ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2023, том 514, № 2, с. 39–48

УДК 004.8

НЕЙРОСЕТЕВОЙ ПОДХОД К КЛАССИФИКАЦИИ АКУСТИЧЕСКИХ СИГНАЛОВ С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИОННЫХ ПРИЗНАКОВ

© 2023 г. П. В. Лысенко[1,*], И. А. Насонов[1,],**
Член-корреспондент РАН А. А. Галяев[1,*], Л. М. Берлин[1,****]**

Представлено академиком РАН А.Л. Семеновым
Поступило 22.08.2023 г.
После доработки 30.08.2023 г.
Принято к публикации 10.09.2023 г.

В работе рассмотрена задача бинарной классификации акустических сигналов биологического происхождения, записанных в естественных условиях. В качестве признакового описания объектов выбраны информационные характеристики, такие как энтропия и статистическая сложность. Методы решения основаны на трех архитектурах нейронных сетей, модифицированных авторами (на ядре Inception, на ядре Inception и технологии Residual, на структуре Self-Attention с блоками LSTM). Использован датасет из соревнования на Kaggle по обнаружению акустических сигнатур китов, и проведено сравнение между моделями по качеству решения рассматриваемой задачи на стандартном наборе метрик. Получено значение AUC ROC более 90%, что говорит об успешном решении задачи обнаружения полезного сигнала и указывает на возможную применимость информационных характеристик к подобным задачам.

Ключевые слова: классификация временных рядов, спектрограмма, статистическая сложность, глубокое обучение

DOI: 10.31857/S2686954323601239, **EDN:** HHQYXE

1. ВВЕДЕНИЕ

Задачи распознавания образов являются важным для практики подразделом теории математической статистики и теории вероятностей [1]. В машинном обучении принято говорить о задаче классификации, когда в результате наблюдения объекта из множества объектов, принадлежащих разным классам, требуется определить, какому классу принадлежит наблюдаемый объект. Рассматриваемая в статье задача детектирования полезного акустического сигнала может рассматриваться как задача бинарной классификации временных рядов, где один класс отвечает случаю наличия полезного сигнала в выборке, а второй класс – его отсутствию.

В задачах глубокого машинного обучения, связанных с анализом временных рядов, часто применяются рекуррентные нейронные сети, в частности сети с долгой краткосрочной памятью LSTM (long-short-term-memory) как для предсказания следующего элемента ряда, так и для классификации [2]. Отдельное место в научной области анализа временных рядов занимают задачи, связанные с исследованием медицинских показателей, таких как ЭКГ и ЭЭГ последовательностей [3]. Чаще всего требуется распознавать аномалии или классифицировать данные, что может помочь в более раннем выявлении симптомов болезни и диагносцировании пациента. Упомянутые выше сети на LSTM демонстрируют высокие показатели в подобных задачах [4, 5].

В последние годы сверточные нейросети стали популярным техническим решением, демонстрирующим лидирующие результаты по выделению признаков в задачах классификации объектов на изображениях [6]. В частности, применительно к аудиосигналам популярным является подход, в котором обучение сверточных нейросетей производится на спектрограммах сигналов [7]. Такой подход показывает высокую эффективность, однако связан с использованием большого объема затрачиваемой памяти и значительным временем обучения. На фоне общей популярности сверточных нейросетей, некоторые исследователи так же применяют их в задачах глубокого обучения, связанных с классификацией временных рядов [8, 9].

Архитектуры нейросетей, основанные на механизме self-attention [10], наиболее известным из

[1]*Институт проблем управления им. В.А. Трапезникова Российской академии наук, Москва, Россия*
**E-mail: pavellysen@ipu.ru*
***E-mail: nasonov.ia18@physics.msu.ru*
****E-mail: galaev@ipu.ru*
*****E-mail: berlin.lm@phystech.edu*

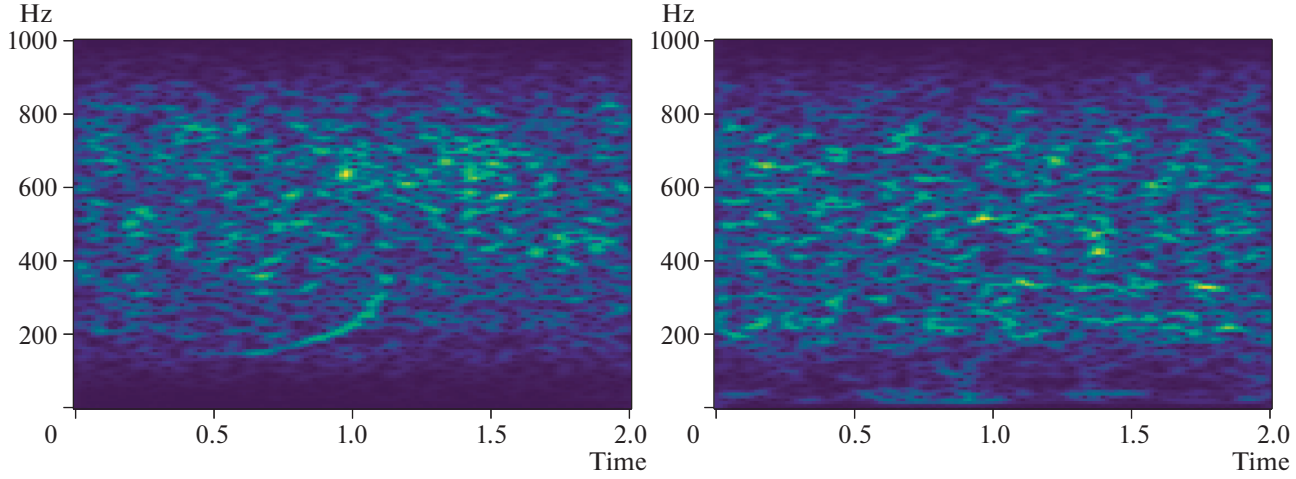

Рис. 1. Примеры записей для класса 1 (слева) и класса 0 (справа).

которых является семейство transformer (BERT, ALBERT, BARD, GPT и др.), вызывают большой интерес исследователей в силу качества результатов, которые они позволяют достичь в задачах обработки текста. При этом возможно совмещение подходов, в частности использование self-attention архитектур совместно с блоками LSTM [11]. Подобная архитектура впервые применена к рассматриваемой задаче классификации временных рядов, что является нетипичным решением, и оказалась перспективной для улучшения результатов обработки данных.

Важным подразделом обработки аудиосигналов являются задачи обработки гидроакустических сигналов, которые в силу физики распространения сигнала в среде и наличию большого количества внешних шумов зачастую являются намного более сложными для решения [12]. В частности, рассмотренная в [13] задача обнаружения аудиосигнатур китов представляет особый интерес для исследователей-биологов, изучающих фауну мирового океана. Обычно в таких задачах невозможно использование сырых данных временных рядов для обучения, и ключевым этапом становится конструирование признаков, на основе которых и будут проводиться анализ и классификация объектов. Обзор таких признаков для задачи классификации аудиосигналов приведен в [14]. Одним из классических признаков является спектральная энтропия или ее аналоги, однако в [15, 16] было показано, что статистическая сложность обладает лучшей способностью по выделению сигнала, а в [17] продемонстрировано использование признаков энтропии/сложности для классификации морских объектов по их гидроакустическому шуму.

Поскольку информационная энтропия и сложность являются перспективными информационными характеристиками временных рядов [18], которые позволяют детектировать появление детерминированных сигналов на фоне сильного шума [15], то эти величины применены в данной работе для решения задачи обнаружения полезных акустических сигналов [19] в естественных шумах методами глубокого обучения.

2. ОПИСАНИЕ ДАТАСЕТА

Использован датасет из соревнования на Kaggle [19] по детектированию акустических сигнатур китов. Датасет состоит из аудиозаписей длительностью $T = 2$ с в формате .aiff с частотой дискретизации $f_s = 2$ кГц. Аудиозаписи делятся на два класса: класс 1 отвечает присутствию в записи аудиосигнатуры кита; класс 0 – отсутствию таковой. При этом во всех записях присутствует шум моря разной интенсивности. На рис. 1 показаны примеры спектрограмм аудиозаписей, соответствующих разным классам. Стоит отметить, что во многих записях присутствуют низкочастотные паразитные шумы, связанные с несовершенством записывающих гидрофонов, поэтому была произведена фильтрация частот ниже 50 Гц.

Набор данных содержит 30 000 таких записей, однако имеет сильный дисбаланс – более 75% примеров относятся к классу 0. Распределение данных по классам изображено на рис. 2. Для достижения более сбалансированного представления классов в датасете была применена техника аугментации данных с использованием стохастического процесса Орнштейна-Уленбека, который является математическим методом моделирования стохастических процессов с длительными корреляциями. Этот процесс создает реалистичные случайные вариации данных, сохраняя при этом статистические свойства и структуру исход-

ных временных рядов. С использованием стохастического процесса Орнштейна-Уленбека было произведено увеличение количества данных класса 1 до 47.5% от всего набора данных. Такое соотношение классов было выбрано для достижения баланса между классами и предотвращения смещения в представлении модели.

Для дополнительного обеспечения сбалансированного распределения классов в обучающей и тестовой выборках данные случайным образом перемешиваются перед разделением, чтобы исключить возможное смещение при разбиении. Кроме того, при разделении датасета на обучающую, тестовую и валидационную выборки пропорции классов в них сохраняются такими же, как и целиком в аугментированных данных. Таким образом, гарантируется, что обе выборки будут содержать соответствующие доли примеров с метками 0 и 1, что способствует более устойчивой и справедливой оценке модели и предотвращает возможные проблемы с переобучением. Выборка разбивается на train, validation и test в отношении 0.8, 0.1 и 0.1 соответственно.

3. ФОРМИРОВАНИЕ ИНФОРМАЦИОННЫХ ПРИЗНАКОВ

В качестве признаков, формирующих временные ряды для распознавания акустических сигналов биологического происхождения, предлагается использовать такие информационные характеристики, как спектральные энтропия и сложность [17]. В работах [15, 16] было показано, что они позволяют выделять детерминированный сигнал в шумовой смеси в задаче различения двух гипотез. При этом разные варианты статистической сложности C_{SQ}, C_{JSD}, C_{TV} достигают своих максимальных значений на разных дискретных распределениях, т.е. лучше всего подходят под детектирование сигналов с разной спектральной структурой [16], что позволяет использовать их совместно с энтропией в качестве признаков в задаче классификации.

Метод получения таких рядов подробно описан в статье [15]. Работа алгоритма, вычисляющего информационные характеристики, осуществляется следующим образом:

1. Синтезированный с частотой дискретизации f_s сигнал длительности T делится на короткие окна, содержащие по $N = 256$ отсчетов.

2. Далее для каждого окна вычисляется спектральная плотность мощности $s(f_i)$ с использованием алгоритма БПФ:

$$s(f_i) = \frac{1}{N}|X(f_i)|^2, \qquad (1)$$

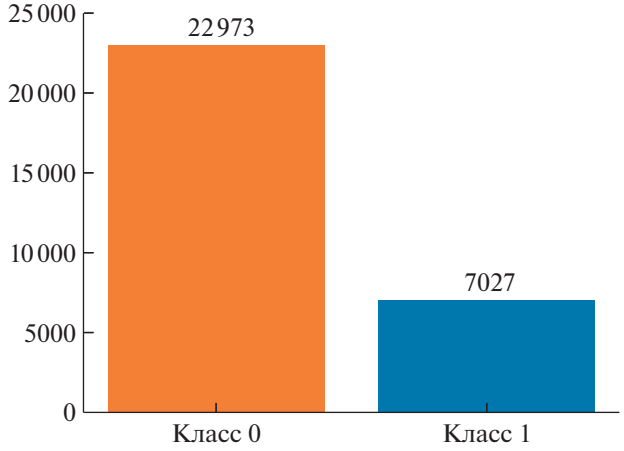

Рис. 2. Распределение данных в датасете.

где $X(f_i)$ – амплитуда спектральной компоненты с частотой f_i.

3. На основе нормированного спектра вычисляются дискретные плотности $p = (p_1, \ldots, p_i, \ldots, p_N)$:

$$p_i = \frac{s(f_i)}{\sum\limits_{k=1}^{N} s(f_k)}, \quad i = 1, \ldots, N. \qquad (2)$$

Полученное дискретное распределение $p = (p_1, \ldots, p_i, \ldots, p_N)$ положительно, нормировано и обладает свойствами ряда распределения дискретной случайной величины

$$\forall p_i \in [0,1], \quad \sum_{i=1}^{N} p_i = 1. \qquad (3)$$

4. Далее для полученного распределения p_i вычисляются информационные характеристики: энтропия Шеннона $H(p)$, статистическая сложность на основе евклидова расстояния $C_{SQ}(p)$, статистическая сложность на основе дивергенции Йенсена-Шеннона $C_{JSD}(p)$, статистическая сложность на основе полной вариации меры со знаком $C_{TV}(p)$ по следующим формулам

$$H(p) = \frac{1}{\log N}\left(-\sum_{i}^{N} p_i \log p_i\right),$$

$$C_{SQ}(p) = \frac{1}{\log N}\left(-\sum_{i=1}^{N} p_i \log p_i\right) \cdot \left(\sum_{i=1}^{N} (p_i - q_i)^2\right), \qquad (4)$$

$$C_{JSD}(p) = \frac{1}{\log N}\left(-\sum_{i=1}^{N} p_i \log p_i\right) \cdot JSD(p,q),$$

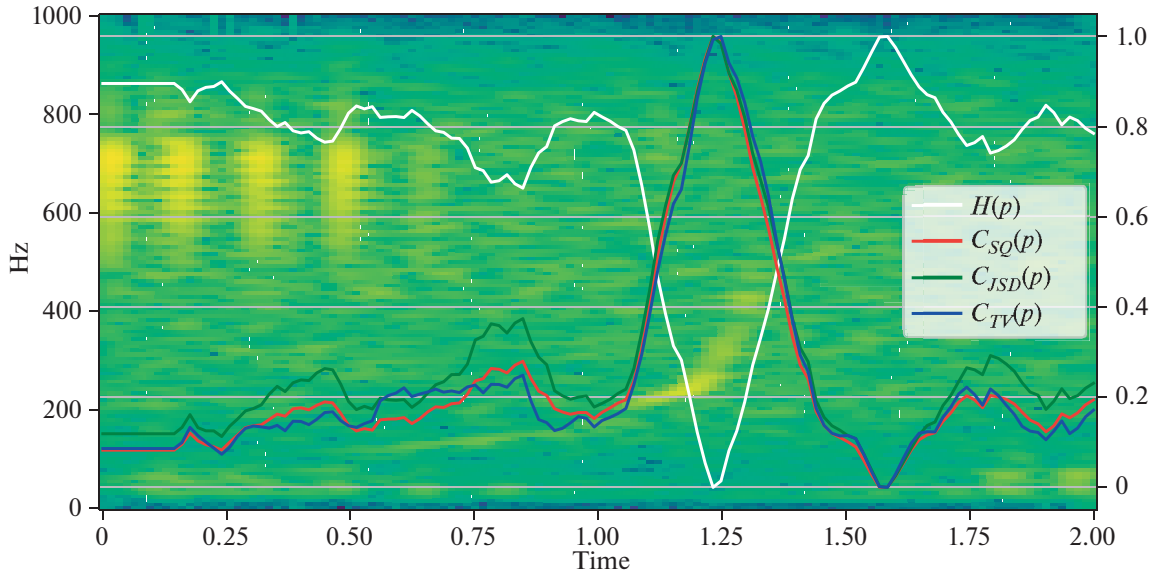

Рис. 3. Временные ряды, изображенные вместе со спектрограммой сигнала из класса 1.

$$C_{TV}(p) = -\frac{1}{4\log N}\left(\sum_{i=1}^{N} p_i \log p_i\right)\cdot\left(\sum_{i=1}^{N}|p_i - q_i|\right)^2,$$

где $q_i = \frac{1}{N}, i = 1,\ldots,N$ – равномерное дискретное распределение, соответствующее спектральному распределению белого шума; $JSD(p,q)$ – дивергенция Йенсена-Шеннона.

5. Полученные для каждого временного окна последовательности значений информационных характеристик сглаживаются методом скользящего среднего для фильтрации шума и формируют временные ряды, которые сохраняются для каждого объекта из датасета.

Таким образом, для каждой записи из датасета вычисляются векторы сложности и энтропии одинаковой длины, которые затем подаются на вход нейронной сети – классификатора. На рис. 3 показан пример спектрограммы аудиосигнала из класса 1 и временных рядов, рассчитанных по вышеприведенному алгоритму. Для удобства презентации ряды нормированы относительно своих минимальных и максимальных значений так, чтобы область их значений лежала в диапазоне [0, 1]. Как видно из графиков, информационные характеристики позволяют выделять появление детерминированного сигнала в шумовой смеси, а форма кривых и превышение пороговых значений могут в теории служить признаками для классификации таких сигналов.

4. АРХИТЕКТУРЫ НЕЙРОННЫХ СЕТЕЙ

Для решения задачи классификации временных рядов информационных характеристик были использованы три нейронных сети, основанные на разных типах архитектур:

1. На ядре Inception [8];

2. На ядре Inception и технологии Residual [6];

3. На структуре Self-Attention с блоками LSTM [10].

Нейронная сеть InceptionTime основана на идее применения модулей Inception для анализа данных временных рядов. Каждый такой модуль содержит три ветви свертки с разными размерами ядер в параллельных ветвях, после чего их выводы объединяются для получения общего признакового представления. Схема сети изображена на рис. 4.

В каждом модуле Inception применяются одномерные свертки с ядрами 1, 2 и 4 для анализа информации на разных уровнях детализации временного ряда. После каждой свертки применяется функция активации ReLU и результат проходит через слой Dropout, который с некоторой вероятностью обнуляет веса сети, что предотвращает переобучение. После обработки данных в каждой ветви, результаты всех ветвей объединяются путем конкатенации в единую матрицу, которая в свою очередь подается в слой MaxPooling для уменьшения размерности и извлечения наиболее значимых признаков из входных данных. Именно способность сети обобщать различные уровни информации из временных рядов позволяет ей более качественно обучаться на таких данных. Завершается модель двумя полносвязными

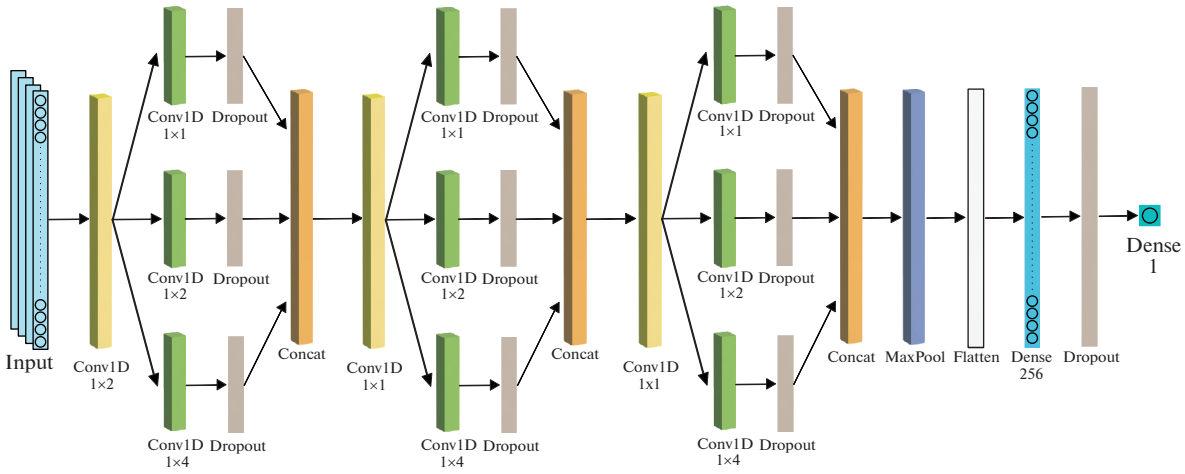

Рис. 4. Архитектура сети InceptionTime.

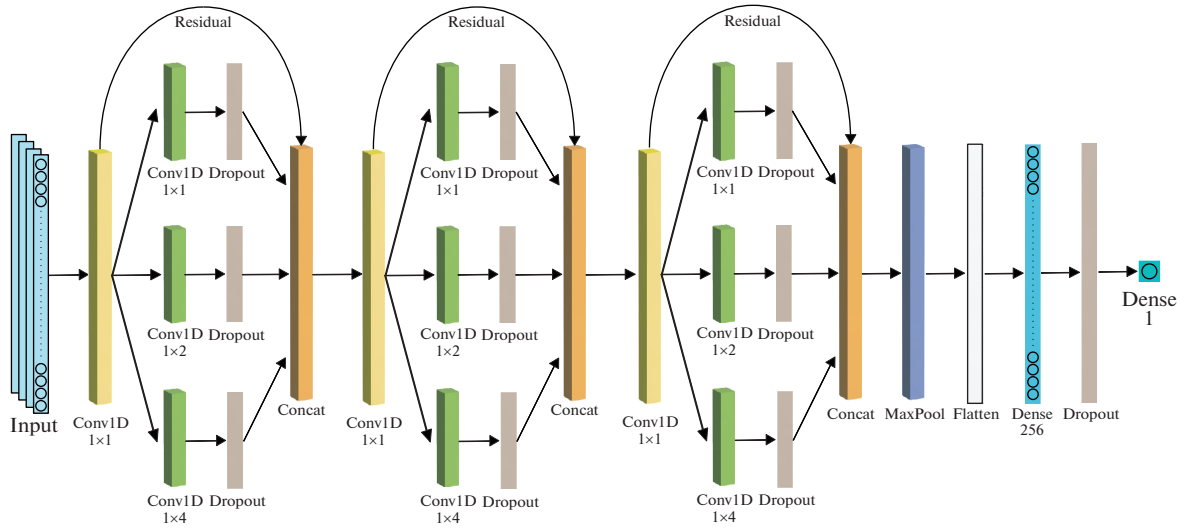

Рис. 5. Архитектура сети InceptionTime с применением блоков Residual.

слоями с применением слоя Dropout. Схема модели сети InceptionTime приведена на рис. 4.

Также была разработана сеть, основанная на InceptionTime, но модифицированная технологией Residual. Данный механизм введен в блоке Inception, в котором входной вектор отдельным каналом передается в конечный слой, добавляясь к данным, прошедшим обработку через слои одномерной свертки. Технология Residual эффективно борется с проблемой затухания градиентов и ускоряет процесс обучения. Главной целью этой модификации является оценка влияния Residual на способность сети классифицировать временные ряды. Схема сети InceptionTime с технологией Residual приведена на рис. 5.

Также была разработана третья нейронная сеть, которая представляет собой модель, состоящую из рекуррентного слоя LSTM (Long Short-Term Memory), слоя Multi-Head Attention и нескольких полносвязных слоев. Рекуррентный слой LSTM используется для анализа последовательности данных во времени. Этот слой применяется для определения долгосрочных зависимостей в рядах информационных характеристик.

Основным слоем для отбора признаков в данной сети является Multi-Head Attention, который позволяет модели учитывать различные части входных данных для выделения важных элементов и установления взаимосвязи между ними. Прежде чем применить механизм Self-Attention, входные данные проходят через три линейных преобразования весовых матриц (W_Q, W_K, W_V) (которые настраиваются как параметры модели во время обучения сети), необходимых для созда-

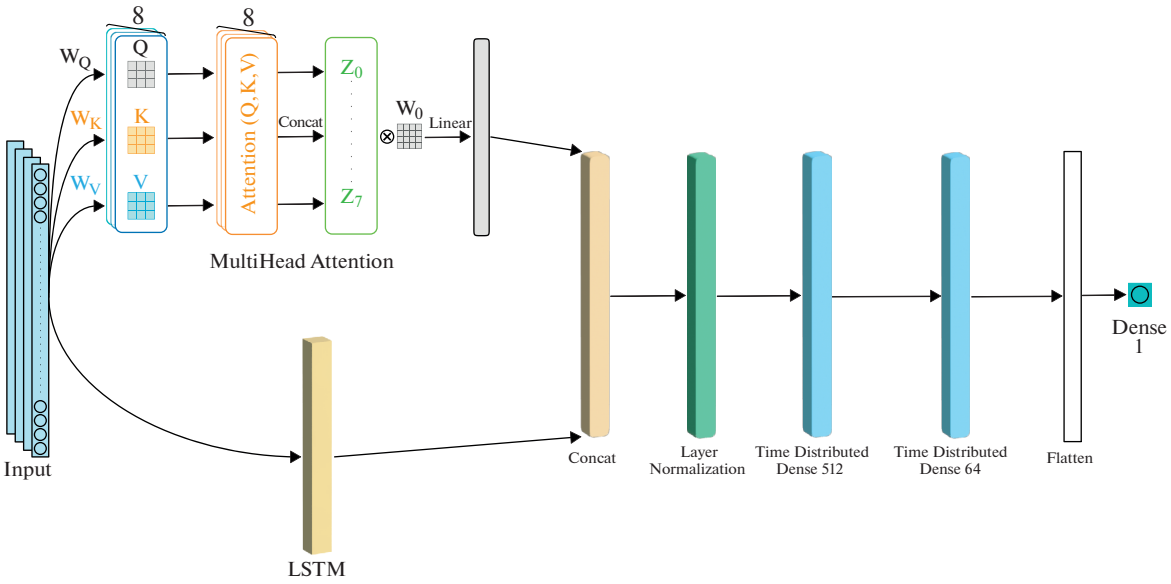

Рис. 6. Архитектура сети на основе технологии Self-Attention.

ния матриц (Q, K, V) для каждого элемента последовательности, чтобы в дальнейшем представить их в пространстве с меньшей размерностью. После этого происходит оценка важности каждого элемента относительно других путем скалярного произведения весов с помощью следующей формулы:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

где d_k – задает размерность пространства K и Q, и может варьироваться в зависимости от конкретной архитектуры и задачи. Деление в формуле выше на $\sqrt{d_k} = 4$ выполняется для нормализации величины весов внимания и управления масштабом значений. Основным свойством данного слоя является использование множественных преобразований для выделения различных аспектов данных. Для получения общего признакового представления данных, выходы каждого преобразования объединяются в единый вектор $Z_{0-7} \otimes W_0$, в котором учтены взаимосвязи между элементами последовательности и выделены наиболее важные характеристики.

В сети данные подаются в блоки LSTM и Multi-Head Attention параллельно, а в дальнейшем их выходы объединяются в единый слой, из которого они проходят через слой нормализации для улучшения стабильности, ускорения сходимости обучения и решения проблемы затухания градиентов. Далее данные подаются в полносвязные TimeDistributed слои, которые обрабатывают данные в каждом временном шаге независимо друг от друга. Схема модели сети на основе архитектуры Self-Attention изображена на рис. 6.

5. ОБУЧЕНИЕ НЕЙРОННЫХ СЕТЕЙ

Для обучения всех моделей использовался оптимизатор Adam с функцией потерь Binary Crossentropy и метриками AUC, Accuracy, F1-score, Recall и Precision для оценки качества классификации.

Во время тренировки сетей значение скорости обучения (Learning Rate) динамически определялось в зависимости от текущей эпохи обучения, начального значения скорости обучения, минимального и максимального значения Learning Rate, а также от параметров управления скоростью обучения, таких как периоды увеличения, удержания и затухания скорости обучения. В начале обучения сети используется повышенное значение Learning Rate для быстрого приближения к оптимальной конфигурации весов. Это помогает избежать застревания в локальных минимумах многомерной плоскости функции потерь и быстро находить верное направление ее минимизации. Затем это значение уменьшается, чтобы стабилизировать процесс обучения и точнее настроить веса для лучшей сходимости к глобальному минимуму. Такой двухэтапный подход позволяет сбалансировать скорость и точность обучения, обеспечивая эффективную и стабильную тренировку сети. Формула расчета коэффициента обучения приведена в выражении ниже.

Таблица 1. Таблица параметров, определяющих зависимость Learning Rate от номера эпохи

Параметр	Значение	Определение
S	10^{-3}	Начальное значение скорости обучения
$LR_{\min}$	10^{-5}	Минимальное значение скорости обучения
$LR_{\max}$	5×10^{-3}	Максимальное значение скорости обучения
I	5	Количество эпох для линейного увеличения скорости обучения с S до $LR_{\max}$
H	0	Количество эпох для удержания максимальной скорости обучения $LR_{\max}$
E	0.85	Коэффициент экспоненциального затухания скорости обучения после периода увеличения и удержания скорости

Таблица 2. Значения метрик качества обучения нейронных сетей

Сеть	AUC ROC	Accuracy	F1-score	Recall	Precision
InceptionTime	0.906	0.811	0.776	0.720	0.841
InceptionTime Residual	**0.909**	**0.814**	**0.788**	**0.763**	**0.813**
Self-Attention	0.894	0.801	0.776	0.762	0.790

$$LearningRate(epoch) = \begin{cases} S + \dfrac{(LR_{\max} - S)\times epoch}{I}, \\ epoch < I \\ LR_{\max}, \quad I \le epoch < I + H \\ (LR_{\max} - LR_{\min})\times E^{(epoch-I-H)} + LR_{\min}. \\ epoch \ge I + H \end{cases} \tag{5}$$

В табл. 1 приведены описания и значения параметров из (5).

Все программные процедуры для предобработки данных и обучения разработаны на языке Python. Модели были реализованы в библиотеке Keras с использованием бэкенда TensorFlow. Обучение производилось на компьютере с процессором Intel i7 и графическим ускорителем Nvidia Geforce RTX 3060.

Графики, иллюстрирующие изменение основных параметров обучения сетей, изображены на рис. 7, 8 и 9.

6. РЕЗУЛЬТАТЫ

В результате обучения нейронных сетей InceptionTime, InceptionTime Residual и сети, основанной на архитектуре Self-Attention с блоками LSTM, для решения задачи бинарной классификации временных рядов информационных признаков были получены следующие значения метрик на тестовом наборе данных, приведенные в табл. 2.

Основным показателем качества решения задачи обнаружения сигнала является ROC-кривая (рабочая характеристика приемника) и площадь под ней AUC ROC. На рис. 10 приведены матрица ошибок и ROC-кривая лучшего обученного классификатора InceptionTime Residual.

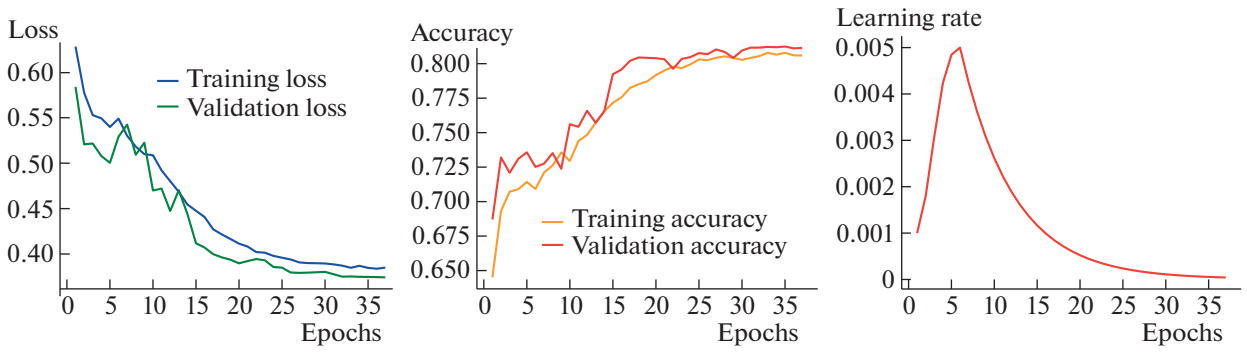

Рис. 7. Графики ошибки, точности и шага обучения на этапе тренировки сети InceptionTime.

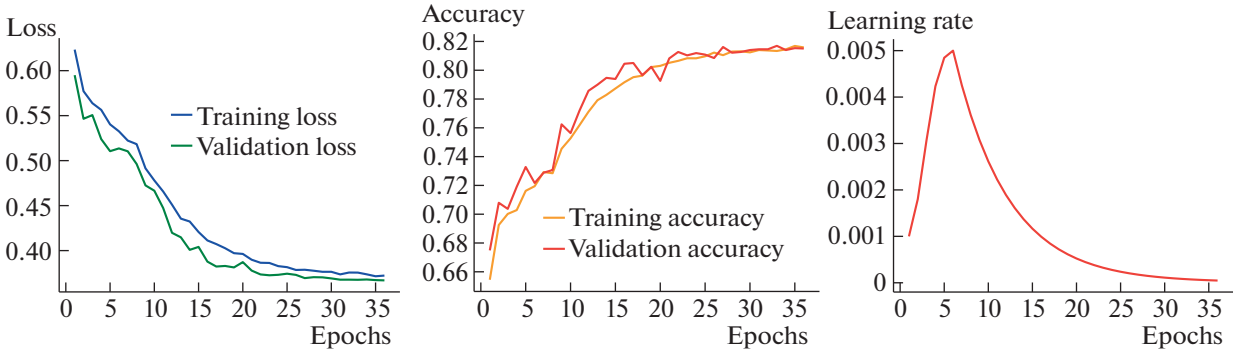

Рис. 8. Графики ошибки, точности и шага обучения на этапе тренировки сети InceptionTime Residual.

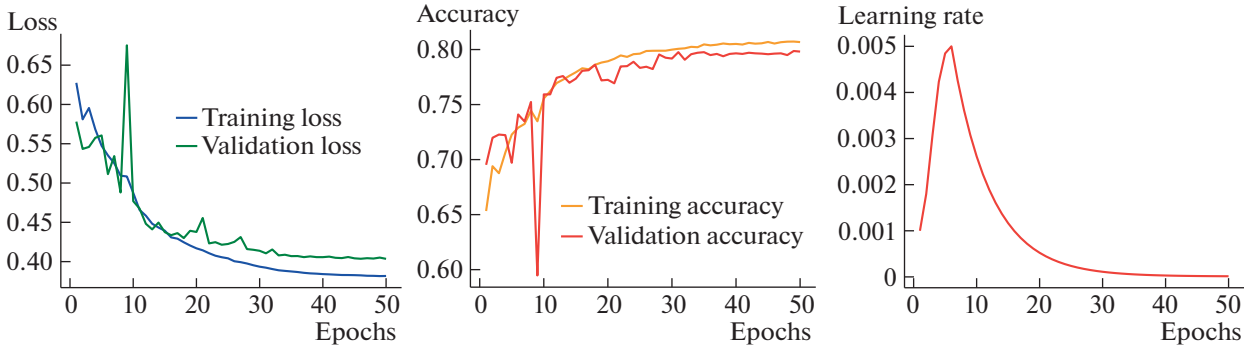

Рис. 9. Графики ошибки, точности и шага обучения на этапе тренировки сети Self-Attention.

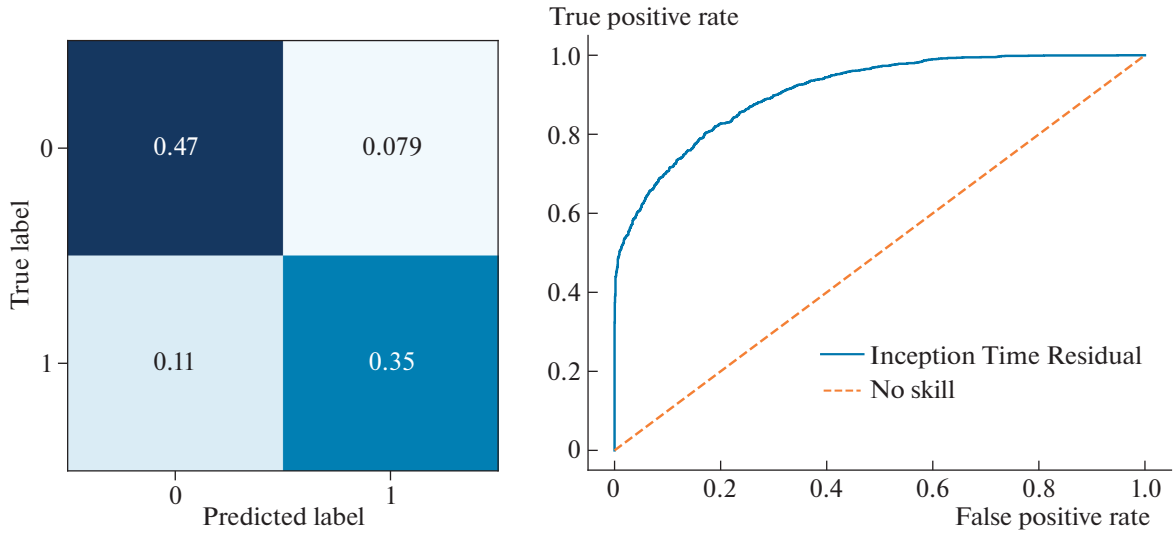

Рис. 10. Матрица ошибок и ROC-кривая для обученной сети InceptionTime Residual.

Анализируя полученные результаты, можно отметить, что все сети показали схожую высокую производительность с близкими значениями AUC ROC. Интересно, что сеть на основе Self-Attention достигает показателей основных метрик качества, сравнимых с сетью InceptionTime, однако имеет лучшие метрики Recall и Precision, что свидетельствует о хорошей способности такой архитектуры классифицировать временные ряды. Сеть InceptionTime с технологией Residual зани-

мает первое место по всем метрикам в табл. 2, что показывает высокую эффективность этого метода.

Таким образом, результаты эксперимента свидетельствуют о том, что все нейронные сети достигли высокого качества решения задачи на тестовых данных и на основе полученного решения может быть создан эффективный детектор полезных сигналов. Кроме того видно, что аугментация данных позволяет успешно решать проблему дисбаланса классов. Однако для дальнейшего улучшения результатов исследования требуется больше реальных данных положительного класса и аналитическая фильтрация данных от реальных внешних помех.

7. ЗАКЛЮЧЕНИЕ

В работе получен набор решений задачи бинарной классификации акустических сигналов биологической природы, применены три архитектуры нейросетей и произведено сравнение метрик качества между ними. Архитектура на self-attention с блоками LSTM применена к подобной задаче впервые и показывает сравнимый с другими архитектурами результат. Получены высокие показатели метрик качества, в частности AUC ROC, что говорит об успешном решении задачи обнаружения полезного сигнала. Показана применимость информационных характеристик энтропии и сложности в качестве признаков для задачи классификации акустических сигналов.

Будущая работа будет направлена на исследование задачи с помощью не только спектральных, но и временных, информационных характеристик. Поскольку перспективным представляется создание более сложного мультимодального классификатора, учитывающего как другие известные признаки в виде временных рядов, так и трехмерные представления в виде спектрограммы и ее производных. Также планируется применить приведенный в работе метод для временных рядов принципиально другой природы, например, данных сейсмической активности или медико-биологических показателей (ЭКГ, ЭЭГ и др.).

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследование выполнено за счет гранта Российского научного фонда 23-19-00134.

СПИСОК ЛИТЕРАТУРЫ

1. *Bishop Christopher M.* Pattern Recognition and Machine Learning (Information Science and Statistics) / Christopher M. Bishop. Berlin, Heidelberg: Springer-Verlag, 2006. https://link.springer.com/book/9780387310732.
2. LSTM Fully Convolutional Networks for Time Series Classification / Fazle Karim, Somshubra Majumdar, Houshang Darabi, Shun Chen // IEEE Access. 2018. V. 6. P. 1662–1669. http://ieeexplore.ieee.org/document/8141873/.
3. Deep Learning-Based ECG Arrhythmia Classiшacation: A Systematic Review / Qiao Xiao, Khuan Lee, Siti Aisah Mokhtar et al. // Applied Sciences. 2023. apr. V. 13. № 8. P. 4964. https://www.mdpi.com/2076-3417/13/8/4964.
4. *Pham Tuan D.* Time–frequency time–space LSTM for robust classification of physiological signals / Tuan D. Pham // Scientific Reports. 2021. mar. V. 11. № 1. P. 6936. https://www.nature.com/articles/s41598-021-86432-7.
5. The Use of Time-Frequency Moments as Inputs of LSTM Network for ECG Signal Classification / Grzegorz Kłosowski, Tomasz Rymarczyk, Dariusz Wójcik et al. // Electronics. 2020. sep. V. 9. № 9. P. 1452. https://www.mdpi.com/2079-9292/9/9/1452.
6. Deep residual learning for image recognition / Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 770–778.
7. Performance of a deep neural network at detecting North Atlantic right whale upcallsa) / Oliver S. Kirsebom, Fabio Frazao, Yvan Simard et al. // The Journal of the Acoustical Society of America. 2020. 04. V. 147. № 4. P. 2636–2646. https://doi.org/10.1121/10.0001132
8. InceptionTime: Finding AlexNet for time series classification / Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier et al. // Data Mining and Knowledge Discovery. 2020. V. 34. № 6. P. 1936–1962. https://link.springer.com/10.1007/s10618-020-00710-y.
9. Convolutional neural networks for time series classification / Bendong Zhao, Huanzhang Lu, Shangfeng Chen et al. // Journal of Systems Engineering and Electronics. 2017. feb. V. 28. № 1. P. 162–169. http://ieeexplore.ieee.org/document/7870510/.
10. Attention is all you need / Ashish Vaswani, Noam Shazeer, Niki Parmar et al. // Advances in neural information processing systems. 2017. V. 30. https://arxiv.org/abs/1706.03762.
11. *Liu Gang*. Bidirectional LSTM with attention mechanism and convolutional layer for text classification / Gang Liu, Jiabao Guo // Neurocomputing. 2019. V. 337. P. 325–338. https://www.sciencedirect.com/science/article/pii/S0925231219301067.
12. *Doan Van-Sang*. Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network / Van-Sang Doan, Thien Huynh-The, Dong-Seong Kim // IEEE Geoscience and Remote Sensing Letters. 2022. V. 19. P. 1–5. https://ieeexplore.ieee.org/document/9229102/.
13. Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics / Peter C. Bermant, Michael M. Bronstein, Robert J. Wood et al. // Scientific Reports. 2019. aug. V. 9. № 1. P. 12588. https://www.nature.com/articles/s41598-019-48909-4.
14. Interpretable features for underwater acoustic target recognition / Junjun Jiang, Zhenning Wu, Junan Lu

et al. // Measurement. 2021. mar. V. 173. P. 108586. https://linkinghub.elsevier.com/retrieve/pii/S0263224120310988.

15. *Berlin Leonid*. Comparison of Information Criteria for Detection of Useful Signals in Noisy Environments / Leonid Berlin, Andrey Galyaev, Pavel Lysenko // Sensors. 2023. V. 23. № 4. https://www.mdpi.com/1424-8220/23/4/2133.
16. *Berlin Leonid*. Statistical complexity as a criterion for the useful signal detection problem / Leonid Berlin, Andrey Galyaev, Pavel Lysenko // Automation and remote control. 2023. V. 84. P. 852–871. http://ait.mtas.ru/en/archive/volume84issue7/AutRemControl2023-07-07Galyaev.pdf.
17. *Wang Junxiong*. Feature Extraction of Ship-Radiated Noise Based on Intrinsic Time-Scale Decomposition and a Statistical Complexity Measure / Junxiong Wang, Zhe Chen // Entropy. 2019. V. 21. № 11. https://www.mdpi.com/1099-4300/21/11/1079.
18. A Complexity-Based Approach for the Detection of Weak Signals in Ocean Ambient Noise / Shashidhar Siddagangaiah, Yaan Li, Xijing Guo et al. // Entropy. 2016. V. 18. № 3. https: //www.mdpi.com/1099-4300/18/3/101.
19. The Marinexplore and Cornell University Whale Detection Challenge. https://www.kaggle.com/competitions/whale-detection-challenge/overview/. 2013.

DEEP LEARNING APPROACH TO CLASSIFICATION OF ACOUSTIC SIGNALS USING INFORMATION FEATURES

P. V. Lysenko[a], I. A. Nasonov[a], Corresponding Member of the RAS A. A. Galyaev[a], and L. M. Berlin[a]

[a] *Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation*

Presented by Academician of the RAS A.L. Semenov

The paper considers the problem of binary classification of acoustic signals of biological origin recorded real environment. Information characteristics such as entropy and statistical complexity are chosen as the characteristic description of objects. The solution methods are based on three neural network architectures modified by the authors (on the Inception core, on the Inception core with the Residual technology, on the Self-Attention structure with LSTM blocks). A dataset from the Kaggle competition for detecting acoustic signatures of whales was used, and a comparison was made between the models in terms of the quality of solving the problem under consideration on a standard set of metrics. The AUC ROC value of more than 90% was obtained, which indicates the successful solution of the problem of detecting a useful signal and indicates the possible applicability of information characteristics to similar tasks.

Keywords: classification of time series, spectrogram, statistical complexity, deep learning