

АВТОМАТИЗИРОВАННОЕ ВЫЯВЛЕНИЕ АНОМАЛИЙ В ГИДРОФИЗИЧЕСКИХ ДАННЫХ: ПОДХОД НА ОСНОВЕ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ

Коротченко Р.А.

*Тихоокеанский океанологический институт им. В.И. Ильичева, г. Владивосток
romankor@mail.ru*

При анализе результатов гидрофизических наблюдений важным является процесс предварительной обработки полученных данных. На этом этапе требуется выполнить оценку качества полученных данных и провести процедуры фильтрации шумов аппаратуры, присутствующих в наблюдениях. Естественной задачей предобработки данных, направленной на удаление аппаратных сбоев, является поиск типичных аномалий, характерных для датчиков регистрации, например, таких как: замирание показаний из-за недостаточного питания и относительно резкие скачки в показаниях, нехарактерных физическим процессам. Если наблюдения проводятся длительное время и множеством датчиков, то актуальным становится разработка автоматизированных методов обнаружения аномалий. Традиционно для обнаружения аномалий используются статистические методы, однако развитие компьютерных технологий и методов машинного обучения показали эффективность обнаружения аномалий с использованием нейронных сетей, позволяющих изучать сложные и неочевидные взаимосвязи [1].

Регулярные сезонные наблюдения гидрологической обстановки на шельфе Японского моря в заливе Посыет и получаемые многосуточные наблюдений с десятков датчиков температуры в виде временных рядов определили актуальность автоматизации обработки данных регистрации. Задача обработки заключается в автоматическом обнаружении, фрагментации и классификации аномалий в гидрофизических временных рядах. Рассматриваются три класса фрагментов данных:

1. «Нормальные» данные: соответствуют ожидаемому поведению датчика;

2. «Замирание» датчика: характеризуется постоянством или ничтожным изменением показаний в течение длительного времени;

3. «Вспышки», блоки с возрастающей дисперсией: участки временного ряда с аномально высоким уровнем шума или резкими хаотическими колебаниями, несоответствующими натурным явлениям.

Под аномалиями понимаются второй и третий тип данных.

Для решения задачи обнаружения, фрагментации и классификации данных были разработаны две нейросети разной архитектуры: сеть типа BiLSTM (двунаправленная рекуррентная сеть с долгосрочной и краткосрочной памятью) и сеть типа CNN (сверточная нейронная сеть) [2].

Обе сети являются классификаторами данных типа «последовательность – последовательность» и позволяют классифицировать каждый временной шаг входных данных. Это позволяет обрабатывать временной ряд, чтобы выделять аномалий указанного типа. Входные данные для сетей содержат сегмент гидрофизических данных и рассчитанные признаки, после обработки которых сеть классифицирует точки поступивших данных и выделяет фрагменты, которые интерпретируются как нормальные и аномальные, класса «замирание» или «вспышка». В качестве признака применялись варианты: информационная энтропия Шеннона [3] для сегмента ряда и амплитуды дифференцирования значений по времени, чтобы выделять изменчивость регистрируемых значений.

На рисунке 1 представлены матрица ошибок по результатам обработки записей данных различными типами сетей, с использованием признака по информационной энтропии Шеннона.

На рисунке 2 представлена матрица ошибок по результатам обработки записей данных нейросетью типа BiLSTM с использованием признака изменчивости значений амплитуды.

Точность классификации аномалий типа «замирание» с помощью BiLSTM сети достигает 99.9%, нейросеть на основе CNN достигает точности 93.2%.

Точность классификации аномалий типа «замирание» с помощью BiLSTM сети достигает 86,6%. Важным развитием обнаружения аномалий различного типа с помощью искусственного интеллекта и нейросетей является поиск в океанологических наблюдениях природных феноменов, таких как температурные боры

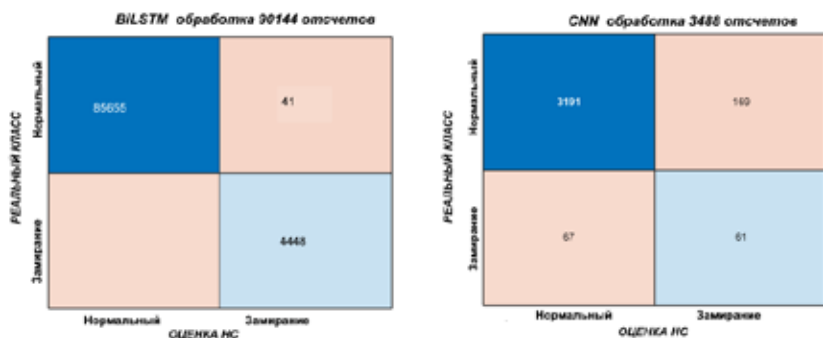


Рис. 1. Матрицы ошибок обработки временных рядов нейросетями разной архитектуры

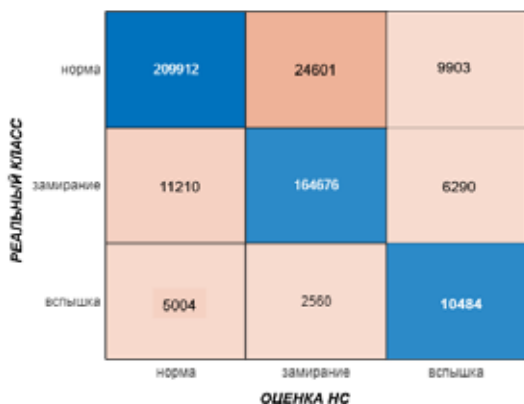


Рис. 2. Матрицы ошибок обработки временных рядов нейросетью BiLSTM

и вихревые волновые процессы.

Работа выполнена в рамках государственного задания по теме № 124022100074-9 «Изучение природы линейного и нелинейного взаимодействия геосферных полей переходных зон Мирового океана и их последствий».

Литература

1. Kang Hango, Kim Dongil, Lim Sungsu Machine Learning-Based Anomaly Detection on Seawater Temperature Data with Oversampling // J. Mar. Sci. Eng. 2024. V. 12. No 5. P. 807;
2. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. ДМК, 2018.
3. Верещагин Н. Информация, кодирование и предсказание. 2012. Издательство: МЦНМО, ФМОП.