

УДК 621.391.2 : 519.24

КЛАССИФИКАЦИЯ ИСТОЧНИКОВ СЕЙСМОАКУСТИЧЕСКОЙ ЭМИССИИ В ОПТОВОЛОКОННЫХ СИСТЕМАХ МОНИТОРИНГА ПРОТЯЖЁННЫХ ОБЪЕКТОВ

© Тимофеев А. В., Грознов Д. И.

ООО «Флагман-Гео»,
197376, г. Санкт-Петербург, Набережная реки Карповки, 5Г, оф. 14
E-mail: timofeev.andrey@gmail.com

Предложен новый метод автоматической классификации источников сейсмоакустической эмиссии в оптоволоконных системах мониторинга, основанных на принципах оптической рефлектометрии во временной области с использованием восстановления фазы интерференционного сигнала во времени. Новизна подхода заключается в оригинальных принципах формирования пространства классификационных признаков и применении классификаторов ансамблевого типа. Такой метод работоспособен при условии применения сравнительно небольшой по объёму обучающей базы данных. Показана высокая практическая эффективность предложенного подхода при работе на реальных данных. Равномерно по различным классам целей достигнута величина надёжности классификации, прежде недоступная для альтернативных методов классификации, реализуемых в серийных оптоволоконных системах мониторинга.

Ключевые слова: классификация, XGBoost, SVM, DAS, ϕ -OTDR.

DOI: 10.15372/AUT20200107

Введение. В последнее время в практику мониторинга состояния протяжённых объектов с длиной контролируемого периметра от 20 км и более активно внедряются оптоволоконные сенсорные системы, основанные на использовании оптических квантовых эффектов, позволяющих измерять динамику сейсмоакустического поля в области прокладки сенсора системы. Сегодня в международной научно-технической литературе устоялось название для систем данного класса DAS (Distributed Acoustic Sensors) — распределённый акустический сенсор [1, 2], а также накоплен значительный опыт практической эксплуатации DAS-систем в различных приложениях, в том числе для контроля охранной зоны магистральных трубопроводов, мониторинга балластной призмы железнодорожных путей, магистральных каналов связи и т. д. [3–8]. С точки зрения деталей физического принципа их работы получившие широкое распространение мониторинговые DAS-системы подразделяются на несколько классов, отличающихся как типом используемого обратнорассеянного излучения (рэлеевского, рамановского или бриллюэновского), так и типом информационной составляющей обратнорассеянного излучения (амплитудой, фазой, поляризацией, частотой). Детальное описание этих классов выходит за рамки данной работы, укажем лишь, что широко применяемые практически системы принадлежат к одному из четырёх типов: C-OTDR (Coherent Optical Time Domain Reflectometer) с использованием амплитуды [9, 10], P-OTDR с использованием поляризации [11], ϕ -OTDR с использованием фазы [12] и BOTDA с использованием частоты [13]. В настоящий момент оптоволоконные мониторинговые системы особенно интенсивно развиваются в направлении фазочувствительных измерений (ϕ -OTDR-системы) [14], что в первую очередь позволяет компенсировать падение чувствительности систем для случая больших длин сенсора (более 20 км) и улучшить ещё целый ряд технических параметров по сравнению с классическими C-OTDR-системами.

К основным преимуществам оптоволоконных систем мониторинга по сравнению с аналогами, построенными, например, на сетях геофонов, следует отнести: их невысокую стоимость при решении мониторинговых задач на объектах большой протяжённости (более 20 км), невосприимчивость чувствительного элемента к электромагнитным воздействиям и помехам, скрытность функционирования системы (нет визуально выявляемых датчиков, система не обнаруживается металлодетекторами, отсутствует излучение), стабильность работы независимо от погодных условий (туман, дождь, снег, ветер). Дополнительным плюсом является возможность совмещения сенсора с линией связи, подсистемами температурного мониторинга или мониторинга деформаций. При этом системы оптоволоконного мониторинга в силу принципа их работы имеют сравнительно узкую полосу пропускания (как правило, лежащую в диапазоне 0–300 Гц), которая тем не менее остаётся достаточной для решения основных задач сейсмоакустического мониторинга. В самых общих чертах принцип работы таких датчиков достаточно прост [15–17]. В оптическое волокно, выполняющее функцию вибросенсора и проложенное по линии мониторинга (по периметру охраняемого объекта или вдоль протяжённого объекта), посылается световой импульс длительностью 50–150 наносекунд и принимается обратнорассеянный сигнал. Как было указано выше, тип DAS-системы может быть амплитудным, фазовым, частотным или поляризационным. В зависимости от её типа для обратнорассеянного сигнала определяются соответствующие характеристические параметры, которые описывают отражающие способности волокна на различных его участках, длина которых лежит в интервале от 1 до 10 метров. В соответствии с типом системы в качестве таких параметров может выступать амплитудный, частотный, фазовый или поляризационный показатель обратнорассеянного интерференционного сигнала. Вследствие влияния внешнего механического воздействия (например, сейсмоакустического), приложенного на конкретном сегменте сенсора, на этом сегменте возникает локальная девиация коэффициента рассеяния. В DAS-системах различных типов обобщённым параметром, объективно характеризующим эту локальную девиацию, считается некоторая аналитическая функция от разности значений параметров, соответствующих одному и тому же сегменту сенсора в смежных циклах зондирования. Вид такой аналитической функции зависит от типа DAS-системы и является ноу-хау производителей DAS-систем. Будем называть этот обобщённый параметр информационным параметром (ИП) или сигналом. Таким образом, для DAS-систем различного типа ИП имеет разную физическую природу. Отражающая способность отдельных участков оптоволоконной линии связана с величиной локального коэффициента рассеяния света на этих участках (сегментах). Локальный коэффициент рассеяния, в свою очередь, определённым образом зависит от частоты и интенсивности вибраций внешнего механического воздействия на поверхность сенсора. В качестве такового воздействия может выступать, например, сейсмоакустическое поле грунта, в который погружён оптоволоконный сенсор. Путём импульсного зондирования оптоволоконного сенсора с определённой частотой (на практике ~ 2000 Гц) при измерении значений соответствующего ИП восстанавливается амплитудно-частотная характеристика (АЧХ) сейсмоакустического поля, воздействующего на сенсор. Физика механизма работы DAS-систем с моноמודовым сенсором довольно сложная. На процесс формирования обратнорассеянного сигнала влияет множество факторов, возникающих вследствие внешнего механического воздействия (например, сейсмоакустического поля) на сенсор. Кроме того, влияют эффекты сжатия/растяжения оптоволоконной линии, ведущие к микрофлуктуациям длины и поперечного сечения сенсора, а также к вариациям локального коэффициента преломления света в волокне. Эффекты изменения локального местоположения центров рэлеевского рассеяния (микровкраплений) влияют на девиацию локального коэффициента рассеяния [16]. С практической точки зрения DAS-системы сейсмоакустического мониторинга отличаются друг от друга типом, ценой, доступностью для потребителя, динамическим диапазоном,

величиной пространственного разрешения и длиной сенсора, которую способен обслуживать один интеррогатор (аналитический блок) системы. На рынке представлены следующие технические решения [5, 6]: «Волк», «Волкодав», «Волна Альфа», «Ворон», «Дунай», «Омега», «Оптолекс», «Сокол», «FiberPatrol», «Optosens», «Silixa iDAS» и система «ОХУ» [5].

Одной из самых важных задач, кардинально влияющей на практическую эффективность системы сейсмоакустического мониторинга, несомненно, является задача высокондёжной классификации типа источника сейсмоакустической эмиссии (ИСЭ), который был предварительно обнаружен в зоне действия системы. К числу факторов, осложняющих решение классификационной задачи динамических ИСЭ (пешеход, автомобиль, гусеничная техника и прочее), следует отнести: сравнительно узкий частотный диапазон измерений (0–300 Гц), который присущ оптоволоконным системам мониторинга; принципиальную невозможность длительного (более 5–30 секунд) накопления данных для принятия классификационного решения; неполную достоверность восстановления АЧХ целевых ИСЭ. Все эти факторы существенно осложняют применение в оптоволоконных мониторинговых системах классических методов распознавания (классификации) целевых объектов, основанных на использовании спектральных характеристик их сейсмоакустических образов, восстановленных по динамике ИП.

В работах [18–29] задача классификации ИСЭ решается на базе использования стандартных методов восстановления АЧХ ИСЭ (часть методов обсуждается далее).

Для предлагаемой в данной работе ϕ -OTDR-системы сейсмоакустического мониторинга этот подход не обеспечивает высокого качества классификации равномерно по всем классам ИСЭ, что обусловлено нелинейными искажениями АЧХ ИСЭ, вносимыми приёмным трактом системы мониторинга. Здесь рассматривается практически эффективный подход, который на основе данных ϕ -OTDR-системы мониторинга обеспечивает высокие показатели надёжности решения классификационной задачи равномерно по всем классам ИСЭ в условиях ограниченной по объёму обучающей базы. Предложенный подход основан на оригинальном наборе классификационных признаков, а также на использовании современных методов машинного обучения, что в совокупности обеспечило его высокую практическую эффективность.

1. Постановка задачи. Обозначим $O = \{o_1, o_2, \dots\}$ множество объектов-источников сейсмоакустической эмиссии, каждый из которых принадлежит ровно к одному классу, $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$ — конечное множество индексов классов, которым принадлежат объекты из множества O . Существует некая априорно неизвестная наблюдателю индикаторная функция χ , которая отображает множество O на множество Θ , сопоставляя любому элементу $o \in O$ ровно один элемент $\theta \in \Theta$. Иначе $\forall o \in O (\exists! \theta \in \Theta) \wedge (\chi(o) = \theta)$. Объект $o \in O$ априорно неизвестного класса генерирует сейсмоакустическую эмиссию, которая регистрируется протяжённым оптоволоконным сенсором ϕ -OTDR-системы мониторинга, основанной на принципе когерентной рефлектометрии с использованием восстановления фазы интерференционного сигнала во времени [1].

Система мониторинга, базируясь на накопленных в течение априорно заданного интервала времени наблюдениях, относящихся к объекту классификации в полностью автоматическом режиме, должна определить класс, к которому из множества Θ принадлежит наблюдаемый объект o . Другими словами, система мониторинга должна аппроксимировать априорно неизвестную индикаторную функцию χ . Предполагается, что ИСЭ $o \in O$, подлежащие автоматической классификации и совместно образующие множество O , находятся на априорно неизвестных расстояниях от сенсора системы мониторинга. Кроме того, априорно неизвестна величина коэффициента поглощения среды (почвы), в которую заложен оптоволоконный сенсор системы мониторинга. Среда распространения сейсмоакустической эмиссии подразумевается анизотропной.

2. Используемые понятия, определения и особенности подхода. Каналом ϕ -OTDR-системы мониторинга называется линейный, ограниченный по длине участок сенсора такой, что система мониторинга в силу принципа её работы и выставленных настроек способна получать и анализировать данные от этого участка сенсора отдельно от других аналогичных участков. Данные, принятые от конкретного канала, в определённой мере характеризуют стохастическую динамику сейсмоакустического поля в области его расположения. Ширина ϕ -OTDR-канала (разрешающая способность по пространству) зависит от длительности зондирующего импульса когерентного лазера, а также от целого ряда других физических параметров системы. Как правило, ширина канала меньше или равна физическому разрешению системы по пространству. Измерительная информация, которая поступает от ϕ -OTDR-канала в систему, оцифровывается и используется в качестве характеристики параметров сейсмоакустического поля в области пространственного расположения канала. Индексы каналов системы совместно образуют множество N . Особенностью предлагаемого подхода к классификации ИСЭ является то, что в его рамках не предполагается непосредственного оценивания параметров АЧХ ИСЭ по отдельным каналам системы мониторинга. В силу физических принципов, заложенных в основу ϕ -OTDR-систем мониторинга, АЧХ сейсмической волны, приходящей от ИСЭ $o \in O$ на группу смежных каналов мониторинговой системы, не могут быть измерены с приемлемой точностью по данным одного канала системы (хотя ряд производителей утверждает обратное). В настоящее время сравнительно низкая точность измерений обусловлена целым комплексом проблем, в том числе и самим принципом когерентной рефлектометрии во временной области с использованием восстановления фазы интерференционного сигнала во времени. В процессе прохождения зондирующего импульса когерентного лазера по телу сенсора и его обратного рассеяния на группах микровкраплений (центрах интерференции), которые всегда присутствуют в оптическом волокне сенсора вследствие специфики его производства, происходит множественная интерференция света на частоте лазера. Результатом этой интерференции являются так называемые интерференционные картинки (спеклы), часть энергии которых фиксируется приёмником системы. Последовательность спеклов, измеренная в последовательные моменты времени, является основой для формирования ИП, каким бы способом он ни определялся. Известно, что спекл, зависящий от взаимного расположения центров интерференции внутри тела конкретного канала, при внешнем воздействии на поверхность сенсора меняет свою конфигурацию по отношению к своему предшествующему состоянию существенно нелинейным образом (практически хаотично). Поэтому даже при малых механических воздействиях на волокно разница спеклов, соответствующая смежным циклам зондирования, может быть непропорционально велика (или, наоборот, непропорционально мала) по отношению к энергии приложенного воздействия. Эта непропорциональность, представляющая собой существенно нелинейное искажение исходного воздействия, неминуемо окажет нелинейное влияние на величину определённого ИП, каким бы способом он ни вычислялся и какие бы математические методы ни использовались для сглаживания неизбежных всплесков его значений, соответствующих смежным интервалам времени измерения. В этой связи АЧХ сейсмического колебания, оказывающего вибрирующее воздействие на сенсор, восстанавливаются довольно приблизительно даже для высоких частот зондирования оптического волокна импульсным лазером системы. Кроме того, из-за сравнительно протяжённых величин ширин каналов (5–10 м) акустическая волна на такой канал приходит с разной фазой, что также вносит существенные искажения в процесс оценивания частоты сейсмоакустического колебания. Таким образом, в случае измерения АЧХ сейсмоакустического поля средствами ϕ -OTDR-системы мониторинга имеем нелинейно искажённую оценку АЧХ ИСЭ. Наш практический опыт работы с ϕ -OTDR-данными показывает, что даже при использовании измерений такого качества можно с высокой достоверностью классифицировать типы ИСЭ.

В рамках предлагаемой концепции каждый канал ϕ -OTDR-системы мониторинга в каждый момент времени t несёт информацию следующего типа: 1 — в канале обнаружен сигнал, 0 — в канале сигнал не обнаружен. Для удобства будем обозначать этот факт с помощью простой индикаторной функции:

$$Ind_j(t) = \begin{cases} 1, & \text{в момент } t \text{ сигнал обнаружен;} \\ 0, & \text{в момент } t \text{ сигнал не обнаружен.} \end{cases}$$

Здесь j — номер канала. Как показывает практика, информации, которую несёт в себе индикаторная функция $Ind_j(t)$, вполне достаточно для надёжной классификации типов ИСЭ. Алгоритмы предварительного, поканального обнаружения ИСЭ, а также принципы выбора порога, отделяющего смесь сигнала с шумом от чистого шума, подробно изложены в [3]. Отметим, что практически использовались два подхода: классический с адаптивной настройкой порога в рамках последовательного анализа, а также подход на базе машинного обучения (ML, Machine Learning), когда в процессе обнаружения цели решается задача бинарной классификации — класс «сигнал + шум» против класса «шум». Для решения этой задачи применялся бинарный SVM-классификатор. Классический подход с адаптивным порогом значительно менее требователен к вычислительным ресурсам, а метод обнаружения на базе ML обеспечивает низкие показатели ошибок первого рода ($< 0,01$).

Пусть в зоне чувствительности системы появился объект o . Обозначим $T = \{0, 1, 2, \dots\}$ как последовательность моментов времени, в которых система мониторинга считывает с оптоволоконного сенсора сигналы от объекта o . В момент времени $t \in T$ сейсмоакустическая эмиссия от объекта o одновременно отражается в одном или сразу в нескольких каналах системы мониторинга, индексы которых совместно образуют множество $N_t = \{n | Ind_n(t) = 1\} \subseteq N, |N_t| \geq 1$. При этом за счёт влияния канального шума, а также воздействия других ИСЭ часть каналов с индексами из множества $N_t \subseteq N$ может не иметь отношения к эмиссии от объекта o . Другими словами, в момент времени $t \in T$ эта группа каналов предоставляет недостоверную информацию, по сути, являясь генератором помех. Эта ситуация является обычной для мониторинговых систем рассматриваемого типа, поэтому система классификации должна быть устойчива к недостоверностям такого рода в наблюдаемых данных.

Координаты j -го канала представим в виде кортежа $\mathbf{c}_j = \langle c_{jl}, c_{jr} \rangle$, где c_{jl}, c_{jr} — координаты его крайних точек; $c_{jl}, c_{jr} \in R^m$; R^m — координатное пространство; $c_{jc} = (c_{jl} + c_{jr}) / 2$ — координаты арифметического центра j -го канала. Координаты арифметического центра группы смежных каналов с индексами из множества $N' \subseteq N$: $c_c(N) = \sum_{j \in N'} c_{jc} |N'|^{-1}$, δ — ширина ϕ -OTDR-канала, $|A|$ — кардинальное число множества A .

Сейсмоакустическая эмиссия от объекта o поступает на поверхность сенсора неравномерно как по времени, так и по пространству. Именно эта неравномерность и несёт в себе классификационные признаки типа объекта o . Введём понятие отметки, которое является центральным в предлагаемой концепции. Отметка с индексом i представляет собой зафиксированное системой мониторинга возбуждение сейсмоакустического поля, которое продолжалось в течение временного интервала $[t, t + \Delta_i]$, было отражено в $|N_i|$ смежных каналах приёмной системы и поэтому имеет ширину $\theta_i = \delta |N_i|$. Длительность отметки Δ_i — крайне важный параметр для решения классификационной задачи. Поэтому разрешение приёмной системы по времени должно быть таким, чтобы обеспечить гарантированный приём сигналов от ИСЭ, имеющих импульсно-ударный характер (удар ломом, удар лопатой, человеческий шаг), длительность которых не превышает 0,1 секунды. Формально будем называть отметкой от объекта o набор параметров $\lambda_i(o) = (t_i, \theta_i, \Delta_i, c_c(N_i))$, где i — индекс отметки, o — ИСЭ (ненаблюдаемый параметр), Δ_i — длительность отметки, θ_i — ширина отметки, t_i — время обнаружения i -й отметки, $c_c(N_i)$ — арифметический

центр группы каналов N_i . Рассмотрим множество $T_o = \{t_1^{(o)}, \dots, t_n^{(o)}\}$, представляющее собой последовательность моментов времени, в течение которых система наблюдала последовательность отметок $\{\lambda_i(o)\}$, соответствующих объекту o , а $I_o = \{1, \dots, n\}$ — множество индексов T_o . В рамках предложенного подхода для решения классификационной задачи использовался поток отметок, соответствующих объекту o , представленный в виде $M(o|T_o) = \{\lambda_i(o)|i \in I_o\}$ от объекта $o \in O$. Распределение временных моментов на множестве T_o также является важным информативным признаком для классификации, неявно характеризующим частотные свойства объекта o .

Для того чтобы отметки $M(o|T_o)$ можно было использовать в алгоритмах классификации, необходимо произвести над ними операцию центрирования: от абсолютных показателей для величин $c_c(N_i)$ и элементов множества T_o перейти к относительным. Введём операцию центрирования применительно к отметке $\lambda_i(o)$ в следующем виде:

$$\mathbf{S}(\lambda_i(o)|d, z) = (t_i - d, \theta_i, \Delta_i, c_c(N_i) - z), d \in R^1, z \in R^m.$$

Результаты применения операции центрирования к последовательности отметок $M(o|T_o)$:

$$M^c(o|T_o, d, z) = \mathbf{S}(M(o|T_o)|d, z) = \{\mathbf{S}(\lambda_i(o)|d, z)|i \in \text{Ind}(T_o)\}, d \in R^1, z \in R^m,$$

где $M^c(o|T_o) = M^c(o|T_o, t_1^{(o)}, \mathbf{c}_c(T_o))$ — центрированная последовательность наблюдений. Здесь $t_1^{(o)}$ — первый момент времени, когда $o \in O$ был обнаружен, а $\mathbf{c}_c(T_o)$ — арифметический центр величин из множества $\{c_c(N_i)|i \in \text{Ind}(T_o)\}$: $\mathbf{c}_c(T_o) = \sum_{i \in \text{Ind}(T_o)} c_c(N_i) |T_o|^{-1}$. Таким образом, $M^c(o|T_o)$ является центрированной версией потока $M(o|T_o)$. Именно поток отметок $M^c(o|T_o)$ используется в качестве исходных данных для решения классификационной задачи.

При решении задачи классификации ИСЭ [18–29] формирование классификационных признаков производится на базе различных методов обработки данных: быстрое преобразование Фурье (БПФ) и аналогичные ему методы спектрального анализа, вейвлет-преобразование, метод главных компонент, модель смеси гауссовых распределений (GMM) и другие. Эти методы для удобства условимся называть синтетическими методами обработки. Например, в работе [18] для первичной обработки данных используется БПФ, в [19] — вейвлет-преобразование, в [20] — комбинация вейвлет-преобразования и метода главных компонент. Зачастую при формировании схемы обработки делаются попытки механистически реализовать те схемы, которые обеспечили приемлемые результаты при решении задач из других прикладных областей. Например, в [21] на этапе выделения признаков используется MFCC-преобразование, которое изначально было разработано для идентификации диктора, традиционно применяется в составе связки MFCC–GMM–SVM и сводится к проецированию спектра сигнала на так называемую mel-шкалу. Этот приём позволяет выделить наиболее значимые для восприятия человеком частоты. Данное преобразование хорошо зарекомендовало себя в задачах идентификации диктора, однако его применение при работе с сейсмоакустическими сигналами выглядит странно: в таком приложении достаточно взять обычный линейный кепстр.

В [22] для решения задачи классификации сейсмоакустических событий берётся схема обработки данных, очевидно, калькированная с известной комбинации БПФ–CNN (здесь CNN — конволюционная нейронная сеть), которая показывает неплохие результаты в задачах классификации оптических изображений. Для того чтобы обеспечить работоспособность этой схемы в приложении к анализу DAS-данных, в [22] наблюдения ϕ -OTDR-системы искусственным образом преобразуются в некое «изображение». В целом данная

схема может быть вполне работоспособной, если не учитывать тот факт, что ϕ -OTDR-система нелинейно искажает АЧХ исходных сейсмоакустических сигналов. В силу специфики функционирования ϕ -OTDR-системы один и тот же ИСЭ на разных участках оптоволоконного сенсора будет неизбежно формировать образы с неполностью идентичными АЧХ, стохастическая динамика ИП которых (в случае ϕ -OTDR-системы, основанной на восстановлении фазы интерференционного сигнала) будет воспроизводиться нестабильно, со значительными нелинейными искажениями. Что происходит, если к последовательности измеренных таким образом ИП применить синтетические процедуры обработки данных? В результате обработки векторы выходных параметров этих процедур в той или иной мере неизбежно отразят искажения, внесённые в последовательность измерений ИП приёмным блоком ϕ -OTDR-системы. Причём различные каналы системы для одного и того же источника сейсмоакустической эмиссии будут давать существенно различный набор величин компонент вектора выходных параметров. Далее этот вектор может быть или непосредственно использован в качестве классификационного признака, или на его основе могут быть сконструированы классификационные признаки более высокого уровня агрегирования. В любом случае полученные таким образом признаки будут в значительной степени искажены. Поэтому классификатор, например CNN, будет обучен на искажённых и нестабильных образах целевых ИСЭ. Этот фактор существенно снижает эффективность любой процедуры классификации, для обучения которой были взяты результаты применения процедур синтетической обработки измерений ИП, вычисляемых отдельно по каждому каналу. В этой связи частотно-временной анализ, реализуемый стандартными методами в отдельных каналах системы, например БПФ, вейвлет-преобразованиями или иными методами, основанными на использовании одноканальных оценок ИП, малоинформативен для задач классификации типа источника сейсмоакустической эмиссии в ϕ -OTDR-системах. Отчасти это является причиной неравномерного качества работы классификаторов, обученных с использованием преобразованных поканальных оценок ИП для различных классов ИСЭ в DAS-системах мониторинга. Так, в [22] для двух из шести классов указана сравнительно невысокая точность классификации (ниже 90 %). В [23] для двух классов из семи определена точность классификации также ниже 90 %. В качестве классификаторов традиционно широко используются: SVM [24], метод ближайшего соседа [25], GMM [18], а также нейронные сети различных топологий [22, 23]. Нейронные сети в рамках развития концепта глубокого обучения становятся всё более популярным средством решения классификационных задач в различных прикладных областях. Особенно впечатляющие результаты достигнуты при применении нейронных сетей в распознавании оптических изображений. В частности, последние годы на международных соревнованиях по распознаванию образов на изображениях (Large Scale Visual Recognition Challenge) выигрывали исключительно нейросетевые алгоритмы. При этом следует отметить, что для полноценного обучения алгоритмов данного класса требуются обучающие корпуса достаточно большой мощности. В работе [23] для обучения конволюционной нейронной сети использовались корпуса мощностью 487 000 образцов. Однако в большинстве практических приложений DAS-систем обучающие корпуса столь значительного объёма практически недоступны ввиду высокой стоимости их создания. Как правило, на практике мощность обучающих корпусов по каждому классу цели редко превышает величину 200–300 образцов. В этом случае практическое использование нейросетевых алгоритмов не представляется возможным, так как рассматриваемый тип алгоритмов не может быть эффективно обучен при малых мощностях обучающих корпусов. В то же время целый ряд современных ансамблевых алгоритмов, например XGBoost [30], сохраняет работоспособность при сравнительно небольших мощностях обучающих корпусов. Именно алгоритмы этого типа показали наибольшую эффективность в целом ряде практических приложений DAS-систем [3–6].

Оригинальность подхода, предложенного в данной работе, заключается в том, что

комплексный, частотно-временной анализ ИСЭ реализуется с помощью специального объекта высокого уровня агрегированности: потока отметок $M^c(o|T_o)$. По отношению к данным одноканальных измерений ИП объект «отметка» является более агрегированным, так как при его конструировании всегда используются измерения нескольких смежных каналов. Такой подход позволяет частично компенсировать некорректность измерения АЧХ сейсмоакустического поля ИСЭ средствами ϕ -OTDR-системы, так как ширина отметки (измеряемая с точностью до ширины ϕ -OTDR-канала) пусть приблизительно, но объективно характеризует мощность источника излучения. Как следует из наших исследований, показатели пространственно-временного распределения отметок в потоке $M^c(o|T_o)$ сами по себе являются очень мощными и оригинальными классификационными признаками. В итоге объект «поток отметок» представляет собой информативную базу для генерации оригинальных классификационных признаков, которые позволяют решить задачу классификации с высокой эффективностью, не требуя значительных затрат вычислительных ресурсов, а использование ансамблевых классификаторов обеспечивает возможность высокоэффективного обучения при применении обучающих корпусов сравнительно небольшой мощности.

3. Классификация типов источников сейсмоакустической эмиссии. Основными этапами построения системы классификации образцов являются:

- сбор размеченной базы «сырых» данных, предназначенной для обучения классификатора методами ML;
- генерация набора признаков классификации;
- сокращение (оптимизация) этого набора специальными методами;
- выбор алгоритма классификации и его обучение методами ML, в результате которого определяются оптимальные значения параметров классификатора, обеспечивающие приемлемые значения показателей качества классификации (метрики).

3.1 Сбор помеченной базы данных. Записываются, собираются и архивируются образцы сырых данных, снимаемых интеррогатором с сенсора ϕ -OTDR-системы, соответствующих определённым классам источников сейсмоакустической эмиссии. При этом каждая запись помечается маркером типа источника. Сбор базы — крайне важный этап, который невозможен без использования дорогостоящего полигона. База данных, которая применялась для обучения системы классификации, собрана с двух полигонов. Сенсор первого полигона был установлен на расстоянии 3–5 метров от края балластной призмы железнодорожных путей на глубине 30–50 см, почва: щебень, песок, глина. Длина сенсора равнялась 1500 м. Полигон был развёрнут по адресу «ПК-2, перегон ст. Сороковая — Астана, 497-й километр» для настройки параметров ϕ -OTDR-системы, предназначенной для железнодорожных применений в рамках пилотного проекта с АО «НК "Казакстан темір жолы"». Сенсор второго полигона был развёрнут вблизи аналога пограничной разделительной полосы на глубине 30–50 см в целях настройки параметров ϕ -OTDR-системы для периметровых приложений. Его длина составляла 5000 м, почва: плотный песок, щебень, суглинок. По этой полосе, кроме пешеходов, перемещались автомобили и лёгкие колёсные трактора. Согласно условиям технического задания длительность каждой записи, соответствующая определённому типу события (классу), равнялась 5 с. Наблюдая обнаруженный объект в течение такого же времени, настроенный классификатор должен определить тип объекта, отнести его к одному из пяти классов. Каждому типу класса соответствовали реализации потока отметок $M^c(o|T_o)$ (запись), наблюдаемые в течение 5 секунд. Настройки ϕ -OTDR-системы позволяли фиксировать сейсмоакустические события в частотном диапазоне от 0 до 500 Гц. Таким образом была собрана база данных, содержащая записи о семи типах событий: «пешеход» (101 запись, дистанция до сенсора от 0 до 7 м), «ручная копка грунта» (50 записей, дистанция до сенсора 1–10 м), «легковой автомобиль» (124 записи, дистанция до сенсора 10–50 м), «колёсный трактор» (145 записей, дистанция до сенсора

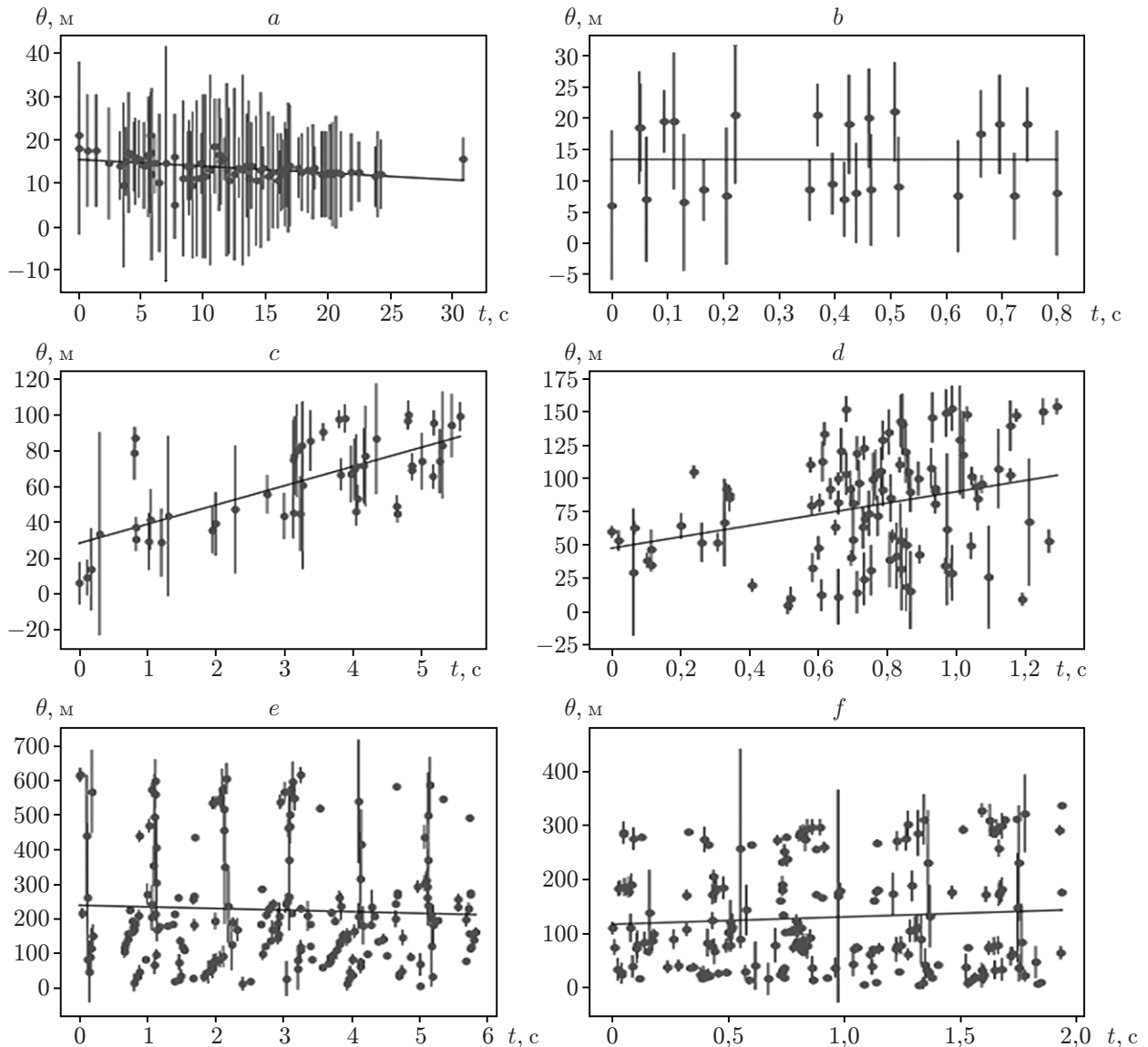


Рисунок. Примеры типичных потоков отметок для шести классов источников сейсмоакустической эмиссии: пешеход (а), ручная копка (b), автомобиль (c), колёсный трактор (d), электричка (e), товарный поезд (f)

10–50 м), «электричка» (150 записей, дистанция до сенсора 3–5 м), «товарный поезд» (150 записей, дистанция до сенсора 3–5 м), «фоновый класс» (200 записей). В «фоновый класс» были включены сейсмоакустические события, которые не являлись целевыми и поэтому не попали в первые шесть классов. В целом события данного класса можно рассматривать как фоновые или помеховые. В него вошли записи работы экскаваторной и гусеничной техники, технологические работы, звуки работы перфораторов, мощные каналные помехи, которые время от времени спонтанно возникали в каналах системы мониторинга, а также всевозможные комбинации этих событий. Визуальные представления типичных потоков отметок для шести классов отражены на рисунке.

3.2 Классификационные признаки. При выборе набора признаков, которые используются для автоматической классификации объектов, решались две последовательные задачи. Сначала на основе параметров потока $M^c(o|T_o)$ осуществлялась генерация набора при-

знаков классификации. Успешность генерации признаков в значительной мере обусловлена глубиной понимания разработчиками физических процессов, лежащих в основе генерации «сырых» данных. Опишем использованные признаки более подробно, так как их выбор — один из центральных вопросов при решении задачи автоматической классификации источников сейсмоакустической эмиссии.

Обозначим $n = |T_o|$ — количество отметок в потоке $M^c(o|T_o)$, а $\Delta T_o = t_n^{(o)} - t_1^{(o)}$ — длительность временного интервала наблюдения потока $M^c(o|T_o)$.

Для произвольного непустого случайного множества $X \subseteq R^1$ определим вектор-функцию $\mathbf{V}(X) = (Me(X), Mn(X), Mh(X), \sigma(X), Range(X)) \in R^5$, компонентами которой являются: $Me(X)$ — медиана множества X , $Mn(X)$ — среднее значение множества X , $Mh(X)$ — среднее гармоническое множества X , $\sigma(X)$ — оценка стандартного отклонения на множестве X , $Rang(X) = Max(X) - Min(X)$. Рассмотрим следующие множества: $T_o^c = \{t_i - t_1^{(o)} | i \in Ind(T_o)\}$, $\theta_o^c = \{\theta(t_i) - c_c(T_o) | i \in Ind(T_o)\}$, $\Delta_o^c = \{\Delta_i | i \in Ind(T_o)\}$, $C_o^c = \{c_c(N_i) - c_c(T_o) | i \in Ind(T_o)\}$.

Элементы множества $T_o^c \otimes \theta_o^c$ могут быть интерпретированы как моменты времени t_i , в которых производятся зашумлённые измерения арифметического центра координат $c_c(N_i)$ динамического объекта $o \in O$. В этом случае символом $\nu(T_o^c \otimes \theta_o^c)$ будем обозначать локальную оценку проекции скорости этого объекта на сенсор, полученную методом линейной интерполяции. Рассмотрим следующие эвристические показатели, которые характеризуют поток $M^c(o|T_o)$: $D_\Delta = n Me(\Delta_o^c) (\Delta T_o \sigma(\Delta_o^c))^{-1}$ — характеристика плотности распределения длительностей отметок во времени; $D_\theta = n (\Delta T_o Me(\theta_o^c))^{-1}$ и $D_{\theta L} = n (\Delta T_o Rang(\theta_o^c))^{-1}$ — параметры пространственно-временного распределения отметок. Были рассмотрены также иные эвристические показатели, характеризующие поток $M^c(o|T_o)$. После того как множество признаков для классификации было сформировано, реализовывалось его обоснованное сужение путём отбора наиболее информативных признаков. Фактически на этом этапе использовалось несколько методов, три так называемых «фильтрующих»: хи-квадрат, корреляция Пирсона и дисперсионный анализ (ANOVA), а также адаптационный метод обратного исключения (Backward Elimination). Результирующий набор признаков, которые в итоге применялись для классификации типа источника сейсмоакустической эмиссии, имеет следующий вид: $\{D_\Delta, D_\theta, D_{\theta L}, \nu(T_o^c \otimes \theta_o^c), \mathbf{V}(\Delta_o^c), \mathbf{V}(C_o^c), \mathbf{V}(\theta_o^c)\}$.

3.3 Алгоритмы классификации. Многочисленные эксперименты показали высокую практическую эффективность хорошо известных методов: мультиклассовый Support Vector Machine (SVM), два ансамблевых метода XGBoost [30] и Gradient-Boosting (GB) [31, 32]. Остальные методы в силу различных причин на использованной базе данных обеспечивали существенно более низкие показатели качества решения классификационной задачи. В частности, классификатор на основе многослойной нейронной сети (ANN) на исследуемом корпусе уступает вышеупомянутой триаде более чем на 5 % по показателю «Вероятность правильной классификации» вследствие недостаточного объёма обучающей выборки. В случае ANN объём обучающей выборки является фактором, радикально определяющим эффективность классификации. Все эти методы, особенно метод SVM (Multi-class SVM), имеют внятное математическое обоснование, которое ввиду большого объёма здесь не приводится.

Для оценивания обобщающей способности классификаторов применялась стандартная схема скользящего контроля, иначе называемая CV (Cross Validation) в варианте LOO (Leave-One-Out). Для классификатора SVM с учётом мультиклассовой постановки задачи использовалась стратегия «один против всех» (one-vs-rest). Основные параметры алгоритмов, которые иначе называются гиперпараметрами, оптимизировались с помощью модуля Hyperopt библиотеки Sklearn, Python 3.6. Например, в классификаторе XGBoost гипер-

параметрами являются «eta» (аналог «learning rate»), «max_depth», «min_child_weight», «gamma». Для оценивания качества работы алгоритмов классификации был взят универсальный показатель $P_{ПК}$ (вероятность правильной классификации), который оценивался численно для каждого класса на этапе тестирования алгоритмов классификации. Используемая для обучения база прецедентов (сигналов) сравнительно неплохо сбалансирована: отсутствует значительная разница в мощности представленных в ней классов. По этой причине показатель $P_{ПК}$ в качестве основной классификационной метрики в данном случае вполне корректен. Тем не менее в процессе исследования активно применялись и иные традиционные метрики качества классификации, в том числе оценки полноты (recall) и параметрическая F -мера, которые в силу хорошей сбалансированности обучающего набора ожидаемо не привнесли дополнительной информации в процесс оценивания качества работы классификационного алгоритма. Именно метрика $P_{ПК}$ в контексте целей настоящей работы является наиболее показательной, хорошо понимается и интерпретируется как разработчиками, так и эксплуатантами мониторинговых систем, а также является объективным интегральным показателем качества работы системы классификации.

Результаты классификации сведены в таблицу. Как следует из приведённых в ней данных, выбранная тройка классификаторов, обученная на помеченной базе данных, содержащей записи сигналов от источников шести типов, уверенно классифицирует события этих типов. Для авторов было полной неожиданностью то, что классификаторы достаточно надёжно отличают по звуковому отрезку длительностью 5 с электричку от товарного поезда, а также то, что класс «пешеход» классифицируется 100 % надёжно: нет путаницы с близким ему по амплитудно-частотным характеристикам классом «ручная копка грунта». В первой секции таблицы представлены результаты классификации семи классов. Во второй секции содержатся результаты классификации для шести классов: классы «электричка» и «товарный поезд» были объединены в единый класс — «поезд». В третьей секции таблицы приведены результаты классификации для пяти классов: классы «лёгкий колёсный трактор» и «легковой автомобиль» были объединены в единый класс — «колёсная техника». В целом результаты свидетельствуют о том, что три отобранных метода классификации уверенно определяют тип класса по его образцу пятисекундной длительности. Например, в случае шести классов мультиклассовый SVM (Multi-class SVM) показал абсолютный по точности результат. Использование методологии CV гарантировало то, что обобщающая способность обученных классификаторов является достаточно высокой, т. е. эффект переобученности (overfitting) не наблюдается. Достаточно надёжно классифицировались и помеховые ИСЭ, которые соответствовали классу «фоновый», что крайне важно для практического использования системы мониторинга в качестве гарантии минимального уровня ошибок первого рода. В частности, в процессе опытной эксплуатации ϕ -OTDR-системы мониторинга, сконфигурированной для железнодорожных приложений, подсистема обнаружения и классификации, оснащённая классификатором XGBoost при постоянном наблюдении в течение шести месяцев в среднем обеспечивала величину параметра «наработка на ложную тревогу» не хуже 920 часов. В данном приложении целевыми ИСЭ считались: поезда пяти типов, пешеходы, ручные технологические работы на железнодорожном полотне, автоматизированные технологические работы на железнодорожном полотне. К помеховым относились: каналные помехи, работа техники рядом с балластной призмой, автомобильный транспорт (рядом с железнодорожным полотном проходит грунтовая автомобильная дорога), домашние животные и прочие ИСЭ. При этом длина оптоволоконного сенсора составляла 1500 метров, а корректность распознавания системой целевых событий контролировалась оператором с использованием видеокамеры, а также априорно известного графика движения поездного состава. С точки зрения удобства и эффективности практического применения рассмотренные методы классификации на выбранном пространстве признаков примерно равноценны, при этом метод Multi-class SVM

Таблица

Классификатор	Секция 1		Секция 2		Секция 3	
	Тип объекта (7 классов)	$P_{ПК}$	Тип объекта (6 классов)	$P_{ПК}$	Тип объекта (5 классов)	$P_{ПК}$
XGBoost	Пешеход	1,0	Пешеход	1,0	Пешеход	1,0
	Ручная копка грунта	1,0	Ручная копка грунта	1,0	Ручная копка грунта	1,0
	Легковая машина	0,99	Легковая машина	0,99	Колёсная техника	0,99
	Лёгкий колёсный трактор	0,99	Лёгкий колёсный трактор	0,99		
	Электричка	0,99	Поезд	1,0	Поезд	1,0
	Товарный поезд	0,98				
	Фоновый класс	0,99	Фоновый класс	1,0	Фоновый класс	0,99
	Среднее $P_{ПК}$	0,99	—	0,99	—	0,99
Gradient- Boosting	Пешеход	1,0	Пешеход	1,0	Пешеход	1,0
	Ручная копка грунта	1,0	Ручная копка грунта	1,0	Ручная копка грунта	0,99
	Легковая машина	0,97	Легковая машина	0,98	Колёсная техника	0,99
	Лёгкий колёсный трактор	0,97	Лёгкий колёсный трактор	0,98		
	Электричка	0,98	Поезд	1,0	Поезд	1,0
	Товарный поезд	0,98				
	Фоновый класс	1,0	Фоновый класс	1,0	Фоновый класс	0,99
	Среднее $P_{ПК}$	0,98	—	0,99	—	0,99
Multi-class SVM	Пешеход	0,99	Пешеход	1,0	Пешеход	1,0
	Ручная копка грунта	0,99	Ручная копка грунта	1,0	Ручная копка грунта	0,99
	Легковая машина	0,98	Легковая машина	1,0	Колёсная техника	0,99
	Лёгкий колёсный трактор	0,98	Лёгкий колёсный трактор	1,0		
	Электричка	0,97	Поезд	1,0	Поезд	1,0
	Товарный поезд	0,97				
	Фоновый класс	1,0	Фоновый класс	1,0	Фоновый класс	1,0
	Среднее $P_{ПК}$	0,98	—	1,0	—	0,99

обучается быстрее, а ансамблевые методы несколько точнее работают при большем количестве классов. Сравнительные эксперименты проводились на обычном ПК (четырёхъядерный Core i7-2630QM со стандартной частотой 2.0 GHz, RAM 16 Гб). На корпусе данных, описанном в разд. 3.1, и пространстве признаков из разд. 3.2 Multi-class SVM в среднем обучался за 0,5 с, GB за 6 с, XGBoost за 16 с.

Заключение. Практические исследования показали, что предложенный метод автоматической классификации источников сейсмоакустической эмиссии по данным мониторинговой DAS-системы показывает высокую надёжность и легко реализуем при современном уровне вычислительных средств. Выбранное оригинальное множество классификационных признаков достаточно информативно для обеспечения собственно классификационного процесса. Наилучшие результаты показали классификаторы следующих типов: Multi-class SVM, XGBoost и Gradient-Boosting, которые равномерно по всем классам обеспечивали величину показателя $P_{ПК} \geq 0,98$ для случаев семи, шести и пяти классов целей. Достигнутый равномерно по всем классам целей уровень надёжности классификации в настоящее время недоступен для альтернативных методов классификации, используемых в ϕ -OTDR-системах [22, 23], где этот показатель для некоторых классов не превышает величину 0,9. По результатам данного исследования была разработана подсистема классификации сейсмоакустических событий, входящая в ϕ -OTDR-систему мониторинга «ОКСИ» (ОХУ). Данная система мониторинга включена в ряд проектов по обеспечению контроля протяжённых объектов на территории Республики Казахстан.

СПИСОК ЛИТЕРАТУРЫ

1. **Pat. GB 2222247 A.** Distributed fibre optic sensor system /J. P. Dakin, C. Lamb. Publ. 28.02.1990.
2. **Choi K. N., Juarez J. C., Taylor H. F.** Distributed fiber optic pressure/seismic sensor for low-cost monitoring of long perimeters // Proc. SPIE. 2003. **5090**. P. 134–141.
3. **Timofeev A. V., Denisov V. M.** Multimodal heterogeneous monitoring of super-extended objects: Modern view // Recent Advances in Systems Safety and Security. Switzerland: Springer-Verlag. 2016. Vol. 62. P. 97–116. DOI: 10.1007/978-3-319-32525-5_6.
4. **Timofeev A. V., Egorov D. V.** Bimodal system for surveillance of super-extended objects // Proc. of the 6th Intern. Conf. on Electronics, Computers and Artificial Intelligence (ECAI). Bucharest, Romania, 23–25 Oct., 2014. P. 57–62. DOI: 10.1109/ECAI.2014.7090211.
5. **Timofeev A. V., Egorov D. V., Denisov V. M.** The rail traffic management with usage of C-OTDR monitoring systems // World Acad. Sci. Eng. Technol. Intern. Journ. Comput. Electr. Autom. Control Inf. Eng. 2015. **9**, N 7. P. 1492–1495.
6. **Timofeev A. V.** Monitoring the railways by means of C-OTDR technology // Intern. Journ. Mech. Aerosp. Ind. Mechatron. Eng. 2015. **9**, N 5. P. 701–704.
7. **Pattern Recognition.** GCPR 2016 /Eds. B. Rosenhahn, B. Andres. Lecture Notes in Computer Science. Vol. 9796. Hannover: Springer, 2016. Train Detection and Tracking in Optical Time Domain Reflectometry (OTDR) Signals /A. Papp, C. Wiesmeyr, M. Litzenberger, H. Garn, W. Kropatsch. P. 320–331.
8. **Korotaev V. V., Denisov V. M., Rodrigues J. J. P. C. et al.** Monitoring industrial facilities using principles of integration of fiber classifier and local sensor networks // Proc. SPIE. 2015. **9525**. 95254J. DOI: 10.1117/12.2184727.
9. **Barrias A., Casas J. R., Villalba S.** A review of distributed optical fiber sensors for civil engineering application // Sensors. 2016. **16**, N 5. P. 56–78.
10. **Takada K., Himeno A., Yukimatsu K.** Phase-noise and shot-noise limited operations of low coherence optical time domain reflectometry // Appl. Phys. Lett. 1991. **59**, Iss. 20. P. 2483–2485.

11. **Rogers A. J.** Polarisation optical time domain reflectometry // *Electron. Lett.* 1980. **16**, Iss. 13. P. 489–490.
12. **Juarez J. C., Taylor H. F.** Polarization discrimination in a phase-sensitive optical time-domain reflectometer intrusion-sensor system // *Opt. Lett.* 2005. **30**, Iss. 24. P. 3284–3286.
13. **Cui Q., Pamukcu S., Xiao W., Pervizpour M.** Truly distributed fiber vibration sensor using pulse base BOTDA with wide dynamic range // *IEEE Photon. Technol. Lett.* 2011. **23**, Iss. 24. P. 1887–1889.
14. **Wang Y., Jin B., Wang Y. et al.** Real-time distributed vibration monitoring system using Φ -OTDR // *IEEE Sensors Journ.* 2017. **17**, N 5. P. 1333–1341.
15. **Liu X., Jin B., Bai Q. et al.** Distributed fiber-optic sensors for vibration detection // *Sensors.* 2016. **16**, Iss. 8. P. 1164–1195. DOI: 10.3390/s16081164.
16. **Марченко К. В., Наний О. Е., Нестеров Е. Т. и др.** Защита ВОЛС распределённым акустическим датчиком на основе когерентного рефлектометра // *Вестн. связи.* 2011. № 9. С. 17–19.
17. **Русанов Ю.** К вопросу о перспективах использования волоконно-оптических средств для охраны периметра // *Системы безопасности.* 2018. № 4. С. 29–32.
18. **Brooks R. R., Ramanathan P., Sayeed A. M.** Distributed target classification and tracking in sensor networks // *Proc. of the IEEE.* 2003. **91**, Iss. 8. P. 1163–1171. DOI: 10.1109/JPROC.2003.814923.
19. **Shen X., Wan Sh. C., Huo H., Fang T.** An improvement on discrete wavelet transform-based algorithm for vehicle classification in wireless sensor networks // *Proc. of the 1st IEEE Conf. on Industrial Electronics and Applications.* Singapore, Singapore, 24–26 May, 2006. P. 1–4. DOI: 10.1109/ICIEA.2006.257218.
20. **Wang X., Qi H., Iyengar S. S.** Collaborative multi-modality target classification in distributed sensor networks // *Proc. of the Intern. Conf. on Information Fusion 2002.* Annapolis, USA, 8–11 July, 2002. P. 285–290.
21. **Kim Y., Jeong S., Kim D.** A GMM-based target classification scheme for a node in wireless sensor networks // *IEICE Trans. Commun.* 2008. E91B. P. 3544–3551.
22. **Aktas M., Akgun T., Demircin M. U., Buyukaydin D.** Deep learning based multi-threat classification for phase-OTDR fiber optic distributed acoustic sensing applications // *Proc. SPIE.* 2017. **10208**. P. 56–70.
23. **Makarenko A. V.** Deep learning algorithms for signal recognition in long perimeter monitoring distributed fiber optic sensors // *Proc. of the IEEE 26th Intern. Workshop on Machine Learning for Signal Processing (MLSP).* Vietri sul Mare, Italy, 13–16 Sept., 2016. P. 1–6. DOI: 10.1109/MLSP.2016.7738863.
24. **Li J., Zhang C., Li Zh.** Battlefield target identification based on improved grid-search SVM classifier // *Proc. of the Intern. Conf. on Computational Intelligence and Software Engineering.* Wuhan, China, 11–13 Dec., 2009. P. 1–4.
25. **Kandpal M., Kakar V. K., Verma G.** Classification of ground vehicles using acoustic signal processing and neural network classifier // *Proc. of the Intern. Conf. on Signal Processing and Communication (ICSC).* Noida, India, 12–14 Dec., 2013. P. 512–518. DOI: 10.1109/ICSPCom.2013.6719846.
26. **Алямкин С. А., Еременко С. И.** Алгоритм распознавания пешехода на основе анализа автокорреляционной функции сейсмического сигнала // *Автометрия.* 2011. **47**, № 2. С. 26–32.
27. **Соколова Д. О., Спектор А. А.** Классификация движущихся объектов по спектральным признакам сейсмических сигналов // *Автометрия.* 2012. **48**, № 5. С. 112–119.

-
28. **Морозов Ю. В., Спектор А. А.** Классификация объектов на основе анализа спектральных характеристик огибающих сейсмических сигналов // Автометрия. 2017. **53**, № 6. С. 49–56.
 29. **Морозов Ю. В., Райфельд М. А., Спектор А. А.** Обработка сейсмических сигналов для оценки траектории движущегося транспортного средства // Автометрия. 2018. **54**, № 3. С. 32–38.
 30. **Chen T., Guestrin C.** XGBoost: A scalable tree boosting system // Proc. of the 22nd ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining. San Francisco, USA, 13–17 Aug., 2016. P. 785–794. DOI: 10.1145/2939672.2939785.
 31. **Mason L., Baxter J., Bartlett P., Frean M.** Boosting algorithms as gradient descent // Proc. of the 12th Intern. Conf. on Neural Information Processing Systems. Denver, USA, 29 Nov. – 4 Dec., 1999. P. 512–518.
 32. **Friedman J. H.** Greedy function approximation: A gradient boosting machine // The Ann. Statist. 2001. **29**, N 5. P. 1189–1232.

Поступила в редакцию 22.05.2019

После доработки 22.07.2019

Принята к публикации 12.11.2019
