НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
сентябрь–октябрь 2023 Том 23 № 5 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
September–October 2023 Vol. 23 No 5 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2023-23-5-1030-1040
УДК 004.852

Доверительные липшицевы классификаторы: инструмент гарантированной надежности

Андрей Владимирович Тимофеев✉

ТОО «Эквалайзум», Астана, 010000, Казахстан
timofeev.andrey@gmail.com✉, https://orcid.org/0000-0001-7212-5230

Аннотация

Введение. Предложен новый метод гарантированного решения задачи мультиклассовой классификации стохастических объектов. **Метод.** В рамках разработанного подхода результат классификации представляет собой конечное множество индексов классов, которое с заданным начальным коэффициентом доверия содержит индекс того класса, которому соответствует классифицируемый объект. При этом классификация реализована на базе использования классификатора нового типа, который назван доверительным липшицевым классификатором. **Основные результаты.** Дано определение доверительного липшицева классификатора и изучены его основные свойства. В том числе исследовано свойство гарантированной надежности классификации, которое выражено в построении доверительного множества ограниченного размера, которое содержит индекс истинного класса с наперед заданным коэффициентом доверия. Приведен случай ассамблеи липшицевых классификаторов, свойства которой оформлены в виде теоремы. Рассмотрен практически важный пример использования предложенного подхода в задачах компенсации динамики шумового процесса в каналах оптоволоконной системы мониторинга. **Обсуждение.** Разработанный подход перспективен для применения в тех классификационных задачах, в которых число классов имеет порядок выше второго. В том числе в широкомасштабных системах биометрической идентификации личности, а также в многоканальных системах мониторинга протяженных объектов.

Ключевые слова

доверительный липшицев классификатор, машинное обучение, гарантированная надежность, оптоволоконная система мониторинга

Ссылка для цитирования: Тимофеев А.В. Доверительные липшицевы классификаторы: инструмент гарантированной надежности // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 5. С. 1030–1040. doi: 10.17586/2226-1494-2023-23-5-1030-1040

Confidence Lipschitz classifiers: an instrument of guaranteed reliability

Andrey V. Timofeev✉

Equalizum LLP, Astana, Kazakhstan, 010000
timofeev.andrey@gmail.com✉, https://orcid.org/0000-0001-7212-5230

Abstract

A new method of guaranteed solution for multiclass classification problem of stochastic objects is proposed. Within the framework of the proposed approach, the classification result is a finite set of class indices which with a predetermined confidence coefficient contains the index of the class to which the object being classified corresponds. In this case, the classification itself is realized on the basis of using a classifier of the new type which is called a confidence Lipschitz classifier. The definition of the confidence Lipschitz classifier is given and its main properties have been studied. Among them, the property of guaranteed reliability of the classification which is expressed in the construction of a confidence set of limited size containing the index of the true class with a predetermined coefficient of confidence, has been studied. The case of the assembly of Lipschitz classifiers, the properties of which are formalized in the form of a theorem, is considered. We consider a practically important example of using the proposed approach in the problems of compensation of the noise process dynamics in the channels of the fiber-optic monitoring system. The proposed approach is promising

for use in those classification tasks in which the number of classes has an order higher than the second, including large-scale biometric identification systems as well as multi-channel systems for monitoring extended objects.

Keywords
confidence Lipschitz classifier, machine learning, guaranteed reliability, fiber optic monitoring system

For citation: Timofeev A.V. Confidence Lipschitz classifiers: an instrument of guaranteed reliability. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 5, pp. 1030–1040 (in Russian). doi: 10.17586/2226-1494-2023-23-5-1030-1040

Введение

В настоящее время существует высокий практический спрос на мультиклассовые системы автоматической классификации, когда мощность множества подлежащих различению (идентификации) классов может достигать сотен тысяч или даже миллионов единиц. В этом случае возникает проблема необходимости выбора единственного варианта решения из множества близких, практически неразличимых альтернатив в ситуации, когда формирование верхней границы для мощности плохоразличимых альтернатив затруднительно ввиду отсутствия формально обоснованного механизма сравнения альтернатив между собой. Типичным примером систем такого рода являются биометрические системы национального масштаба, предназначенные для идентификации личности по ее биометрическим параметрам и лежащие в основе различных криминалистических комплексов, а также систем биометрической паспортизации населения. Работая с представительными корпусами реальных биометрических образцов, исследователи постоянно сталкиваются с проблемой принципиальной биометрической неразличимости, т. е. с ситуацией, когда у группы, состоящей из двух, или более физических лиц, первичные биометрические признаки оказываются практически неразличимыми в рамках используемой системы автоматической классификации. Природа этого явления многогранна и исследована в соответствующих разделах науки, но отметим, что ни одна из известных биометрических мод не дает 100 % надежной идентификации личности. Исследователи постоянно совершенствуют алгоритмическое обеспечение, выделяют все более информативные первичные признаки, используют репрезентативные корпуса для обучения систем, но идеальная биометрическая идентификация пока недостижима. В результате суммарная ошибка первого и второго рода равная 0,1 %, для базы, состоящей из миллиона образцов, имплицирует 1000 неверных идентификаций. Существуют практические методы нивелирования этой особенности широкомасштабных систем биометрической идентификации личности, сводящиеся к привлечению дополнительной информации и позволяющие сузить целевое множество альтернативных вариантов. При этом сравнительно небольшую по объему часть работы по биометрической идентификации может взять на себя эксперт-биометрист. Но на практике идентификация личности в мультиклассовой постановке при значительном числе классов сводится к дополнительной логической обработке данных, относящихся к некоторому множеству классов, различить которые автоматически биометрическими методами не представляется возможным. Заметим, что до сих пор отсутствует *научно обоснованная методика*, согласно которой упомянутое множество формируется в принципе. Аналогичная ситуация возникает, например, в задаче компенсации канальной шумовой динамики в оптоволоконных системах мониторинга Distributed Acoustic Sensor Monitoring Systems (DAS) [1–5], которая будет подробно рассмотрена в разделе «Пример использования CLC в системах обработки данных оптоволоконной системы мониторинга протяженных объектов». Настоящая работа посвящена поиску решений на перечисленные проблемы.

В работе предложен принципиально новый тип классификаторов, предназначенных для отображения пространства первичных признаков объектов классификации в конечную систему подмножеств множества индексов возможных классов. В этом случае результатом классификации является не единичный индекс класса, а некоторое непустое множество индексов с ограниченной мощностью, которое с заданным начальным коэффициентом доверия содержат искомый класс. Назовем этот тип классификаторов — доверительными. Впервые подобный подход к исследованию свойств липшициевых классификаторов был описан в работе [6]. Липшицевы классификаторы [7] образуют достаточно общий класс алгоритмов, который, в том числе, включает в себя большинство известных топологий нейронных сетей [8, 9], классические Support Vector Machines (SVM) [10], Relevance Vector Machines и Linear Programming Machines. Как показано в [6], свойства липшицевых классификаторов позволяют гарантированно оценивать точность их работы, что является решающим фактором для их использования в качестве фундамента при построении класса доверительных классификаторов.

Цель настоящей работы — выполнить описание нового типа классификаторов и понятия об ассамблее доверительных липшициевых классификаторов (Confidence Lipschitz Classifier, CLC), а также обсудить практический пример их использования.

Краткий экскурс в теорию липшицевых классификаторов

Липшицевы классификаторы представляют собой емкий класс алгоритмов, широко используемых в целом ряде практических приложений, в том числе в биометрии, системах обработки шумовых сигналов, автоматическом анализе текстов, в различных эконометрических приложениях. Рассмотрим основы концепции построения липшицевых классификаторов детально, а также основы теоретического фундамента, на который опирается эта перспективная концепция. В работе [10]

описаны основы концепции разделения классов посредством гиперплоскости с широким зазором. В [10] было впервые показано, что использование принципа максимизации зазора (маржи) между элементами различных классов обучающего множества и гиперплоскостью, их разделяющую, позволяет минимизировать верхнюю границу величины эмпирического риска и, тем самым, увеличить обобщающую способность классификатора с зазором «Margin Classifier» или «Large Margin Classifier» [11]). В настоящей работе будем использовать термин «маржинальный классификатор».

Заметим, что конструирование маржинальных классификаторов состоит в достаточно близкой ассоциации с проблемой изоморфного изометрического погружения метрических пространств в банахово или гильбертово пространства [12]. В этом случае компактное пространство первичных признаков (Z, d), где $d()$ — метрика этого пространства; Z — множество значений первичных признаков, изометрически погружено в целевое банахово пространство $(B, \|\cdot\|)$ посредством отображения (feature mapping) φ: $Z \to B$. Маржинальный классификатор строится в подпространстве пространства $(B, \|\cdot\|)$, в котором выпуклые множества разделяются гиперплоскостью согласно теореме разделения Хана–Банаха. В частности, SVM-алгоритм классификации соответствует случаю, когда метрическое пространство погружается в гильбертово пространство. Итак, маржинальный классификатор представляет собой такое классификационное правило, при котором элементы обучающего множества разносятся «далеко» от разделяющей классы гиперплоскости H в пространстве B (в результате обучения достигается максимальное значение величины зазора (маржи) между классами). Пусть $\{(z_i, y_i)|i = 1, \ldots, n\} \subset Z \times \{\pm 1\}$ — обучающее множество.

Представим выпуклую оболочку множества F в виде:

$$C(F) = \left\{ \sum_{i\in I} \alpha_i z_i | \sum_{i\in I} \alpha_i = 1, \ \forall_{i\in I} \alpha_i > 0, z_i \in F, |I| < \infty \right\}.$$

Пусть F^+, F^- — обучающие множества для двух классов «+» и «–», при условии $C(F^+) \cap C(F^-) = \emptyset$; α_i — коэффициенты выпуклой оболочки; z_i — элементы множества F; I — множество индексов.

В работе [13] показано, что конструирование маржинального классификатора эквивалентно определению расстояния (отступа или маржи) $d(C(F^+), C(F^-))$ между множествами $C(F^+)$ и $C(F^-)$. В этом случае $d(C(F^+), C(F^-)) = \underset{p^+\in C(F^+), p^-\in C(F^-)}{Inf} \|p^+ - p^-\|$. В [13] доказано, что $d(C(F^+), C(F^-)) = \sup_{T\in B'} \underset{p^+\in C(F^+), p^-\in C(F^-)}{Inf} \langle T, p^+ - p^-\rangle \|T\|^{-1}$, а конструирование маржинального классификатора, в итоге эквивалентно решению следующей оптимизационной задачи:

$$\underset{T\in B', b}{Inf} \|T\| \text{ при условии } \forall_{i=1}^{n} (y_i(\langle T, z_i\rangle + b) \geq 1), \qquad (1)$$

где T — элементы пространства B'; B' — сопряженное к B пространство; b — элемент одномерного вещественного пространства. Маржинальный классификатор, служащий решением задачи (1), представим в форме функции следующего вида: $f(z) = \langle T, z\rangle + b$.

В этом случае величина маржи равна $\|T\|^{-1}$. Обозначим: $AE(Z)$ — пространство Аренса–Эльса [14]; $AE(Z)'$ — пространство, сопряженное к $AE(Z)$ (пространство всех непрерывных линейных форм на $AE(Z)$); $Lip(Z)$ — множество липшицевых функций:

$$Lip(Z) := \{f | \forall f \exists\, L \in]0, \infty[: \forall x, y \in Z : |f(x) - f(y)| \leq L d(x, y)\}. \qquad (2)$$

Определим как $L(f)$ минимальную константу L, для которой выполняется неравенство $|f(x) - f(y)| \leq L d(x, y)$ при всех $x, y \in Z$. В [7] предложено изометрически погружать Z в соответствующее банахово пространство $AE(Z)$:

$$\Phi: Z \to AE(Z), \ \Psi: Lip(Z) \to AE(Z)',$$

где Φ — отображение, которое представляет собой изометрическое погружение пространства Z в пространство $AE(Z)$; Ψ — отображение между пространствами $Lip(Z)$ и $AE(Z)'$ при условии, что $Lip(Z)$ из (2) является изометрически изоморфным пространству $AE(Z)'$. Конструировать маржинальный классификатор предложено посредством решения следующей оптимизационной задачи:

$$\underset{f\in Lip(Z)}{Inf} L(f) \text{ при условии } \forall_{i=1}^{n} (y_i f(x_i) \geq 1). \qquad (3)$$

Решением задачи (3) является, так называемый, жесткий маржинальный липшициев классификатор (Hard Margin Lipschitz Classifier) [7]. В (3) получена величина маржи $1/L(f)$. С другой стороны, мягкий маржинальный липшициев классификатор (Soft Margin Lipschitz Classifier) [7] представляет наибольший практический интерес и является решением следующей оптимизационной задачи:

$$\underset{f\in Lip(Z)}{Inf} \lambda L(f) + n^{-1} \sum_i l(y_i f(z_i)),$$

$$\text{где } l(y_i f(z_i)) = \max\{0, 1 - y_i f(z_i)\},$$

где λ — регуляризирующая константа; l — функция невязки.

Фактически, широко известный SVM-классификатор является мягким маржинальным липшицевым классификатором. Кроме рассмотренного изометрического погружения (Z, d) в $AE(Z)$, существуют другие изометрические погружения, также пригодные к использованию в качестве платформы для построения маржинального классификатора. Например, в работе [15] изучено погружение метрического пространства Z в пространство непрерывных функций, известное как погружение Куратовского. Каким-либо образом построенный липшицев классификатор по определению обладает решающей (дискриминирующей) функцией с малой константой Липшица, что полностью согласуется с принципом регуляризации, который предполагает избегать использования в классификаторах решающих функций с высокой вариацией. Интуитивно хорошо

воспринимается и факт того, что величина маржи, достигнутая при разделении классов липшицевым классификатором, обратно пропорциональна константе Липшица решающей функции этого классификатора.

Точечная классификация посредством липшицевых классификаторов

Пусть, аналогично с разделом «Краткий экскурс в теорию липшицевых классификаторов», (Z, d) представляет собой компактное метрическое пространство первичных признаков. Обозначим: кардинальное число некоторого множества X символом $|X|$; $Diam(Z)$ — диаметр множества Z: $Diam(Z) = \sup\limits_{z_1, z_2 \in Z} \|z_1 - z_2\|$, где $\|\cdot\|$ — норма пространства признаков (Z, d). Каждый объект, подлежащий классификации, имеет индекс θ и описан наблюдаемым набором первичных признаков $z(\theta) \in Z$.

Множество возможных классов конечно, а их индексы совместно образуют множество $\Theta = \{\theta_k\} \subseteq R^1$, $|\Theta| < \infty$. При этом $\theta_2 \neq \theta_1 \in \Theta$: $d(z(\theta_2), z(\theta_1)) > 0$. Примем, что $d(z(\theta_2), z(\theta_1))$ эквивалентно $d(\theta_2, \theta_1)$. Обучающая выборка представляет собой совокупность конечных множеств $\mathfrak{M}(\Theta) = \bigcup\limits_{\theta \in \Theta} \Omega^{(\theta)}$, где

$$\forall(\theta \in \Theta, \Omega^{(\theta)} = \mathfrak{M}(\Theta)): \Omega^{(\theta)} =$$
$$= \left\{(z_j(\theta), \theta) | z_j(\theta) \in Z; j \in \{1, \ldots |z_j(\theta)|\}; |z_j(\theta)| < \infty\right\}.$$

На основании изучения свойств наблюденного образца $\mathbf{z}^{(\theta^*)} \in Z$, автоматический классификатор должен выбрать из множества Θ индекс класса, которому, с точки зрения классификатора, принадлежит представленный для классификации образец $\mathbf{z}^{(\theta^*)}$. Здесь $\theta^* \in \Theta$ — ненаблюдаемый индекс целевого класса, от значения которого наблюдение $\mathbf{z}^{(\theta^*)}$ зависит имплицитно.

Отметим, что в настоящей работе рассмотрены только классификаторы, работающие со стационарными (независящими от времени) объектами классификации. Любой классификатор оценивает степень соответствия опытного образца $\mathbf{z}^{(\theta^*)}$ произвольному классу с индексом $\theta \in \Theta$ посредством исследования величины специальной случайной функции $f(\theta|\mathbf{z}^{(\theta^*)})$, $f(\theta|\mathbf{z}^{(\theta^*)}) \in R^1$. Эта функция называется решающей или *дискриминирующей функцией*. По значению функции $f(\theta|\mathbf{z}^{(\theta^*)})$ можно сделать вывод о том, в какой степени образец $\mathbf{z}^{(\theta^*)}$ соответствует гипотетическому классу с индексом $\theta \in \Theta$. Обозначим индекс класса, которому реально принадлежит полученный образец $\mathbf{z}^{(\theta^*)}$, $\chi(\mathbf{z}^{(\theta^*)})$. Очевидно, что $\theta^* = \chi(\mathbf{z}^{(\theta^*)})$.

Классификатор f: $Z \to \Theta$ представляет собой функцию, которая разделяет пространство (Z, d) на $m = |\Theta|$ классов. В свою очередь, $f(\theta|\mathbf{z}^{(\theta^*)})$ — стохастическая функция, эксплицитно зависящую от индекса проверяемой гипотезы $\theta \in \Theta$ и имплицитно от индекса целевого класса $\theta^* \in \Theta$.

Определение 1. *Классификатор $f(\theta|\mathbf{z}^{(\theta^*)})$ называется точечным липшицевым классификатором (PLC), если одновременно выполняются следующие условия:*

— *классификатор является мягким маржинальным липшицевым классификатором*;
— $|\Theta| \gg 2$, $|\Theta| < \infty$;
— *классификационное решение представляет собой множество* $\tilde{\theta} = \left\{\theta | \operatorname{Arg\,Max}\limits_{\theta \in \Theta}(f(\theta|\mathbf{z}^{(\theta^*)}))\right\}$, $|\tilde{\theta}| = 1$.

Очевидно, что дискриминирующая функция $f(\theta|\cdot)$ PLC липшицируема на Θ, а выходное множество индексов классов $\tilde{\theta}$ состоит из единственного элемента. Именно поэтому PLC и называются точечными.

Понятие доверительного липшицевого классификатора

Приведем определение и рассмотрим некоторые свойства нового типа классификаторов — CLC. CLC, в отличие от точечных классификаторов, обладают гарантированными точностными показателями и удобны в задачах с большим числом классов $|\Theta| \gg 1$. Особенно CLC полезны тогда, когда часть классов из множества Θ принципиально плохоразличима в пространстве первичных признаков (Z, d). Часто, плохоразличимость этих групп классов представляет собой внутренне присущее свойство изучаемой объектовой совокупности. И поэтому никакие дополнительные манипуляции с пространствами первичных признаков и/или перебор различных типов точечных классификаторов неспособны существенно повысить точность решения классификационной задачи. Для решения классификационных задач в этом случае разумно использовать классификатор CLC.

Определение 2. *Кортеж* $\Psi(\theta|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c)) = (f(\theta|\mathbf{z}^{(\theta^*)}), c(P_c))$ *называется CLC, если одновременно выполняются следующие условия:*

— *классификатор* $(f(\theta|\mathbf{z}^{(\theta^*)}), \theta \in \Theta, \mathbf{z}^{(\theta^*)} \in Z$ *является мягким маржинальным липшицевым классификатором, отображающим пространство первичных признаков* (Z, d) *на множество индексов классов* Θ;
— $|\Theta| \gg 2$, $|\Theta| < \infty$;
— *скалярный параметр P_c, называемый коэффициентом доверия, таков, что* $0 < P_c < 1$;
— *скалярный параметр $c(P_c)$, называемый порогом доверия и зависящий от величины коэффициента доверия, таков, что* $0 < c(P_c) < \infty$;
— *результатом использования* $\Psi(\theta|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$ *для классификации образца* $\mathbf{z}^{(\theta^*)}$ *является непустое замкнутое множество* $\Xi(\tilde{\theta}|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c)) \subseteq \Theta$ *такое, что*
  — $\Xi(\tilde{\theta}\mathbf{z}^{(\theta^*)}|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c)) = \{\theta \in \Theta \| f(\tilde{\theta}(\mathbf{z}^{(\theta^*)})|\mathbf{z}^{(\theta^*)}) - f(\theta|\mathbf{z}^{(\theta^*)})| \le c(P_c)\}$, *где* $\tilde{\theta}(\mathbf{z}^{(\theta^*)}) = \operatorname{ArgMax}\limits_{\theta \in \Theta}(f(\theta|\mathbf{z}^{(\theta^*)}))$;
  — $\mathbf{P}(\theta^* \in \Xi(\tilde{\theta}|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c)) \ge P_c$.

В отличие от точечного, CLC гарантированно проецирует пространство первичных признаков на непустое множество индексов объектов, которое с заданным начальным коэффициентом доверия содержит индекс целевого класса. Точечный классификатор, наоборот, проецирует пространство первичных признаков в единственный индекс класса, причем точностные характеристики этого преобразования никак не гарантированы. CLC $\Psi(\theta|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$ называется построенным на базе PLC $f(\cdot)$.

Рассмотрим PLC стационарных объектов $f(\theta|\mathbf{z}^{(\theta^*)})$. Решающая функция PLC допускает следующее представление в виде:

$$f(\theta|\mathbf{z}^{(\theta^*)}) = \mathbf{E}_{\theta^*} f(\theta|\mathbf{z}^{(\theta^*)}) + \eta(\theta|\mathbf{z}^{(\theta^*)}),$$

где $\mathbf{E}_{\theta^*} f(\theta|\mathbf{z}^{(\theta^*)})$ — условное математическое ожидание случайной функции $f(\theta|\mathbf{z}^{(\theta^*)})$ с фиксированными параметрами θ^*, $\theta \in \Theta$; $\eta(\theta|\mathbf{z}^{(\theta^*)})$ — случайная функция, зависящая от параметров θ^*, $\theta \in \Theta$.

В работе [1] исследованы свойства CLC. В частности, доказана следующая теорема.

Теорема 1 [5]. Пусть выполнены следующие условия:

1) $\forall\theta\theta^* \in \Theta$: $\mathbf{E}_{\theta^*}(\eta(\theta|\mathbf{z}^{(\theta^*)})) = 0$;
2) $\forall\theta^* \in \Theta$: $\Big\{\operatorname{Max}_{\theta\in\Theta}(\mathbf{E}_{\theta^*} f(\theta|\mathbf{z}^{(\theta^*)})) = \mathbf{E}_{\theta^*} f(\theta^*|\mathbf{z}^{(\theta^*)}),$
$\big|\big\{\theta^{\#} \in \Theta|\mathbf{E}_{\theta^*} f(\theta^{\#}|\mathbf{z}^{(\theta^*)}) = \mathbf{E}_{\theta^*} f(\theta^*|\mathbf{z}^{(\theta^*)})\big\}\big| = 1\Big\}$;
3) $\forall\theta_1,\theta_2,\theta^* \in \Theta\ \exists\ L \in]0, \infty[$: $\|f(\theta_1|\mathbf{z}^{(\theta^*)}) - f(\theta_2|\mathbf{z}^{(\theta^*)})\| \le$
$\le Ld(\theta_1, \theta_2)$ п.н.;
4) $c(P_c) = LDiam(Z)(1 - P_c)^{-0,5}$ и $P_c \in]0, 1[$.

Тогда имеет место утверждение:

$$\forall\theta^* \in \Theta: \mathbf{P}_{\theta^*}(\theta^* \in \Xi(\tilde{\theta}(\mathbf{z}^{(\theta^*)})|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))) \ge P_c.$$

Условие 2 Теоремы 1 требует, чтобы математическое ожидание от дискриминирующей функции достигало максимума на индексе целевого объекта. Это необходимое условие для успешной классификации с помощью выбранной дискриминирующей функции $f(\cdot)$. На стадии обучения классификатора его параметры выберем таким образом, чтобы условие 2 Теоремы 1 достигалось хотя бы на тренировочном наборе данных. Если достичь этого условия не удается, тогда определим положительную константу $\vartheta^*(\Theta) \in R^1$, определяющую минимальную верхнюю грань для величины $\Big(\operatorname{Max}_{\theta\in\Theta}(\mathbf{E}_{\theta^*} f(\theta|\mathbf{z}^{(\theta^*)})) - \mathbf{E}_{\theta^*} f(\theta^*|\mathbf{z}^{(\theta^*)})\Big)$, $\theta^* \in \Theta$.

Тогда:

$$\vartheta^*(\Theta) = Inf\Big\{\vartheta \in R^1|\forall\theta^* \in \Theta:$$
$$\Big(\Big(\operatorname{Max}_{\theta\in\Theta}(\mathbf{E}_{\theta^*} f(\theta|\mathbf{z}^{(\theta^*)})) - \mathbf{E}_{\theta^*} f(\theta^*|\mathbf{z}^{(\theta^*)})\Big) \le \vartheta\Big)\Big\}.$$

Величину $\vartheta^*(\Theta)$ оценим численными методами на $\mathfrak{M}(\Theta)$. Нетрудно показать, что утверждения Теоремы 1 выполнены при условии $c(P_c) = \vartheta^*(\Theta) + LDiam(Z)(1 - P_c)^{-0,5}$.

Таким образом, результатом классификации образца $\mathbf{z}^{(\theta^*)}$ посредством CLC является ограниченное множество $\Xi(\tilde{\theta}|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$, которое с заданным коэффициентом доверия P_c содержит индекс целевого класса θ^*.

Определение 3. Процедуру применения CLC $\Psi(\theta|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$, построенного на базе PLC $f(\cdot)$ для классификации образца $\mathbf{z}^{(\theta^*)}$, назовем Ψ-проектором, а множество $\Xi(\tilde{\theta}(\mathbf{z}^{(\theta^*)})|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$, являющееся результатом применения Ψ-проектора к образцу $\mathbf{z}^{(\theta^*)}$ — множеством P_c-неразличимости целевого класса θ^*.

Практически, класс P_c-неразличимости $\Xi(\tilde{\theta}(\mathbf{z}^{(\theta^*)})|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$ представляет собой множество индексов классов, которые фактически невозможно различить в рамках используемого алгоритма классификации. Важно то, что сама по себе неразличимость является не только, и не сколько, следствием используемого метода точечной классификации, но в значительной степени имплицируется различающими характеристиками пространства первичных признаков. Иначе говоря, чем хуже различимы классы в пространстве первичных признаков, тем больше мощность класса P_c-неразличимости. Напомним, что плохая различимость классов в пространстве первичных признаков часто является следствием естественно высокой идентичности классов, подлежащих классификации. Например, достаточно трудно различить однояйцовых, однополых близнецов по оцифрованному фотографическому изображению их лиц, какие бы первичные признаки, автоматически выделяемые из оцифрованных изображений, не использовались бы для этой цели.

Качество классификации, реализованной в Ψ-проекторе, оценим путем использования усредненной по всей совокупности классов величины шенноновской энтропии, естественно сопровождающей принятие решения о значении истинного индекса целевого класса. В этом случае необходимо сделать выбор из множества альтернатив, составляющих класс P_c-неразличимости. Другими словами, чем больше мощность класса P_c-неразличимости, тем выше энтропия этого класса. Полагая равновероятными гипотезы из любого класса P_c-неразличимости $\Xi(\tilde{\theta}(\mathbf{z}^{(\theta^*)})|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$, $\theta^* \in \Theta$, получаем простой вид для величины реализовавшейся в Ψ-проекторе энтропии:

$$h(\theta^*) = \mathbf{E}\big(\ln(|\Xi(\tilde{\theta}(\mathbf{z}^{(\theta^*)})|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))|)\big). \qquad (4)$$

В выражении (4) усреднение выполнено по всем возможным реализациям $\mathbf{z}^{(\theta^*)}$ при фиксированном значении целевого параметра $\theta^* \in \Theta$. В силу мультиклассовости постановки задачи, желательно иметь ориентиры для достигнутого качества классификации по всему множеству возможных значений целевого параметра. Обозначим символом $\mathbf{Z}(\Theta) = \{\mathbf{z}^{(\theta^*)}|\theta^* \in \Theta\}$ — множество всех возможных реализаций элементов множества первичных признаков из пространства (Z, d). Если Ψ-проектор применен ко всему множеству реализаций из $\mathbf{Z}(\Theta)$, то в результате будет получена совокупность $\boldsymbol{\Xi}(\Theta) = \big\{\Xi(\tilde{\theta}(\mathbf{z}^{(\theta^*)})|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))|\theta^* \in \Theta, \mathbf{z}^{(\theta^*)} \in \mathbf{Z}(\Theta)\big\}$ множеств P_c-неразличимости. Реализовавшаяся совокупность $\boldsymbol{\Xi}(\Theta)$, совместно характеризует качество использованного Ψ-проектора. Однако, с практической точки зрения, существенно важнее знание верхней границы для достигнутой величины усредненной шенноновской энтропии, соответствующей некому обобщенному классу из Θ. Легко видеть, что эта граница имеет следующий простой вид: $\mathbf{H}(\Theta) = |\Theta|^{-1} \sum_{\theta^*\in\Theta} h(\theta^*)$, причем, чем меньше величина $\mathbf{H}(\Theta)$, тем более высокое качество классификации достигнуто Ψ-проектором по всему множеству классов Θ. Получим, что величина $\mathbf{H}(\Theta)$ представляет собой верхнюю границу для величины усредненной шенноновской энтропии классификационного решения, полученного в проекторе $\Psi(\theta|f(\cdot), \mathbf{z}^{(\theta^*)}, c(P_c))$. В [6] приведены результаты практического использования CLC для решения задачи идентификации диктора по цифровой записи его голоса. Для уменьшения громоздкости записи Ψ-проектор обозначим как $\Psi(\theta|f, P_c)$.

CLC по ассамблее PLC

Ассамблеи классификаторов широко используются для стабилизации (баггин-схема, [16]) и повышения эффективности (бустинг-схема, [17]) классификационных процедур. При этом формирование интегрального классификационного решения строится по различным схемам комбинирования результатов классификации, полученным отдельными классификаторами, входящими в ассамблею. Рассмотрим построение CLC по ассамблее PLC.

Пусть $\mathbf{Z} = \{(Z_k, d_k)_k | k \in \{1, \ldots, m\}\}$ представляет собой множество компактных метрических пространств первичных признаков, где d_k — метрика k-го пространства из $\mathbf{Z}$; Z_k — множество значений первичных признаков k-го пространства Z_k.

Обозначим символом $\mathbf{F}(\mathbf{Z}|\theta^*) = \{f_k(\theta|\mathbf{z}_k^{(\theta^*)}) | k = 1, \ldots, m\}$ ассамблею статистически независимых PLC мощностью m, сформированную при значении искомого класса $\theta^* \in \Theta$. Каждая дискриминирующая функция $f_k(\theta|\mathbf{z}_k^{(\theta^*)})$ соответствует определенному пространству $(Z_i, d_i)_i \in \mathbf{Z}$, и допустимо следующее представление:

$$f_k(\theta|\mathbf{z}_k^{(\theta^*)}) = \mathbf{E}_{\theta^*} f_k(\theta|\mathbf{z}_k^{(\theta^*)}) + \eta_k(\theta|\mathbf{z}_k^{(\theta^*)}),$$

где $k \in \{1, \ldots, m\}$; $\mathbf{z}_k^{(\theta^*)} \in Z_k$; $\mathbf{E}_{\theta^*} f_k(\theta|\mathbf{z}_k^{(\theta^*)})$ — условное математическое ожидание случайной функции $f_k(\theta|\mathbf{z}_k^{(\theta^*)})$ с фиксированными параметрами θ^*, $\theta \in \Theta$; $\eta_k(\theta|\mathbf{z}_k^{(\theta^*)})$ — случайная функция, зависящая от параметров θ^*, $\theta \in \Theta$.

Символом $f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})$ обозначим результат нормализации дискриминирующей функции $f_k(\theta|\mathbf{z}_k^{(\theta^*)})$ при условии $f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}) \in [0, 1]$.

CLC по ассамблее $\mathbf{F}(\mathbf{Z}|\theta^*)$ представляет собой пару

$$\Psi(\theta|F(\cdot), \mathbf{F}(\mathbf{Z}|\theta^*), c(P_c)) = (F(\theta|\mathbf{F}(\mathbf{Z}|\theta^*)), c(P_c)),$$

где $F(\theta|\mathbf{F}(\mathbf{Z}|\theta^*))$ — интегральная дискриминирующая функция, построенная для ассамблеи $\mathbf{F}(\mathbf{Z}|\theta^*)$.

Найдем функцию $F(\theta|\mathbf{F}(\mathbf{Z}|\theta^*))$, соответствующую ассамблее PLC $\mathbf{F}(\mathbf{Z}|\theta^*)$, в виде выпуклой оболочки $F(\theta|\mathbf{F}(\mathbf{Z}|\theta^*)) = \sum\limits_k \alpha_k f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})$, где $\sum\limits_k \alpha_k = 1$, $\forall \alpha_k \geq 0$.

Набор коэффициентов $\{\alpha_k | k \in \{1, \ldots, m\}\}$ сформируем на стадии оптимизации интегральной дискриминирующей функции в соответствие с выбранным критерием оптимизации. В качестве критерия оптимизации выберем математическое ожидание от функции потерь интегрального классификатора, аргументом которой является сумма классификационных ошибок первого и второго рода. Примем, что каким-либо образом набор коэффициентов $\{\alpha_k\}$ — фиксирован.

Рассмотрим множество:

$$\Xi(\tilde{\theta}(\mathbf{Z}|\theta^*|F(\cdot), \mathbf{Z}, c(P_c)) = \\ = \left\{\theta \in \Theta \middle| |F(\tilde{\theta}(\mathbf{Z}|\theta^*)|\mathbf{F}(\mathbf{Z}|\theta^*)) - F(\theta|\mathbf{F}(\mathbf{Z}|\theta^*))| \leq c(P_c)\right\},$$

где $\tilde{\theta}(\mathbf{Z}|\theta^*) = \underset{\theta \in \Theta}{\mathrm{ArgMax}}(\mathbf{F}(\theta|\mathbf{Z}|\theta^*))$.

Теорема 2. *Пусть одновременно выполнены следующие условия*:

1) *Случайные величины* $\left\{\eta_k(\theta|\mathbf{z}_k^{(\theta^*)}) | k \in \{1, \ldots, m\}\right\}$ *взаимно независимы*;
2) $\forall k$; θ, $\theta^* \in \Theta$: $\mathbf{E}_{\theta^*}(\eta_k(\theta|\mathbf{z}_k^{(\theta^*)})) = 0$;
3) $\forall k \; \exists \, a_k, b_k \in R^1$: $(f_k(\theta|\mathbf{z}_k^{(\theta^*)}) \in [a_k, b_k] \subseteq R^1, |a_k - b_k| < \infty)$;
4) $f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}) = (f_k(\theta|\mathbf{z}_k^{(\theta^*)}) - b_k)|a_k - b_k|^{-1}$;
5) $\forall k$; $\theta^* \in \Theta$: $\left\{\underset{\theta \in \Theta}{\mathrm{Max}}(\mathbf{E}_{\theta^*} f_k(\theta|\mathbf{z}_k^{(\theta^*)})) = \mathbf{E}_{\theta^*}(f_k(\theta^*|\mathbf{z}_k^{(\theta^*)}),\right.$ $\left.|\{\theta^{\#} \in \Theta | \mathbf{E}_{\theta^*} f_k(\theta|\mathbf{z}_k^{(\theta^*)}) = \mathbf{E}_{\theta^*} f_k(\theta^*|\mathbf{z}_k^{(\theta^*)})\}| = 1\right\}$;
6) $\forall k$; $\theta_1, \theta_2, \theta^* \in \Theta \; \exists \, L_k \in]0, \infty[$: $\|f_k(\theta_1|\mathbf{z}_k^{(\theta^*)}) - f_k(\theta_2|\mathbf{z}_k^{(\theta^*)})\| \leq$ $\leq L_k d_k(\theta_1, \theta_2)$ п. н.;
7) $c(P_c) = \left(\sum\limits_k \alpha_k^2 (L_k^{(N)} Diam(Z_k))^2\right)^{0,5} (1 - P_c)^{-0,5}$, где $P_c \in]0, 1[$ и $L_k^{(N)} = L_k |a_k - b_k|^{-1}$.

Тогда истинно следующее утверждение:

$$\forall \theta^* \in \Theta: \mathbf{P}_{\theta^*}(\theta^* \in \Xi(\tilde{\theta}(\mathbf{Z}|\theta^*)|F(\cdot), \mathbf{Z}, c(P_c))) \geq P_c.$$

Доказательство. В тексте доказательства использованы обозначения — запись «X ⇒ Y» обозначает то, что «X имплицирует Y»; «ω(1) ⊂ ω(2)» — «событие ω(1) влечет событие ω(2)»:

— $\tilde{\theta} \equiv \tilde{\theta}(\mathbf{Z}|\theta^*)$;
— $\eta_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}) = \eta_k(\theta|\mathbf{z}_k^{(\theta^*)})|a_i - b_i|^{-1}$;
— $\Delta\eta^{(i,N)}(\tilde{\theta}, \theta) = (\eta_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - \eta_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}))$;
— $\delta = c(P_c) - \sum\limits_k \alpha_k \Delta\eta^{(i,N)}(\tilde{\theta}, \theta)$;
— $\Phi(\tilde{\theta}, \theta|\theta^*) = \sum\limits_k \alpha_k (\mathbf{E}_{\theta^*}(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)})) - \mathbf{E}_{\theta^*}(f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})))$.

Очевидно следующее представление:

$$\sum_k \alpha_k f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - \sum_k \alpha_k f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}) = \\ = \sum_k \alpha_k (\mathbf{E}_{\theta^*}(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)})) - \mathbf{E}_{\theta^*}(f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}))) + \\ + \sum_k \alpha_k ((\eta_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - \eta_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}))) = \\ = \Phi(\tilde{\theta}, \theta|\theta^*) + \sum_i \alpha_i \Delta\eta^{(k,N)}(\tilde{\theta}, \theta).$$

Вследствие условия 5 Теоремы 2 запишем:

$$\tilde{\theta} \in \Theta: \Phi(\tilde{\theta}, \theta|\theta^*) = \\ = \sum_k \alpha_k (\mathbf{E}_{\theta^*}(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)})) - \mathbf{E}_{\theta^*}(f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}))) \leq 0.$$

Величина $|\Phi(\tilde{\theta}, \theta|\theta^*)|$ не может быть равна нулю, когда $\sum\limits_i \alpha_i \Delta\eta^{(k,N)}(\tilde{\theta}, \theta) \neq 0$. В случае, когда $|\Phi(\tilde{\theta}, \theta|\theta^*)| = 0$, основываясь на условии 5 Теоремы 2, имеем: $|\Phi(\tilde{\theta}, \theta|\theta^*)| = 0$ и $\tilde{\theta} = \theta^*$, так как $\tilde{\theta} = \underset{\theta \in \Theta}{\mathrm{ArgMax}}(F(\theta|(\mathbf{Z}, \theta^*))$. Таким образом, получим следующую импликацию:

$$(\Phi(\tilde{\theta}, \theta|\theta^*) = 0) \Rightarrow (\theta = \theta^*). \qquad (5)$$

Допустимо следующее представление:

$$\Xi(\tilde{\theta}|F(\cdot), \mathbf{Z}, c(P_c)) = \\ = \left\{\theta \in \Theta \middle| |\Phi(\tilde{\theta}, \theta|\theta^*) + \sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)| \leq c(P_c)\right\}.$$

Рассмотрим события:

$$\omega_0: \left\{\left|\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)\right| \le c(P_c)\right\},$$
$$\omega_1: \{0 < |\Phi(\tilde{\theta}, \theta|\theta^*)| \le c(P_c)\},$$
$$\omega_2: \{\Phi(\tilde{\theta}, \theta|\theta^*) = 0\}, \omega_3: \{\theta^* \in \Xi(\tilde{\theta}|\mathbf{Z}, c(P_c))\}.$$

Учитывая (5), на множестве $\Xi(\tilde{\theta}|\mathbf{Z}, c(P_c))$ введем условие:

$$\omega_0 \subset \omega_1 \supset \omega_2 \subset \omega_3. \quad (6)$$

Далее, с учетом взаимной независимости величин $\left\{\eta_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})|k \in \{1, \ldots, m\}\right\}$, имеем:

$$\mathbf{E}_{\theta^*}\left(\sum_k \alpha_k f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - \sum_k \alpha_k f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})\right)^2 =$$
$$= \mathbf{E}_{\theta^*}(\Phi(\tilde{\theta}, \theta|\theta^*))^2 + \mathbf{E}_{\theta^*}\left(\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)\right)^2.$$

Для любого $k \in \{1, \ldots, m\}$ допустимо аналогичное представление:

$$\mathbf{E}_{\theta^*}\left(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})\right)^2 = \left(\mathbf{E}_{\theta^*}\left(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)})\right) -\right.$$
$$\left. - \mathbf{E}_{\theta^*}(f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}))\right)^2 + \mathbf{E}_{\theta^*}(\Delta\eta^{(k,N)}(\tilde{\theta}, \theta))^2.$$

Заметим, что:

$$\forall k;\ \theta_1, \theta_2, \theta^* \in \Theta: |f_k^{(N)}(\theta_1|\mathbf{z}_k^{(\theta^*)}) - f_k^{(N)}(\theta_2|\mathbf{z}_k^{(\theta^*)})| =$$
$$= \left|\frac{f_k(\theta_1|\mathbf{z}_k^{(\theta^*)}) - f_k(\theta_2|\mathbf{z}_k^{(\theta^*)})}{|a_k - b_k|}\right| \le \frac{L_k}{|a_k - b_k|} d_k(\theta_1, \theta_2) = \quad (7)$$
$$= L_k^{(N)} d_k(\theta_1, \theta_2).$$

С учетом, что $(\mathbf{E}_{\theta^*}\left(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)})\right) - \mathbf{E}_{\theta^*}\left(f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})\right))^2 \ge 0$ и (7), найдем:

$$\forall k: \mathbf{E}_{\theta^*}(\Delta\eta^{(k,N)}(\tilde{\theta}, \theta))^2 \le \mathbf{E}_{\theta^*}\left(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})\right)^2 \le$$
$$\le \int_{-\infty}^{\infty} \left(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})\right)^2 \rho(x|\tilde{\theta}, \theta, \theta^*) dx \le$$
$$\le \int_{-\infty}^{\infty} \left(L_k^{(N)}\right)^2 d^2(\tilde{\theta}, \theta) \rho(x|\tilde{\theta}, \theta, \theta^*) dx \le$$
$$\le \left(L_k^{(N)}\right)^2 d^2(\tilde{\theta}, \theta) \int_{-\infty}^{\infty} \rho(x|\tilde{\theta}, \theta, \theta^*) dx \le \left(L_k^{(N)}\right)^2 (Diam(Z_k))^2,$$

где $\rho(x|\tilde{\theta}, \theta, \theta^*)$ — плотность распределения случайной величины $(f_k^{(N)}(\tilde{\theta}|\mathbf{z}_k^{(\theta^*)}) - f_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)}))$.

В результате получим:

$$\forall\forall \tilde{\theta}, \theta \in \Theta: \mathbf{E}_{\theta^*}(\Delta\eta^{(k,N)}(\tilde{\theta}, \theta))^2 \le \left(L_k^{(N)}(Diam(Z_k))\right)^2.$$
$$k$$

Отсюда, с учетом взаимной независимости случайных величин $\left\{\eta_k^{(N)}(\theta|\mathbf{z}_k^{(\theta^*)})|k \in \{1, \ldots, m\}\right\}$:

$$\mathbf{E}_{\theta^*}\left(\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)\right)^2 = \sum_k \alpha_k^2 \mathbf{E}_{\theta^*}(\Delta\eta^{(k,N)}(\tilde{\theta}, \theta))^2 \le$$
$$\le \sum_k \alpha_k^2 \left(L_k^{(N)} Diam(Z_k)\right)^2. \quad (8)$$

Учитывая второе условие доказываемой теоремы, найдем:

$$\mathbf{E}_{\theta^*}\left(\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)\right) = 0.$$

На основании неравенства Чебышева и выражения (8), для любого $C > 0$ имеет место следующее неравенство:

$$\forall(\tilde{\theta}, \theta, \theta^* \in \Theta, C > 0): \mathbf{P}_{\theta^*}\left(\left|\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)\right| \le C\right) >$$
$$> 1 - \frac{\mathbf{E}_{\theta^*}(\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta))^2}{C^2} > 1 - \frac{\sum_k \alpha_k^2 (L_k^{(N)} Diam(Z_k))^2}{C^2}.$$

Далее,

$$\forall(\tilde{\theta}, \theta, \theta^* \in \Theta): \mathbf{P}_{\theta^*}\left(\left|\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)\right| \le c(P_c)\right) >$$
$$> 1 - \frac{\sum_k \alpha_k^2 (L_k^{(N)} Diam(Z_k))^2}{c(P_c)^2}.$$

Учитывая условие 7 Теоремы 2 получим:

$$1 - \frac{\sum_k \alpha_k^2 (L_k^{(N)} Diam(Z_k))^2}{c(P_c)^2} = P_c.$$

В этом случае

$$\forall(\tilde{\theta}, \theta, \theta^* \in \Theta): \mathbf{P}_{\theta^*}\left(\left|\sum_k \alpha_k \Delta\eta^{(k,N)}(\tilde{\theta}, \theta)\right|, c(P_c)\right) \ge P_c.$$

С учетом (6)

$$\mathbf{P}_{\theta^*}(\omega_3) = \mathbf{P}_{\theta^*}(\theta^* \in \Xi(\tilde{\theta}|F(\cdot), \mathbf{Z}, c(P_c))) \ge P_c.$$

Теорема доказана ■

Условия Теоремы 2 во многом аналогичны условиям Теоремы 1. В этом случае CLC $\Psi(\theta|F(\cdot), \mathbf{F}(\mathbf{Z}|\theta^*), c(P_c))$ реализована на базе PLC $F(\theta|\mathbf{F}(\mathbf{Z}|\theta^*))$.

Можно показать, что при увеличении величины m (мощности ассамблеи статистически независимых классификаторов $\mathbf{F}(\mathbf{Z}|\theta^*)$), при фиксированной величине параметра P_c, величина порога $c(P_c)$ уменьшается, что, в свою очередь, служит причиной уменьшения математического ожидания мощности соответствующего доверительного множества $\Xi(\cdot)$. Следовательно, уменьшается и величина $\mathbf{H}(\Theta)$, представляющая собой верхнюю границу для величины усредненной шенноновской энтропии классификационного решения, полученного в Ψ-проекторе. Другими словами, при увеличении мощности ансамбля независимых классификаторов будет, в среднем, улучшаться и точность доверительного оценивания истинного значения класса.

Пример использования CLC в системах обработки данных оптоволоконной системы мониторинга протяженных объектов

В качестве примера практического использования **CLC** рассмотрим задачу выбора параметров компонента идентификации целевых событий (КИЦС) системы обработки данных оптоволоконной системы мониторинга протяженных объектов (DAS-система монито-

ринга) [1–5] таким образом, чтобы стабилизировать процесс идентификации целевых событий в условиях нестабильной шумовой обстановки. DAS-системы мониторинга предназначены для контроля инфраструктуры значительной протяженности, в том числе трубопроводов, железнодорожного полотна, периметров режимных объектов, подводных коммуникаций и др. Одной из основных задач DAS-систем мониторинга является обнаружение и классификация сейсмоакустических (виброакустических) событий, которые были реализованы в области чувствительности оптоволоконного сенсора системы. Данные от оптоволоконного сенсора (линии) в систему DAS-мониторинга поступают от пространственно-разнесенных каналов $CH = \{ch_i\}$, ширина которых имеет величину 5–10 м. Как правило, величина $|CH|$ (число каналов) в системах данного класса имеет третий порядок. Причем различные каналы расположены в различных областях контролируемого пространства и могут быть расположены на расстояния в десятки километров друг от друга, находясь в различных шумовых сейсмоакустических (виброакустических) контекстах. Иначе говоря, шумовая обстановка, которая типична для двух любых, пространственно-разнесенных каналов, различна. В некоторых случаях, эта разница — кардинальна, в других — невелика, но она всегда объективно существует. Таким образом, хотя шумовой контекст в области пространственного расположения различных каналов различен, стохастическая динамика этого шума подчиняется определенной системе. Эта система определяется технологическими режимами эксплуатации объекта, временем суток, временем года, наличием постоянных источников сейсмоакустической эмиссии, расположенных в определенных областях расположения оптоволоконного сенсора системы мониторинга. В общем случае система вероятностного распределения шумов вдоль оптоволоконного сенсора вполне может быть описана с точностью до совокупности групп пространственно-смежных каналов (ГПСК) DAS-системы. Важно то, что стохастическая динамика шумовых сигналов в каналах конкретной ГПСК стационарна (стабильна) в течение определенных временных интервалов. Множество ГПСК всего оптоволоконного сенсора, элементы которого соответствуют определенному временному интервалу, определяет конфигурацию пространственного шума вдоль сенсора. Эта конфигурация может считаться стабильной на протяжении определенных интервалов времени в смысле сохранения стационарности вероятностных распределений шумового процесса в каналах ГПСК в течение этих интервалов. Так как общее число каналов системы имеет третий порядок, а длина сенсора достигает пятидесяти километров, то имеет место случай значительного числа ГПСК. В итоге получается значительное число различных конфигураций пространственного шума вдоль сенсора. С целью повышения эффективности функционирования компоненты идентификации целевых сигналов на фоне шумов данная компонента может быть адаптирована к каждой из возможных конфигураций пространственного шума, но это практически нереально в силу их значительного числа. Таким образом логично определить на множестве шумовых конфигураций систему непересекающихся подмножеств, состоящих из конфигураций, близких по определенному критерию, и адаптировать компонент идентификации к каждому из этих подмножеств. Число таких подмножеств (классов), как правило, не превышает нескольких десятков единиц. При этом возникают следующие проблемы: создание научно обоснованного критерия включения конфигурации в шумовой класс, а также идентификация наиболее близкого шумового класса по наблюдениям текущего фонового процесса. Так как КИЦС обрабатывает информацию от каждого канала $ch_i \in CH$, становится актуальной задача определения порога, отделяющего целевой сигнал (целевое сейсмоакустическое или виброакустическое событие) от фонового шума, так как некорректно установленный порог, как на этапе обучения КИЦС, так и на этапе его эксплуатации, является причиной снижения эффективности функционирования КИЦС. Один из простых, но практически эффективных подходов сводится к выполнению следующих шагов.

Шаг 1. В течение определенного интервала времени T, как правило, в течении 40–60 ч, набирается массив Υ статистики по фоновым сигналам, который содержит данные по наблюденному фоновому шуму для каждого $ch_i \in CH$, по нескольким частотным полосам в диапазоне [0, 300] Гц.

Шаг 2. Период T разбивается на малые, непересекающиеся временные интервалы $\mathbf{T} = \{\mathbf{t}_p\}$ равной длины $\forall_p: |\mathbf{t}_p| = \Delta$, в течении которых фоновые сигналы по $ch_i \in CH$ коррелированы. Как правило, на практике величина Δ не превышает 1 мин.

Шаг 3. В каждом канале $ch_i \in CH$ формируется статистическая модель фонового помехового процесса $m_{noise}(t|ch_i)$, $t \in T$, которая может иметь или регрессионный, или авторегрессионный тип [6].

Шаг 4. С использованием моделей, описанных в [6], для $\{m_{noise}(t|ch_i)\}$ формируется множество поканальных наборов порогов $\mathbf{Tr} = \{tr(\mathbf{t}_p)|\mathbf{t}_p \in \mathbf{T}\}$ такое, что для заданной величины $P_T \in [0, 1]$ имеет место утверждение:

$$\forall(ch_i \in CH, \mathbf{t}_p \in \mathbf{T}): \forall t \in \mathbf{t}_p: \mathbf{P}(m_{noise}(t|ch_i)| < tr(\mathbf{t}_p)) \geq P_T.$$

Шаг 5. Методами кластерного анализа набор данных **Tr** разбивается на N кластеров, образующих множество $\{Tr_k\}$. На практике величина N, как правило, не менее 100 и не более 200. Для каждого $\mathbf{tr} \in \{Tr_k\}$ определяются группы смежных каналов, внутри которых величины порогов tr не отличаются друг от друга на заранее заданную величину $\delta > 0$. Группы каналов такого типа называются *однородными*.

Для конкретного интервала $\mathbf{t}_p \in \mathbf{T}$, число групп однородных каналов, как правило, не более пятидесяти на один километр оптоволоконного сенсора, инсталлированного в области расположения протяженного объекта. Результат разбиения линии на однородные канальные группы называется пространственной конфигурацией фона на линии (ПКФЛ) и обозначается символом con_h, где h — индекс конфигурации. Полученные таким образом конфигурации образуют множество $\mathbf{Con} = \{con_h\}$. Каждая однородная группа каналов внутри con_h обозначает $\lambda(i, h)$, где i — индекс группы.

Шаг 6. Множество **Con**, методами кластерного анализа, разбивается на M кластеров, каждый из которых содержит условно близкие конфигурации разбиения линии, и формируется множество классов: Con_θ, где $\theta \in \mathbf{\Theta}$, $\mathbf{\Theta} = \{1, 2, \ldots, M\}$ — множество индексов классов. Классы Con_θ совместно образуют множество *Con_Set*. Для каждого класса $Con_\theta \in Con\_Set$, методом усреднения, формируется типичная ПКФЛ, которую обозначим символом $\mathbf{c}_\theta$. Величина $|\mathbf{c}_\theta|$ равна количеству групп однородных каналов в $\mathbf{c}_\theta$.

Шаг 7. Для **Con** определяется компактное метрическое пространство первичных признаков (Z, d), основанных на элементарной геометрии в пространстве R^1. В результате получим множество значений первичных признаков $\mathfrak{M}(\mathbf{\Theta}|\mathbf{Con})$ в (Z, d), соответствующее множествам $\mathbf{\Theta}$ и **Con**, $\mathfrak{M}(\mathbf{\Theta}) = \bigcup_\theta \Omega^{(\theta)}$.

Шаг 8. Множество $\mathfrak{M}(\mathbf{\Theta}|\mathbf{Con})$ разбивается на обучающую и тестовую выборки, которые используются для обучения и тестирования CLC $\Psi(\theta|f, P_c)$ соответственно. В качестве f используется обычный SVM, $P_c = 0{,}95$. Для каждого $\Omega^{(\theta)}$ формируется соответствующая конфигурация системы КИЦС. Конфигурация КИЦС представляет собой набор классификаторов целевых сейсмоакустических (виброакустических) событий, соответствующий ПКФЛ $\Omega^{(\theta)}$. Таким образом, формируется M конфигураций КИЦС, каждая из которых актуализируется в зависимости от текущей ПКФЛ. Для класса $Con_\theta \in Con\_Set$ верно следующее правило: чем меньше величина $|\mathbf{c}_\theta|$, тем удобнее и практичнее конфигурация $\mathbf{c}_\theta$, соответствующая классу Con_θ.

Шаг 9. В процессе определения текущей конфигурации КИЦС, для поканального набора $tr(\mathbf{t})$ определенного в течение интервала времени $\mathbf{t}$, вычисляется его Ψ-проекция на Θ в виде множества индексов $\Xi(\tilde{\theta}(tr(\mathbf{t}))|f, P_c)$. Таким образом, получен класс P_c-неразличимости ПКФЛ. Фактически, все предвычисленные ПКФЛ, включенные в $\Xi(\tilde{\theta}(tr(\mathbf{t}))|f, P_c)$, являются *равновозможными*, при этом каждая ПКФЛ подразумевает соответствующую конфигурацию КИЦС.

Шаг 10. Для стабилизации функционирования КИЦС, в качестве оценки текущей ПКФЛ выбирается следующее «осторожное» решение:

$$\theta^* = \mathrm{Arg} \max_{\theta \in \Xi(\tilde{\theta}(tr(\mathbf{t}))|f, P_c))} (|\mathbf{c}_\theta|).$$

Тогда ПКФЛ $\mathbf{c}_{\theta^*}$, состоящая из максимального на множестве $\Xi(\tilde{\theta}(tr(\mathbf{t}))|f, P_c)$ числа однородных групп $\{\lambda(i, \theta^*)\}$, отвечает *наиболее сложной* конфигурации КИЦС, что выражается в том, что в данном случае для обработки данных по линии используется наибольшее число автономно настроенных классификаторов (каждый классификатор настраивается на соответствующую группу $\lambda(i, \theta^*)$). Такая конфигурация является гарантом сохранения устойчивости КИЦС в условиях общей неопределенности относительно реальной конфигурации ПКФЛ.

На рисунке приведена диаграмма типа Waterfall, соответствующая тридцати каналам системы DAS-мониторинга, предназначенной для контроля трубопроводной конструкции в условиях динамичной шумовой обстановки. Диаграмма соответствует частотному диапазону [0, 10] Гц и построена по данным DAS-системы мониторинга, полученным в процессе полигонных испытаний в зимне-весенний период. На приведенной диаграмме отсутствуют целевые виброакустические сигналы, а вся зафиксированная в данном частотном диапазоне суточная виброакустическая активность вдоль оптоволоконного сенсора (линии) является следствием влияния фоновых шумовых процессов. Con(1), …, Con(6) обозначают схематические (иллюстративные) изображения последовательностей ПКФЛ (вертикальные полосы), а $\lambda(1)$, $\lambda(2)$, $\lambda(3)$ — однородные группы каналов внутри ПКФЛ Con(4) (цветные фрагменты полос).

На рисунке, в разделе цветовой шкалы, символ k обозначает третий порядок при характеристике интенсивности вибрационной динамики трубопроводной конструкции. Данная цветовая шкала имеет иллю-

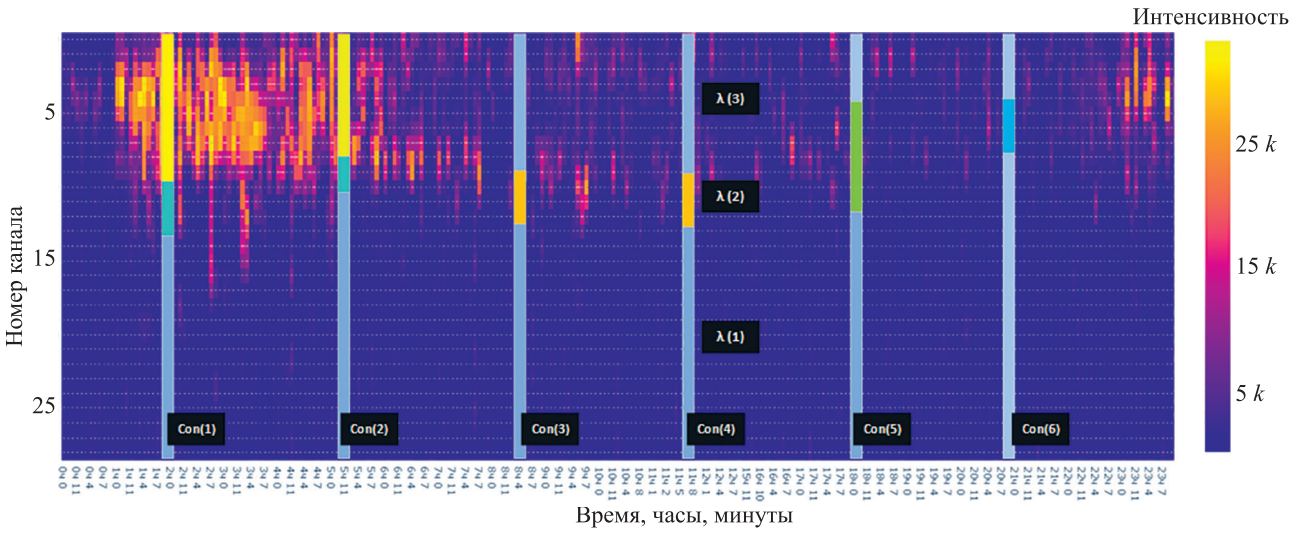

Рисунок. Суточная временная развертка интенсивности DAS-сигналов в полосе [0, 10] Гц по 30 каналам

Figure. Daily time plot of DAS signal intensity in the band [0, 10] Hz for 30 channels

стративный характер, который определяется простым правилом: цвета, с большими значениями интенсивности (справа), соответствуют большим амплитудам виброакустической динамики. Легко представить, что при $|CH| = 6000$ число однородных канальных групп, соответствующих конкретной ПКФЛ, вполне может составлять несколько сотен единиц. Таким образом, задача идентификации текущей ПКФЛ с точностью до класса из множества **Con** становится нетривиальной. Применение к данному случаю технологии, описанной шагами 1–10, показало, что, в среднем, при $|CH| = 2100$ (длина сенсора 10 200 м, ширина канала 5 м) величина $|\Xi(\tilde{\theta}(tr(\mathbf{t}))|f, P_c)|$ не превышает 10. Применение CLC на шаге 10, при среднем темпе изменения ПКФЛ на линии один раз в 5 мин, обеспечивает величины показателей качества классификации TPR (полнота) и PPV (точность) не хуже 0,95. Тестирование выполнено на наборе данных, собранном по результатам работы реальной DAS-системы мониторинга в течение трех суток. Список целевых виброакустических события для данного случая: механические удары по трубопроводной конструкции, пиление материала трубопроводной конструкции, имитация утечки транспортируемого агента. Данные типы целевых событий генерировались в дневное время, на фоне интенсивных фоновых шумов (отношение сигнал/шум не более 2–3 дБ). Среднее число целевых событий в течение часа: не менее 30. В условиях неиспользования CLC (в этом случае на шаге 10 применялся обычный точечный классификатор на базе SVM) величины показателей TPR и PPV не превышали 0,9. Таким образом, использование модели CLC позволило значительно повысить качество классификации целевых событий в условиях динамического шума высокой интенсивности.

Заключение

В работе рассмотрены основы нового подхода к решению задачи классификации стохастических объектов в мультиклассовой постановке для случая, когда часть классов плохоразличимы в пространстве первичных признаков. Предложенный подход позволил определить конечное множество классов, которое с заданным коэффициентом доверия содержит индекс класса, которому соответствует классифицируемый объект. Мощность этого множества, в которое *гарантированно* попадает искомый класс, зависит от множества факторов, в том числе от плотности распределения классов в пространстве первичных признаков, от степени различимости классов в этом пространстве, от величины заданного коэффициента доверия. Свойства предложенных алгоритмов строго доказаны, а экспериментальные исследования показали их практическую работоспособность. Величина кардинального числа $|\Xi|$ доверительного множества Ξ в значительной мере определена величиной константы Липшица L: чем меньше L, тем меньше величина $|\Xi|$. Для хорошо обученных систем величина L невелика, что на практике обуславливает приемлемо невысокое значение величины $|\Xi|$. Например, для качественно обученных систем классификации широкополосных шумоподобных сигналов, при $P_c = 0{,}9$, величина $|\Xi|$, как правило, не превышает 2–3 % от величины $|\Theta|$. Описанный в работе подход перспективен для применения в многоканальных системах обработки данных мониторинга протяженных объектов, в практической биометрии, а также и иных областях в случае, когда множество альтернативных классов состоит из сотен или даже тысяч элементов при условии, что некоторые классы являются плохоразличимыми в силу своей природы.

Литература

1. Korotaev V.V., Denisov V.M., Timofeev A.V. Analysis of seismoacoustic activity based on using optical fiber classifier // Proc. of the Latin America Optics and Photonics Conference. OSA, 2014. P. LM4A.22. https://doi.org/10.1364/LAOP.2014.LM4A.22
2. Aktas M., Akgun T., Demircin M.U., Buyukaydin D. Deep learning based multi-threat classification for phase-OTDR fiber optic distributed acoustic sensing applications // Proceedings of SPIE. 2017. V. 10208. P. 102080G. https://doi.org/10.1117/12.2262108
3. Korotaev V.V., Denisov V.M., Rodrigues J.J.P.C., Serikova M.G., Timofeev A. Monitoring of deep-sea industrial facilities using fiber optic cable // Proceedings of SPIE. 2015. V. 9525. P. 95253Q. https://doi.org/10.1117/12.2184741
4. Timofeev A.V., Denisov V.M. Multimodal heterogeneous monitoring of super-extended objects: modern view // Studies in Systems, Decision and Control. 2016. V. 62. P. 97–116. https://doi.org/10.1007/978-3-319-32525-5_6
5. Тимофеев А.В., Грознов Д.И. Классификация источников сейсмоакустической эмиссии в оптоволоконных системах мониторинга протяжённых объектов // Автометрия. 2020. Т. 56. № 1. С. 59–73. https://doi.org/10.15372/AUT20200107
6. Timofeev A.V. The guaranteed estimation of the Lipschitz classifier accuracy: Confidence set approach // Journal of the Korean Statistical Society. 2012. V. 41. N 1. P. 105–114. https://doi.org/10.1016/j.jkss.2011.07.005
7. Von Luxburg U., Bousquet O. Distance-based classification with Lipschitz functions // Journal of Machine Learning Research. 2009. V. 5. P. 669–695.

References

1. Korotaev V.V., Denisov V.M., Timofeev A.V. Analysis of seismoacoustic activity based on using optical fiber classifier. *Proc. of the Latin America Optics and Photonics Conference*, OSA, 2014, pp. LM4A.22. https://doi.org/10.1364/LAOP.2014.LM4A.22
2. Aktas M., Akgun T., Demircin M.U., Buyukaydin D. Deep learning based multi-threat classification for phase-OTDR fiber optic distributed acoustic sensing applications. *Proceedings of SPIE*, 2017, vol. 10208, pp. 102080G. https://doi.org/10.1117/12.2262108
3. Korotaev V.V., Denisov V.M., Rodrigues J.J.P.C., Serikova M.G., Timofeev A. Monitoring of deep-sea industrial facilities using fiber optic cable. *Proceedings of SPIE*, 2015, vol. 9525, pp. 95253Q. https://doi.org/10.1117/12.2184741
4. Timofeev A.V., Denisov V.M. Multimodal heterogeneous monitoring of super-extended objects: modern view. *Studies in Systems, Decision and Control*, 2016, vol. 62, pp. 97–116. https://doi.org/10.1007/978-3-319-32525-5_6
5. Timofeev A.V., Groznov D.I. Classification of seismoacoustic emission sources in fiber optic systems for monitoring extended objects. *Optoelectronics, Instrumentation and Data Processing*, 2020, vol. 56, no. 1, pp. 50–60. https://doi.org/10.3103/s8756699020010070
6. Timofeev A.V. The guaranteed estimation of the Lipschitz classifier accuracy: Confidence set approach. *Journal of the Korean Statistical Society*, 2012, vol. 41, no. 1, pp. 105–114. https://doi.org/10.1016/j.jkss.2011.07.005
7. Von Luxburg U., Bousquet O. Distance-based classification with Lipschitz functions. *Journal of Machine Learning Research*, 2009, vol. 5, pp. 669–695.

8. Fazlyab M., Robey A., Hassani H., Morari M., Pappas G.J. Efficient and accurate estimation of Lipschitz constants for deep neural networks // NIPS'19: Proc. of the 33rd International Conference on Neural Information Processing Systems. 2019. P. 11427–11438.
9. Latorre F., Rolland P., Cevher V. Lipschitz constant estimation of Neural Networks via sparse polynomial optimization // Proc. of the 8th International Conference on Learning Representations (ICLR 2020).
10. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов: Статистические проблемы обучения. М.: Наука, 1974. 415 с.
11. Advances in Large Margin Classifiers / ed. by A. Smola, P. Barlett, B. Scholkopf, D. Schuurmans. MIT Press, 2000. 412 p.
12. Hein M., Bousquet O., Schölkopf B. Maximal margin classification for metric spaces // Journal of Computer and System Sciences. 2005. V. 71. N 3. P. 333–359. https://doi.org/10.1016/j.jcss.2004.10.013
13. Bennett K.P., Bredensteiner E.J. Duality and geometry in SVM classifiers // Proc. of the International Conference on Machine Learning (ICML). 2000. P. 57–64.
14. Arens R., Eells J. On embedding uniform and topological spaces // Pacific Journal of Mathematics. 1956. V. 6. N 3. P. 397–403. https://doi.org/10.2140/pjm.1956.6.397
15. Hein M., Bousquet O. Maximal margin classification for metric space // Lecture Notes in Computer Science. 2003. V. 2777. P. 72–86. https://doi.org/10.1007/978-3-540-45167-9_7
16. Breiman L. Bagging predictors // Machine Learning. 1996. V. 24. N 2. P. 123–140. https://doi.org/10.1007/bf00058655
17. Rangel P., Lozano F., García E. Boosting of support vector machines with application to editing // Proc. of the 4nd International Conference on Machine Learning and Applications (ICMLA'05). 2005. https://doi.org/10.1109/icmla.2005.13

8. Fazlyab M., Robey A., Hassani H., Morari M., Pappas G.J. Efficient and accurate estimation of Lipschitz constants for deep neural networks. *NIPS'19: Proc. of the 33rd International Conference on Neural Information Processing Systems*, 2019, pp. 11427–11438.
9. Latorre F., Rolland P., Cevher V. Lipschitz constant estimation of Neural Networks via sparse polynomial optimization. *Proc. of the 8th International Conference on Learning Representations (ICLR 2020)*.
10. Vapnik V.N., Chervonenkis A.Ia. *Pattern Recognition Theory: Statistical Problems of Learning*. Moscow, Nauka Publ., 1974, 415 p. (in Russian)
11. *Advances in Large Margin Classifiers*. Ed. by A. Smola, P. Barlett, B. Scholkopf, D. Schuurmans. MIT Press, 2000, 412 p.
12. Hein M., Bousquet O., Schölkopf B. Maximal margin classification for metric spaces. *Journal of Computer and System Sciences*, 2005, vol. 71, no. 3, pp. 333–359. https://doi.org/10.1016/j.jcss.2004.10.013
13. Bennett K.P., Bredensteiner E.J. Duality and geometry in SVM classifiers. *Proc. of the International Conference on Machine Learning (ICML)*, 2000, pp. 57–64.
14. Arens R., Eells J. On embedding uniform and topological spaces. *Pacific Journal of Mathematics*, 1956, vol. 6, no. 3, pp. 397–403. https://doi.org/10.2140/pjm.1956.6.397
15. Hein M., Bousquet O. Maximal margin classification for metric space. *Lecture Notes in Computer Science*, 2003, vol. 2777, pp. 72–86. https://doi.org/10.1007/978-3-540-45167-9_7
16. Breiman L. Bagging predictors. *Machine Learning*, 1996, vol. 24, no. 2, pp. 123–140. https://doi.org/10.1007/bf00058655
17. Rangel P., Lozano F., García E. Boosting of support vector machines with application to editing. *Proc. of the 4nd International Conference on Machine Learning and Applications (ICMLA'05)*, 2005. https://doi.org/10.1109/icmla.2005.13

Автор

Тимофеев Андрей Владимирович — доктор технических наук, научный директор, ТОО «Эквалайзум», Астана, 010000, Казахстан, sc 56689367600, https://orcid.org/0000-0001-7212-5230, timofeev.andrey@gmail.com

Author

Andrey V. Timofeev — D.Sc., Scientific Director, LLP "EqualiZoom", Astana, 010000, Kazakhstan, sc 56689367600, https://orcid.org/0000-0001-7212-5230, timofeev.andrey@gmail.com

Статья поступила в редакцию 21.05.2023
Одобрена после рецензирования 24.06.2023
Принята к печати 17.09.2023

Received 21.05.2023
Approved after reviewing 24.06.2023
Accepted 17.09.2023