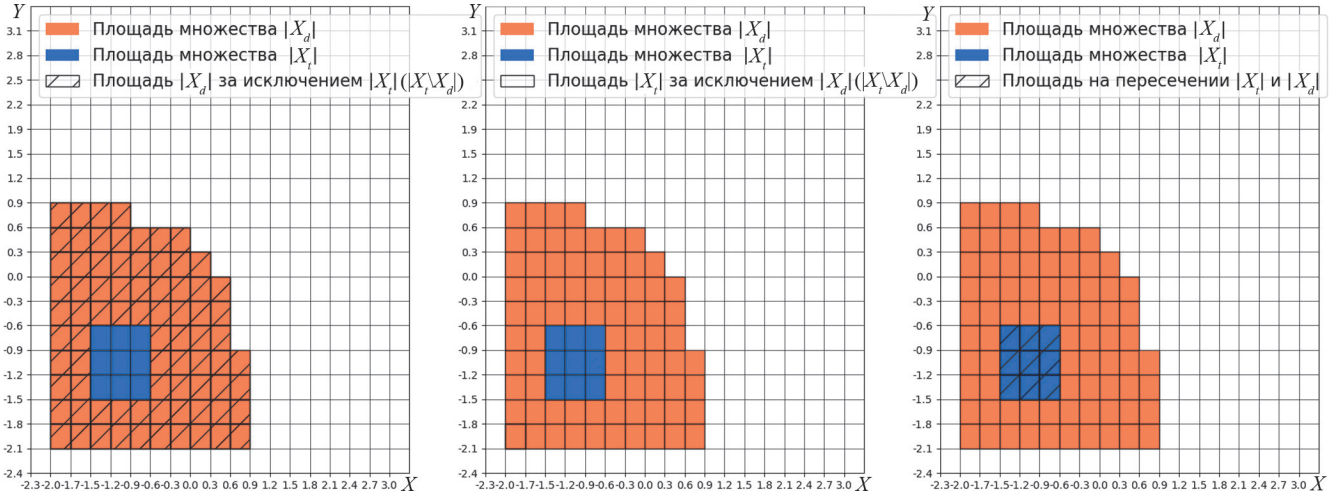
Из рисунка 6, *б* следует, что величина Deficit = 0. Это означает нулевую вероятность ошибки второго рода β внутри обучающего множества.

На рисунке 7 даны аналогичные данные для расчета показателей качества классификации объектов первого класса классификатором SVM2.

Можно заметить, что классификатор построил границы класса в полном соответствии с целевым множеством, а все величины приняли идеальные значения (см. (4)).

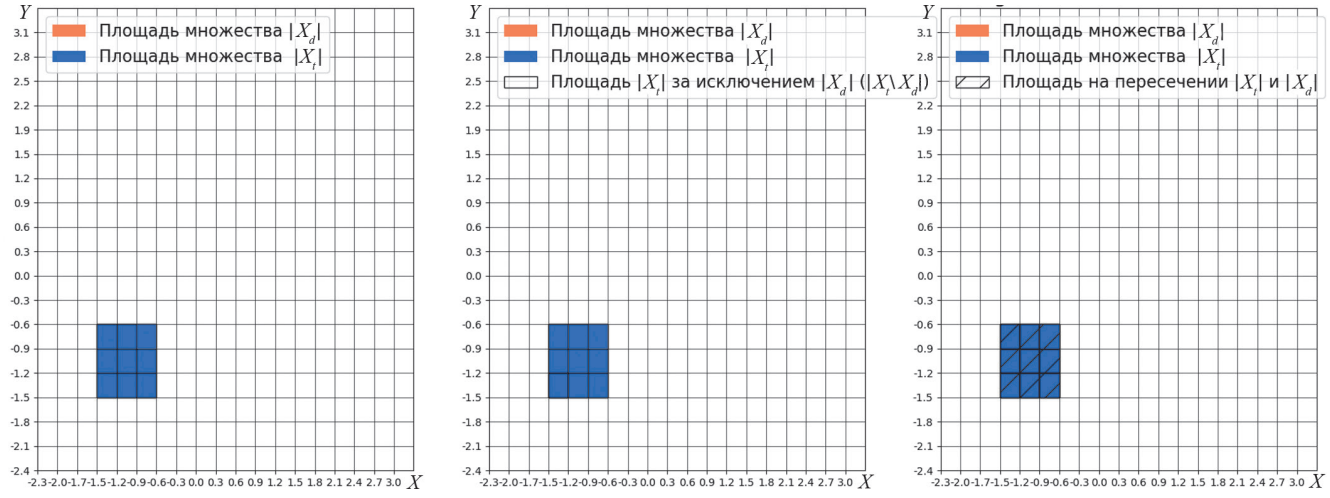
Соответствующим образом по рис. 8 — 11 сравним показатели качества классификации оставшихся двух классов, обеспечиваемые классификаторами SVM1 и SVM2.

Таким образом, с помощью введённых показателей можно рассчитать объем дискретных областей и оценить, насколько деформированная область каждого



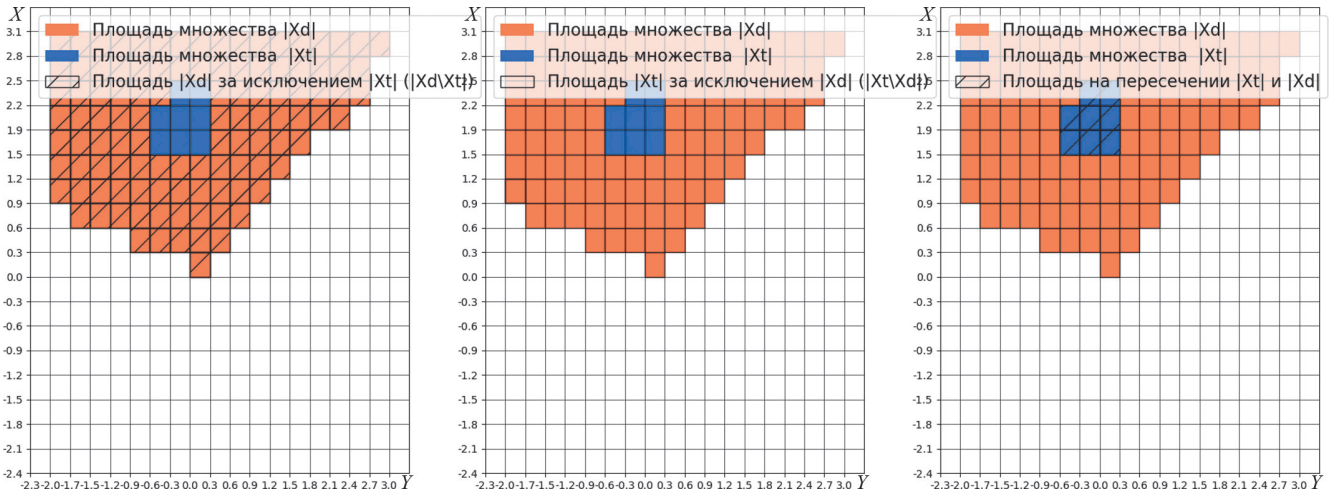
а *б* *в*

Рис. 6. Визуализация дискретных областей, необходимых для расчета показателей Excess (*а*), Deficit (*б*) и Coating (*в*) (SVM1. Gamma = 0,0001. Класс 1)



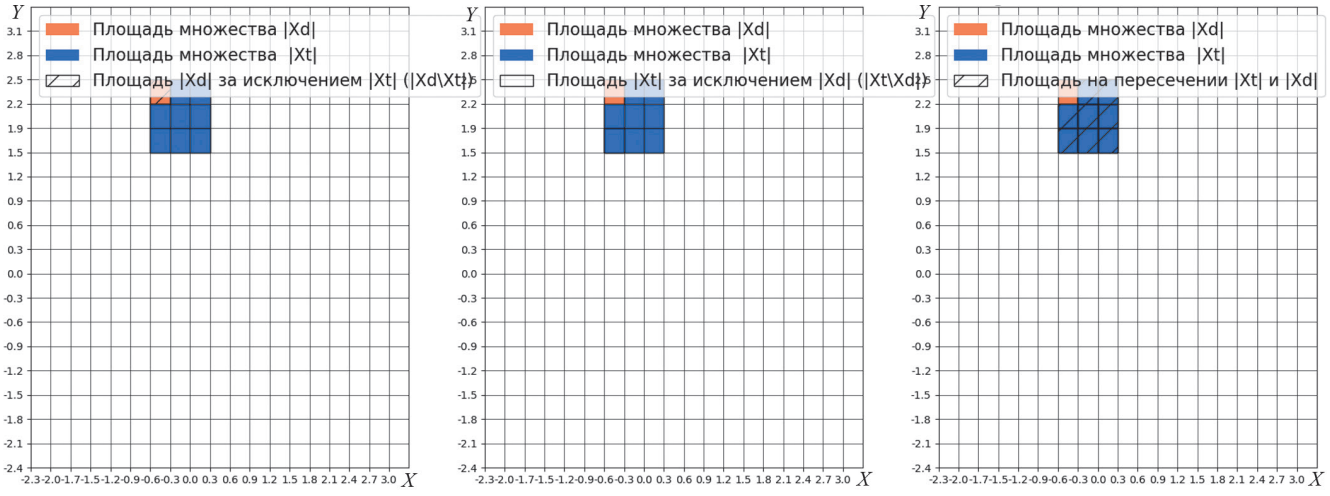
а *б* *в*

Рис. 7. Визуализация дискретных областей, необходимых для расчета показателей Excess (*а*), Deficit (*б*) и Coating (*в*) (SVM2. Gamma = 900. Класс 1)



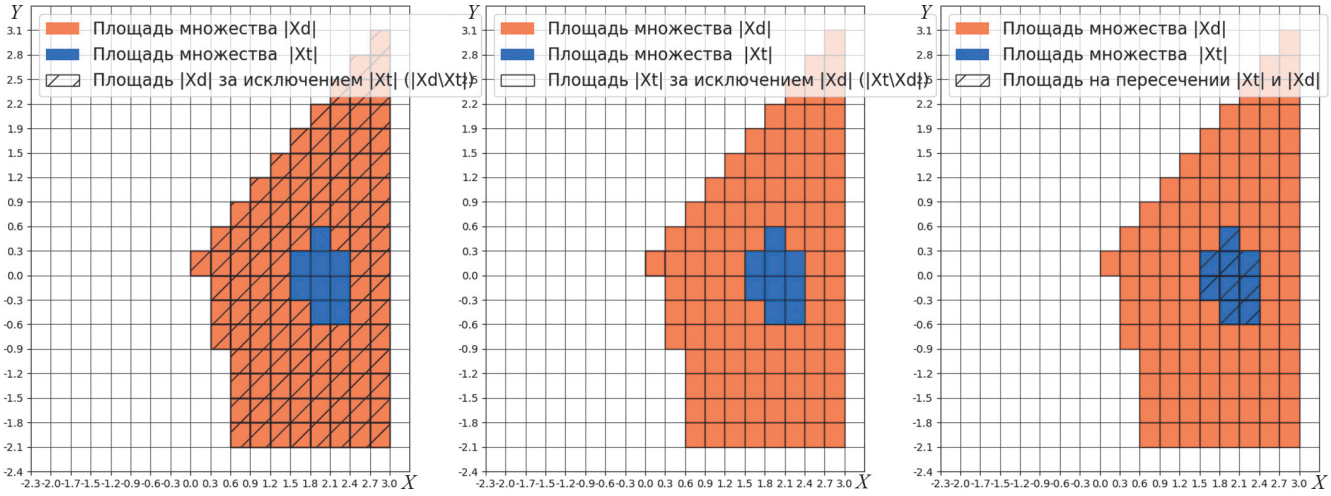
а *б* *в*

Рис. 8. Визуализация дискретных областей, необходимых для расчета показателей Excess (*а*), Deficit (*б*) и Coating (*в*) (SVM1. Gamma = 0,0001. Класс 2)



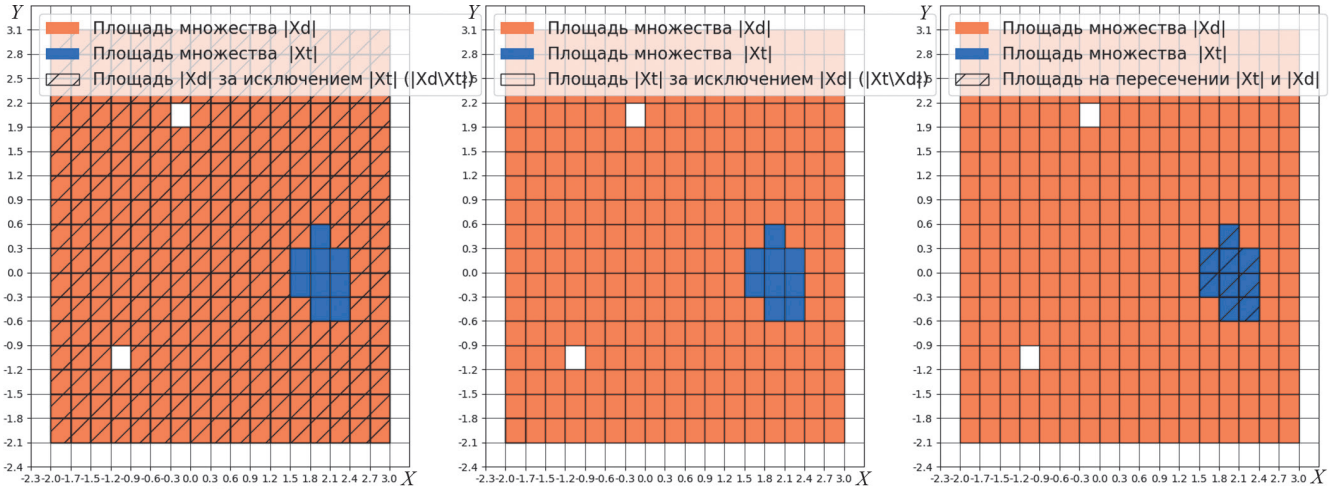
а *б* *в*

Рис. 9. Визуализация дискретных областей, необходимых для расчета показателей Excess (*а*), Deficit (*б*) и Coating (*в*) (SVM2. Gamma = 900. Класс 2)



а *б* *в*

Рис. 10. Визуализация дискретных областей, необходимых для расчета показателей Excess (*а*), Deficit (*б*) и Coating (*в*) (SVM1. Gamma = 0,0001. Класс 3)



а *б* *в*

Рис. 11. Визуализация дискретных областей, необходимых для расчета показателей Excess (*а*), Deficit (*б*) и Coating (*в*) (SVM2. Gamma = 900. Класс 3)

класса, построенная классификатором после обучения, не соответствует идеалу — области, занимаемой точками представительной обучающей выборки. Ценность такой оценки особенно возрастает в пространствах высокой размерности.

Анализ полученных результатов эксперимента подтверждает полезность введенных показателей для оценки качества классификаторов, а также преимущество по сравнению с традиционным подходом к оценке качества классификации.

Обсуждение

Предложенный подход к оценке качества классификации представляется более объективным и информативным по сравнению с общеизвестными критериями качества классификации, основной недостаток которых в неполноте оценки. Они позволяют оценить качество только внутри обучающего и тестового множеств и не дают возможности оценить риски ошибок классификаторов на новых данных.

Преимущество и уникальность предлагаемого подхода в том, что он оперирует не количеством примеров в обучающей и тестовой выборках, а областями, покрываемыми точками обучающего и деформированного множеств. Введенные показатели позволяют учитывать качество классификации объектов каждого класса даже в многомерных пространствах, когда невозможно визуализировать области обучающих и классифицируемых точек.

Можно предположить, что новый критерий качества классификации в виде совокупности показателей Approx, Excess, Deficit, Coating заслуживает большего доверия на этапе тестирования и позволит повысить доверие к результатам классификации. Указанные показатели можно использовать для сравнения, целенаправленной оптимизации бинарных и многоклассовых классификаторов, а также для управления границами, формируемыми классификаторами на этапе обучения. Это позволит строить классификаторы повышенной надежности.

В то же время, нельзя не отметить объективные недостатки подхода:

- необходимость расчета показателей для каждого класса;
- экспоненциальный рост вычислительной сложности и потребляемой памяти с ростом размерности пространства признаков X;
- высокая вычислительная сложность сканирования пространства для определения деформированного множества X_D^*, растущая экспоненциально с уменьшением шага сетки разбиения;
- экспоненциально увеличивающийся с уменьшением шага сетки объём памяти для выполнения операций над дискретизированными множествами;
- зависимость величин рассчитываемых показателей от шага сетки.

Вычислительная сложность определения X_T^* линейно зависит от размерности обучающего множества $\hat{X}_T$.

Заключение

Рассмотрены проблема классификации за пределами обучающей выборки и недостатки традиционных показателей оценки качества классификации. Сформулирован новый критерий качества для бинарных и многоклассовых классификаторов, позволяющий оценивать качество классификации объектов каждого класса в отдельности на основе теоретико-множественных операций в предположении о непрерывности целевых и классифицированных множеств. Качество классификации объектов каждого класса характеризуется показателями Excess, Deficit, Coating, Approx (EDCA), основанными на оценке объемов множеств в пространстве признаков. Подтверждена хорошая интерпретируемость значений показателей, продемонстрированы преимущества и большая точность введенных показателей относительно общепринятых критериев качества.

Литература

1. **Шпрингер Е.** 17 примеров применения машинного обучения в 5 отраслях бизнеса // Cloud Solutions [Электрон. ресурс] www.mcs.mail.ru/blog/17-primerov-mashinnogo-obucheniya. (дата обращения 03.06.2021).

2. **From Roadblock** to Scale: The Global Sprint Towards AI. New Research Commissioned by IBM in Partnership with Morning Consult. [Электрон. ресурс] www.filecache.mediaroom.com/mr5mr_ibmnews/183710/Roadblock-to-Scale-exec-summary.pdf (дата обращения 03.06.2021).

3. **Pike S.** Почему одного только машинного обучения недостаточно [Электрон. ресурс]

References

1. **Shpringer E.** 17 Primerov Primeneniya Mashinnogo Obucheniya v 5 Otraslyakh Biznesa. Cloud Solutions [Elektron. Resurs] www.mcs.mail.ru/blog/17-primerov-mashinnogo-obucheniya. (Data Obrashcheniya 03.06.2021). (in Russian).

2. **From Roadblock** to Scale: The Global Sprint Towards AI. New Research Commissioned by IBM in Partnership with Morning Consult. [Elektron. Resurs] www.filecache.mediaroom.com/mr5mr_ibmnews/183710/Roadblock-to-Scale-exec-summary.pdf (Data Obrashcheniya 03.06.2021).

3. **Pike S.** Pochemu Odnogo Tol'ko Mashinnogo Obucheniya Nedostatochno [Elektron. Resurs] www.kaspersky.

www.kaspersky.ru/blog/ai-fails/18678/ (дата обращения 03.06.2021).

4. **Goodfellow J.I., Shlens J., Sze Ch.** Explaining and Harnessing Adversarial Examples. Mountain View: Google Inc., 2015.

5. **Hern A.** Want to Beat Facial Recognition? Get Some Funky Tortoiseshell Glasses. [Электрон. ресурс] www.theguardian.com/technology/2016/nov/03/how-funky-tortoiseshell-glasses-can-beat-facial-recognition (дата обращения 03.06.2021).

6. **Гурина А.О., Елисеев В.Л.** Нейросетевой метод классификации в условиях нестационарного множества классов // Информационные системы и технологии: Материалы XXVI Междунар. науч.-техн. конф. Н. Новгород : Изд-во Нижегородского гос. техн. ун-та им. Р.Е. Алексеева, 2020. С. 750—764.

7. **Forman G.** An Extensive Empirical Study of Feature Selection Metrics for Text Classification // J. Machine Learning Research. 2003. V. 3. Pp. 1287—1305.

8. **Powers D.** Evaluation: From Precision, Recall and F-Factor to ROC. Techn. Rep. Informedness, Markedness&Correlation, 2007.

9. **Fawcett T.** An Introduction to ROC Analysis // Pattern Recognition Letters. 2006. V. 27. Pp. 861—874.

10. **Szegedy Ch. e. a.** Intriguing Properties of Neural Networks // Proc. Computer Vision and Pattern Recognition Conf. 2014. Pp. 248—255.

11. **Harris M.** Researchers Find a Malicious Way to Meddle with Autonomous Cars. [Электрон. ресурс] www.caranddriver.com/news/a15340148/researchers-find-a-malicious-way-to-meddle-with-autonomous-cars 2 (дата обращения 03.06.2021).

12. **Robin J., Liang P.** Adversarial Examples for Evaluating Reading Comprehension Systems. Computation and Language [Электрон. ресурс] www.arxiv.org/pdf/1707.07328.pdf (дата обращения 03.06.2021).

13. **Kurakin A., Goodfellow I., Bengio S.** Adversarial Examples in the Physical World [Электрон. ресурс] www.arxiv.org/abs/1607.02533 (дата обращения 03.06.2021).

14. **Mahmood S. e. a.** Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-art Face Recognition // Proc. ACM SIGSAC Conf. 2016. Pp. 1528—1540.

15. **Shafahi A. e. a.** Are Adversarial Examples Inevitable? [Электрон. ресурс] www.arxiv.org/pdf/1809.02104.pdf (дата обращения 03.06.2021).

16. **Гурина А.О., Елисеев В.Л.** Эмпирический критерий качества одноклассовой классификации // Информационные системы и технологии: Материалы XXVII Междунар. науч.-техн. конф. Н. Новгород: Изд-во Нижегородского гос. техн. ун-та им. Р.Е. Алексеева, 2021. С. 673—682.

ru/blog/ai-fails/18678/ (Data Obrashcheniya 03.06.2021). (in Russian).

4. **Goodfellow J.I., Shlens J., Sze Ch.** Explaining and Harnessing Adversarial Examples. Mountain View: Google Inc., 2015.

5. **Hern A.** Want to Beat Facial Recognition? Get Some Funky Tortoiseshell Glasses. [Elektron. Resurs] www.theguardian.com/technology/2016/nov/03/how-funky-tortoiseshell-glasses-can-beat-facial-recognition (Data Obrashcheniya 03.06.2021).

6. **Gurina A.O., Eliseev V.L.** Neyrosetevoy Metod Klassifikatsii v Usloviyakh Nestatsionarnogo Mnozhestva Klassov. Informatsionnye Sistemy i Tekhnologii: Materialy XXVI Mezhdunar. Nauch.-tekhn. Konf. N. Novgorod: Izd-vo Nizhegorodskogo Gos. Tekhn. Un-ta im. R.E. Alekseeva, 2020:750—764. (in Russian).

7. **Forman G.** An Extensive Empirical Study of Feature Selection Metrics for Text Classification. J. Machine Learning Research. 2003;3:1287—1305.

8. **Powers D.** Evaluation: From Precision, Recall and F-Factor to ROC. Techn. Rep. Informedness, Markedness&Correlation, 2007.

9. **Fawcett T.** An Introduction to ROC Analysis. Pattern Recognition Letters. 2006;27:861—874.

10. **Szegedy Ch. e. a.** Intriguing Properties of Neural Networks. Proc. Computer Vision and Pattern Recognition Conf. 2014:248—255.

11. **Harris M.** Researchers Find a Malicious Way to Meddle with Autonomous Cars. [Elektron. Resurs] www.caranddriver.com/news/a15340148/researchers-find-a-malicious-way-to-meddle-with-autonomous-cars 2 (Data Obrashcheniya 03.06.2021).

12. **Robin J., Liang P.** Adversarial Examples for Evaluating Reading Comprehension Systems. Computation and Language [Elektron. Resurs] www.arxiv.org/pdf/1707.07328.pdf (Data Obrashcheniya 03.06.2021).

13. **Kurakin A., Goodfellow I., Bengio S.** Adversarial Examples in the Physical World [Elektron. Resurs] www.arxiv.org/abs/1607.02533 (Data Obrashcheniya 03.06.2021).

14. **Mahmood S. e. a.** Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-art Face Recognition. Proc. ACM SIGSAC Conf. 2016:1528—1540.

15. **Shafahi A. e. a.** Are Adversarial Examples Inevitable? [Elektron. Resurs] www.arxiv.org/pdf/1809.02104.pdf (Data Obrashcheniya 03.06.2021).

16. **Gurina A.O., Eliseev V.L.** Empiricheskiy Kriteriy Kachestva Odnoklassovoy Klassifikatsii. Informatsionnye Sistemy i Tekhnologii: Materialy XXVII Mezhdunar. Nauch.-tekhn. Konf. N. Novgorod: Izd-vo Nizhegorod-skogo Gos. Tekhn. Un-ta im. R.E. Alekseeva, 2021. C. 673—682. (in Russian).

Сведения об авторах:

Гурина Анастасия Олеговна — аспирантка кафедры управления и интеллектуальных технологий НИУ «МЭИ», e-mail: asya.gurina001512@yandex.ru
Елисеев Владимир Леонидович — кандидат технических наук, руководитель центра научных исследований и перспективных разработок АО «ИнфоТеКС», доцент кафедры управления и интеллектуальных технологий НИУ «МЭИ», e-mail: vlad-eliseev@mail.ru

Information about authors:

Gurina Anastasiya O. — Ph.D-student of Control and Intelligent Technologies Dept., NRU MPEI, e-mail: asya.gurina001512@yandex.ru
Eliseev Vladimir L. — Ph.D. (Techn.), Head of the Center for Scientific Research and Advanced Development of JSC «InfoTeCS», Assistant Professor of Control and Intelligent Technologies Dept., NRU MPEI, e-mail: vlad-eliseev@mail.ru

Работа выполнена при поддержке: РФФИ (проект № 20-37-90073)
The work is executed at support: RFBR (Project No. 20-37-90073)

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов
Conflict of interests: the authors declare no conflict of interest

Статья поступила в редакцию: *21.07.2021*
The article received to the editor: *21.07.2021*