

Обнаружение эффектов солнечных вспышек в вариациях амплитуды СДВ радиосигналов с применением автоэнкодера

Хак Хоанг Зыонг Нгуен, А. С. Полетаев

Иркутский национальный исследовательский технический университет

Аннотация: Обнаружение аномалий в вариациях амплитуды и фазы радиосигналов является одним из инструментов исследования динамических свойств канала радиосвязи. Развитие технологий машинного обучения при этом позволяет по-новому подойти к решению актуальных задач радиофизики. Целью настоящей работы является разработка алгоритма для поиска эффектов солнечных вспышек в вариациях амплитуды СДВ радиосигналов с использованием методов машинного обучения. Представлены разработанные модели нейронной сети на основе автоэнкодера для автоматического обнаружения аномальных изменений амплитуды по данным распространения СДВ радиоволн на трассе JJ1 (Япония, 22.2 кГц) – Иркутск (пункт приема п. Теплоэнергетик). После обучения нейронная сеть выполняет реконструкцию дневного хода амплитуды сигнала так, чтобы выходные данные максимально соответствовали распространению радиоволн в спокойных условиях. Аномальные изменения сигнала, связанные с возмущением ионосферы во время солнечных вспышек, обнаруживаются путем определения ошибки реконструкции и ее сравнения с рассчитанным пороговым значением. Порог определяется путем расчета среднего квадратичного значения ошибки восстановления набора обучающих данных. Результаты применения предложенных авторами автоэнкодеров для поиска аномалий показали свою эффективность (общая точность оказалась близка к 90%). Важным преимуществом метода является возможность поиска аномальных вариаций амплитуды в большом массиве экспериментальных данных с использованием измерений только для рассматриваемого дня. При этом исключается влияние аномалий, присутствующих в соседние дни, как это происходит при использовании усреднения реализаций суточного хода по нескольким дням до и после рассматриваемого события.

Ключевые слова: сверхдлинные волны, автоэнкодер, солнечные вспышки, аномалия, нейронная сеть, ВИВ, ионосфера.

Для цитирования: Нгуен Х. Х. З., Полетаев А. С. Обнаружение эффектов солнечных вспышек в вариациях амплитуды СДВ радиосигналов с применением автоэнкодера // Вестник СибГУТИ. 2025. Т. 19, № 1. С. 28–44. <https://doi.org/10.55648/1998-6920-2025-19-1-28-44>.



Контент доступен под лицензией
Creative Commons Attribution 4.0
License

© Нгуен Х. Х. З., Полетаев А. С., 2025

Статья поступила в редакцию 13.01.2024;
переработанный вариант 21.04.2024;
принята к публикации 18.06.2024.

1. Введение

Солнечные вспышки являются мощными взрывами на поверхности Солнца, которые вызывают выбросы энергии и вещества в космос. Они могут оказывать значительное

влияние на Землю, ионосферу и различные технические системы. Одним из способов изучения солнечных вспышек и физических процессов, происходящих в нижней ионосфере, является наблюдение за вариациями амплитуды и фазы СДВ радиосигналов. Автоматизированное обнаружение эффектов солнечных вспышек в экспериментальных данных наблюдений за распространением СДВ радиосигналов является крайне сложной задачей. Это связано с тем, что такие вариации зависят от мощности вспышек, зенитного угла, протяженности трассы, длительности вспышек. Помимо этого, значительные амплитудные изменения могут быть вызваны и другими факторами: магнитные бури, высыпание частиц из магнитосферы, красные спрайты, импульсные радиопомехи искусственного и естественного происхождения и др.

Традиционно для ассоциирования аномальных изменений амплитуды и фазы сигналов с возмущениями ионосферы во время солнечных вспышек используют данные о рентгеновском излучении Солнца. Однако проблема заключается в том, что при проведении долговременных исследований массивы экспериментальных данных очень большие (более 10 лет наблюдений), а количество анализируемых радиосигналов может исчисляться десятками. При этом распространение радиоволн от различных навигационных СДВ передатчиков происходит по трассам, находящимся в совершенно разных гелиогеофизических условиях. И помимо всего прочего, подобный массив данных возрастает многократно при организации синхронных наблюдений в нескольких пунктах приема.

Обнаружение аномалий относится к классу задач по идентификации экземпляров данных, которые имеют статистические отличающиеся особенности по сравнению с другими доступными данными, считающимися нормальными наблюдениями (в данном случае это распространение радиоволн в спокойных условиях). Существует много методов обнаружения аномалий [1-4], которые применяются в различных областях, таких как обнаружение медицинских биомаркеров, интеллектуальный анализ данных, обнаружение мошенничества, кибербезопасность и промышленный контроль качества. Есть также традиционные алгоритмы машинного обучения, которые используются для обнаружения аномалий, например: древовидная модель, оценка плотности, кластерный метод, вероятностная оценка и т. д. Однако их применение для обнаружения аномалий в радиофизических исследованиях является сложной задачей, поскольку аномальные явления могут встречаться редко или имеют разнообразные формы проявления. При этом сам процесс обучения предполагает использование большого количества размеченных образцов с аномалиями и без них.

В настоящей работе предлагается новый метод обнаружения эффектов солнечных вспышек в вариациях амплитуды СДВ радиосигналов с применением методов машинного обучения. Этот метод основан на использовании автоэнкодеров, которые являются нейронными сетями и фокусируются при обучении детекторов аномалий на большом количестве нормальных данных. Экспериментальные результаты тестирования показывают, что предлагаемый метод позволяет эффективно обнаруживать эффекты солнечных вспышек в вариациях амплитуды СДВ радиосигналов.

2. Описание данных и методики обработки

2.1. Описание данных

Исходными данными являются результаты долговременных измерений амплитуды и фазы сверхдлинноволновых радиоволн [5], которые излучаются радионавигационными станциями. Излученная волна распространяется в пространстве между земной поверхностью и ионосферой (волновод Земля – ионосфера). Таким образом, состояние ионизированных слоев атмосферы сильно влияет на параметры принимаемых сигналов. Такие изменения в ионосфере можно косвенно исследовать, выполняя статистический анализ и обработку накопленных данных измерений. Одним из факторов, влияющих на ионосферу, являются

солнечные вспышки. В результате увеличения плотности потока энергии во время солнечной вспышки изменяется электронная концентрация в ионосфере и эффективная высота отражающего слоя, что в свою очередь приводит к изменению амплитуды и фазы сверхдлинноволновых радиосигналов в точке приема [6]. В ИрНИТУ в течение многих лет проводится регистрация СДВ сигналов, на основе их вариаций осуществляется долговременный мониторинг состояния ионосферы [7]. За период с 2015 по 2020 год было отмечено 2139 солнечных вспышек [8].

На рисунке 1 приведена временная диаграмма, показывающая влияние солнечных вспышек на амплитуду СДВ сигнала станции JJI по данным наблюдений в п. Теплоэнергетик 1 и 2 апреля 2017 г. Верхняя панель диаграммы показывает вариации амплитуды сигнала, а нижняя – рентгеновское излучение по данным спутников GOES 13 и 15. Первый день характеризуется спокойными условиями распространения сигнала, второй день сопровождался возмущениями от трех вспышек С и М классов. Из диаграммы видно, что каждая вспышка привела к резкому увеличению амплитуды СДВ сигнала пропорционально ее классу мощности. Это увеличение было наиболее заметным в первые несколько минут после вспышки. После этого амплитуда сигнала постепенно снижалась, но все еще оставалась выше, чем до вспышки. Рентгеновское излучение, зарегистрированное аппаратурой GOES, также указывает на то, что вспышка 3 была значительной. Излучение достигло своего пика в первые несколько минут после вспышки, а затем постепенно снижалось. При этом схожая вспышка М-класса 1 апреля в 22:00 UTC воздействия на сигнал не оказала ввиду малого зенитного угла (большая часть трассы не была освещена) [9].

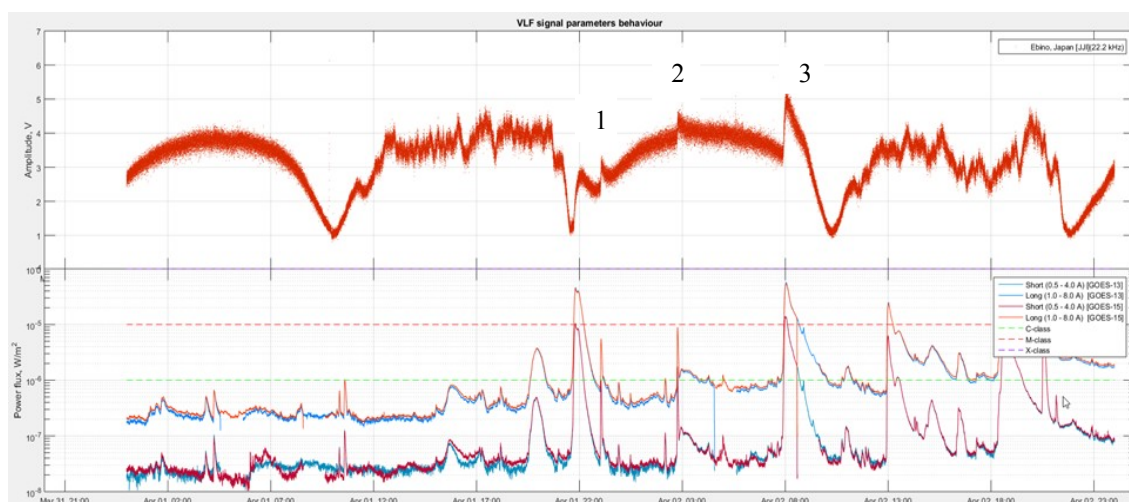


Рис. 1. Вариации амплитуды сигнала во время солнечных вспышек:
верхняя панель – вариации амплитуды сигнала JJI (22.2 кГц);
нижняя панель – рентгеновское излучение Солнца по данным GOES 13, 15

2.2. Подготовка данных

Предварительная обработка – этап, обязательный для анализа и обработки данных. Её цель – обеспечить точность, надежность и эффективность последующего анализа [10-12]. К особенностям сырых радиофизических данных измерений амплитуды относятся выбросы и искажения, вызванные импульсными радиопомехами, скачками в сети электропитания, атмосферными электромагнитными шумами и наводками. Также для исходных данных характерны неэквидистантность отсчетов (наличие пропусков), суточная вариативность флуктуаций амплитуды сигнала в соответствии с изменениями параметров волновода Земля-ионосфера. Предварительная обработка данных выполняется в несколько этапов: визуализация и фильтрация данных, усреднение, уменьшение размера, декомпозиция (нормализация) данных.

Визуализация

Данные представляют собой серию значений амплитуды СДВ сигнала во времени с частотой записи 1 раз в секунду. Данные визуализируются с помощью пользовательского интерфейса (рисунок 2), разработанного на языке Python. С помощью этого интерфейса можно визуализировать, выделять необходимые фрагменты данных, которые будут использоваться в процессе обучения и тестирования нейронной сети.

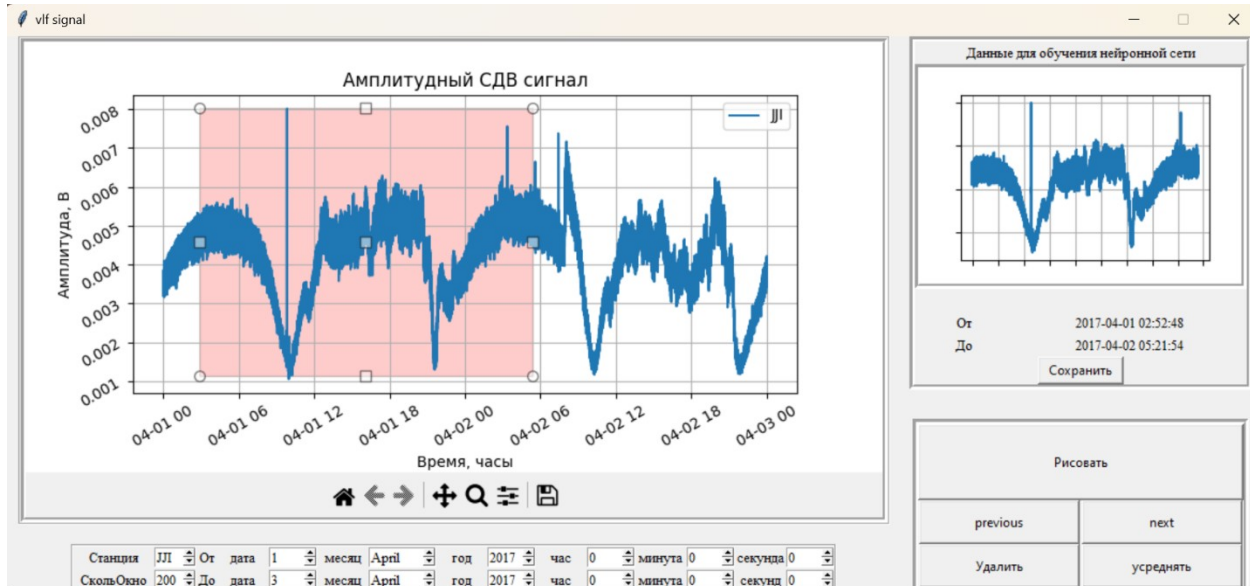


Рис. 1. Визуализация сигнала с помощью пользовательского интерфейса

Усреднение данных

Нейронные сети обучаются на данных, которые должны быть подготовлены перед обучением. Эта подготовка включает в себя удаление ошибок, пропусков и других аномалий, а также преобразование данных в формат, соответствующий входным требованиям модели нейронной сети. Один из способов очистки данных перед обучением нейронной сети – это усреднение данных. Усреднение данных заключается в вычислении среднего значения для набора данных в пределах окна усреднения. Это позволяет помочь уменьшить шум и другие помехи в данных, что улучшает качество обучения нейронной сети. Чтобы сохранить аномальные изменения амплитуды сигнала, ассоциированные с солнечными вспышками, размер окна усреднения выбирается на порядок меньше, чем длительность солнечных вспышек.

В пользовательском интерфейсе нейронной сети добавлена функция усреднения данных. Эта функция позволяет пользователю выбирать размер окна, используемого для усреднения данных. Размер окна определяет, сколько данных будет использоваться для вычисления среднего значения. В программе реализован метод экспоненциального сглаживания [13], который вычисляет значения сглаженного ряда путем взвешенного сложения отсчетов. Информация предыдущего и текущего шагов берётся с разными весами, которыми можно управлять. В простейшем варианте одинарного сглаживания i -ый отсчет сглаженного ряда находится как:

$$s_i = \alpha x_i + (1 - \alpha) s_{i-1},$$

где $\alpha = \frac{2}{N+1}$ – коэффициент отношения, $0 \leq \alpha \leq 1$; N – количество отсчетов сглаживания (размер окна); x_i – исходный ряд; s_i – сглаженный ряд.

Параметр α определяет соотношение между несглаженным значением на текущем шаге и сглаженным значением предыдущего шага. При $\alpha = 1$ берутся только точки исходного ряда, т. е. никакого сглаживания не будет. При $\alpha = 0$ суммируются только сглаженные

значения с предыдущих шагов, т. е. ряд превратится в константу. Для предварительной обработки выбрано окно $N = 200$.

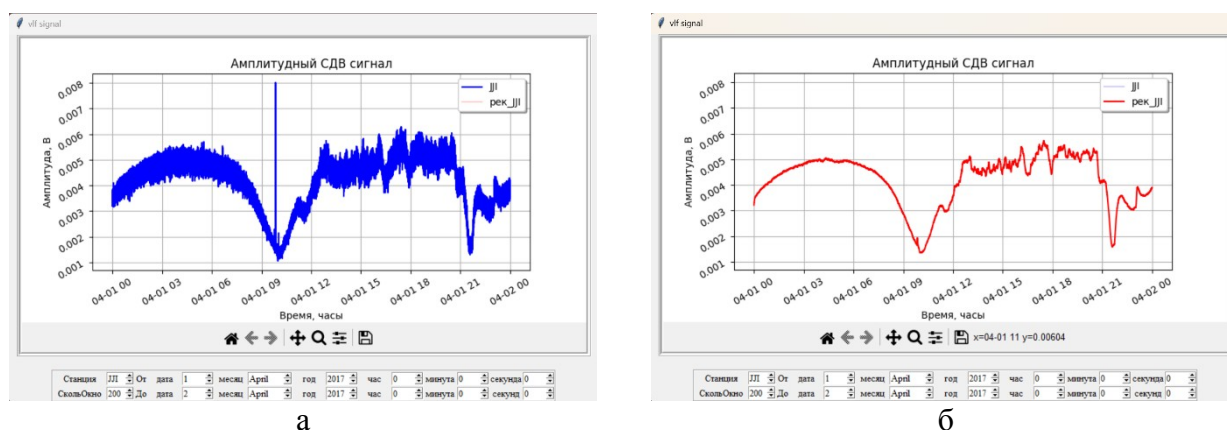


Рис. 2. Сигнал, полученный от радиостанции JJ1 1 апреля 2017 г. до усреднения (а) и после усреднения (б)

На рис. 3а показан сигнал, полученный от радиостанции в Японии JJ1 1 апреля 2017 г. На рис. 3б показан этот же сигнал, но после усреднения. Видно, что сигнал до усреднения (синий) имеет большую дисперсию, а сигнал после преобразования (красный) становится гладким, без выброса в центре, лучше прослеживаются флуктуации, вызванные волноводным характером распространения радиосигнала. Это облегчает работу нейронной сети с данными.

Уменьшение размера данных

Исходные экспериментальные данные, используемые для обучения, имеют табличную форму. Каждый столбец представляет значения амплитуды сигнала в определенный момент времени. Интервал в 8 часов соответствует 28800 столбцам, поскольку период измерений амплитуды составляет 1 с. Каждая строка соответствует дню измерений и является выборкой, используемой для обучения нейронной модели. Столбец при этом представляет собой атрибут модели. Для построения модели с использованием этих данных можно использовать нейронную сеть с 28800 входными нейронами. Однако такой подход создает сложности для обучения и является избыточным для целей данной задачи. Поэтому можно уменьшить количество атрибутов, не оказав существенного воздействия на результаты работы нейросети.

Статистика Бельгийского института солнечных исследований (Solar-Terrestrial Centre of Excellence) [14] показывает, что солнечные вспышки обычно длятся от нескольких минут до нескольких десятков минут или даже нескольких часов в зависимости от их класса. На основе этой статистики предлагается взять значение сигнала раз в 2 минуты, что приведет к уменьшению количества точек данных в 120 раз. Следовательно, количество входных нейронов составит 240. Таким образом, на вход алгоритма поступает фрагмент данных измерений для дневного интервала времени длительностью 8 часов (28800 отсчетов), а после децимации с коэффициентом прореживания 120 нейросетевая модель получает уже 240 отсчетов (на каждый отсчет создается отдельный входной нейрон). Такой подход позволяет упростить модель, уменьшить количество входных параметров и сохранить при этом ключевую информацию для достижения целей исследования. На верхней панели рисунка 4 представлен дневной ход амплитуды сигнала до прореживания, а на нижней – после. Заметно, что между ними нет существенных отличий, флуктуации амплитуды и всплеск в центре, произошедший в результате воздействия на ионосферу солнечной вспышки, сохранены.

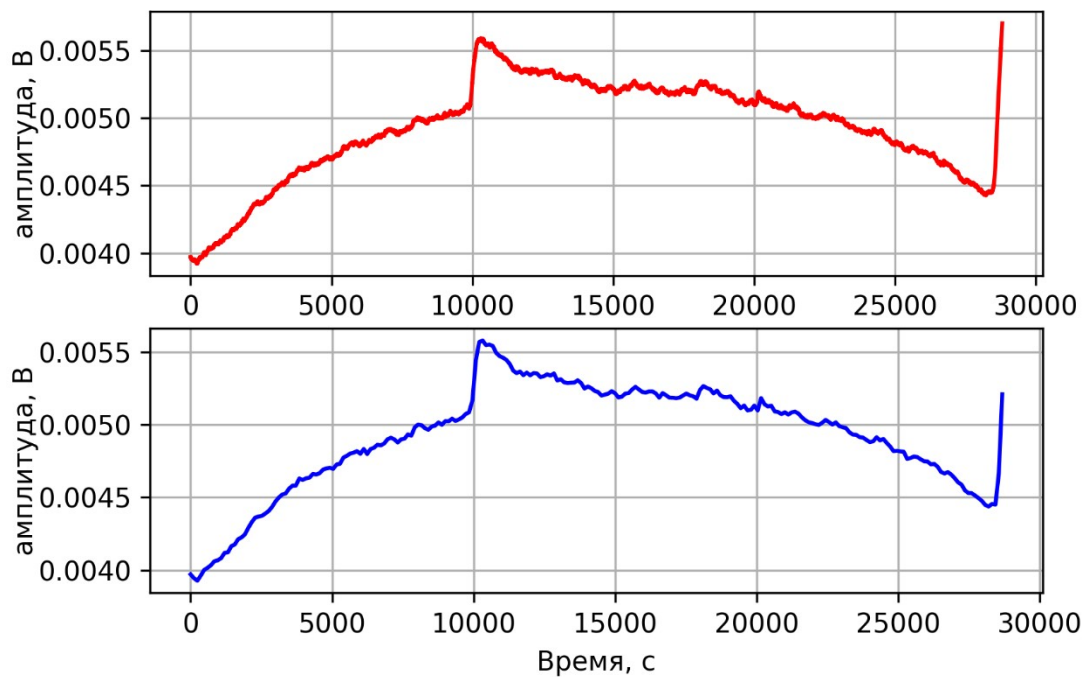


Рис. 2. Сигнал до (красный) и после (синий) уменьшения размера

Нормализация данных

Нормализация данных перед обучением нейросети является важным этапом для обеспечения эффективности и стабильности работы модели. Основная цель этого процесса – привести значения признаков к одному и тому же диапазону или распределению. Это помогает модели эффективнее изучать взаимосвязи между ними и избежать влияния различных масштабов между признаками.

В случае регистрации СДВ радиосигналов есть несколько факторов, которые влияют на динамический диапазон вариаций амплитуды. Во-первых, помимо того, что различные передающие СДВ радиостанции имеют разную номинальную мощность излучения, изменение мощности возможно в рамках работы одного передатчика. Во-вторых, регистрируемые сигналы имеют разные частоты и неодинаковые условия распространения (географические широты, коэффициенты затухания, протяженности трасс), которые, как правило, также имеют выраженный сезонный характер изменений. Кроме того, могут отличаться частотные характеристики приемных трактов регистраторов сигналов.

Одним из распространенных методов нормализации данных является приведение значений по максимальному или минимальному уровню (Min-Max Scaling). Обычно данные приводятся к диапазону значений $[0; 1]$. В результате нормализации данные получают в относительных единицах (отн. ед.).

2.3. Автоэнкодер

Автоэнкодер [15, 16] – это нейронная сеть прямого распространения, которая обучается без учителя. Целью применения автоэнкодера является анализ входных данных и их реконструкции на выходе. Автоэнкодер обучается путем сведения к минимуму различий между входными данными и выходными данными реконструкции, то есть выделяет общие тренды в данных. Применительно к решаемой задаче такая нейросеть может выполнить реконструкцию суточного хода амплитуды радиосигнала, убирая при этом всплески, ассоциированные с солнечными вспышками. По величине ошибки реконструкции можно судить о наличии солнечной вспышки в рассматриваемый день.

Архитектура автоэнкодера

Архитектура нейросети такого типа включает в себя 3 основные части: энкодер, скрытый слой, декодер (рис. 5). Энкодер – модуль, который сжимает входные данные. Обычно количество нейронов на его выходе на несколько порядков меньше размера исходных данных. Скрытый слой – модуль, содержащий основные характеристики входных данных, он связан с выходом кодировщика. Это наиболее важная часть сети, поскольку она содержит основные характеристики входных данных, которые используются для получения характеристик сигнала и его воспроизведения. Декодер – сетевой модуль, который распаковывает представленные в скрытом слое входные данные и реконструирует структуры данных из их закодированной формы.

Принцип работы автоэнкодера состоит из двух этапов: кодирование (сжатие) и декодирование (восстановление). На вход автоэнкодера подаются нормализованные данные об амплитуде сигнала, которые он должен сжать и восстановить. В процессе сжатия происходит потеря всплесков, вызванных возмущениями ионосферы в результате солнечных вспышек. На выходе автоэнкодера получается такая реконструкция исходных данных, как если бы вспышек не было.

Для сравнения реконструированных данных автоэнкодера и их исходных версий часто используется функция потерь. Например, при использовании в качестве функции потерь среднеквадратичной ошибки (MSE) реконструированные данные сравниваются с исходными данными по их среднеквадратичному отклонению.

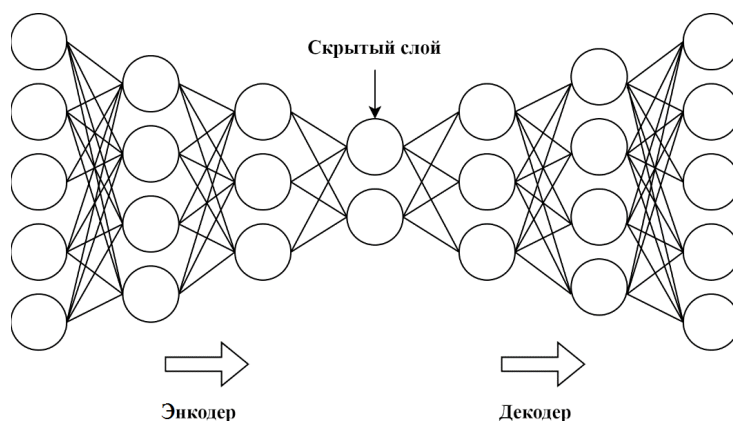


Рис. 3. Структура автоэнкодера

Для того, чтобы создать автоэнкодер, выбраны 4 суперпараметра, а именно:

1. Размер кода (скрытого слоя), который определяет степень сжатия информации.
2. Количество слоев для настройки глубины кодировщика и декодера. Чем больше глубина (количество слоев), тем сложнее модель и меньше скорость обработки.
3. Количество узлов на слой – параметр, который определяет вес, используемый на каждом уровне. Обычно количество узлов уменьшается с каждым последующим слоем.
4. Функция потерь, которая зависит от типа вывода и ввода модели [17]. После завершения прямого расчета выходной прогноз сравнивается с входной меткой, и разница рассчитывается с помощью функция потери. Потери позволяют модели посредством обратного распространения ошибки обновлять параметры модели так, что последующие прогнозы выполняются в правильном направлении и достигают улучшенных результатов.

2.4. Алгоритм обнаружения аномалий и оценка точности

Алгоритм обнаружения аномалий включает несколько этапов. Сначала модель автоэнкодера обучается на нормальных данных, то есть на вход нейронной сети подаются

только сигналы без аномалий. Это позволяет обучить модель основным закономерностям и характеристикам нормальных данных для восстановления суточного хода амплитуды в спокойные дни.

После того, как модель освоила правильное закодированное представление скрытого пространства и восстановила нормальные сигналы, анализируется среднее значение ошибки. Это необходимо для того, чтобы установить эффективный порог, характеризующий распределение потерь для нормальных и для аномальных данных. Этот порог устанавливается на одно стандартное отклонение выше средней потери, получаемой при обучении на нормальных данных. В результате получается задача двоичной классификации, так как полученный порог определяет, являются данные нормальными или нет. Если ошибка восстановления входных данных превышает пороговое значение, то они классифицируются как аномальные.

Ошибка реконструкции, также называемая ошибкой прогнозирования, может быть вычислена с помощью статистической оценки, такой как среднеквадратическая ошибка (MSE), средняя абсолютная ошибка (MAE), квадратичная ошибка прогнозирования (SPE) или абсолютное значение. Поскольку эти средства оценки просто вычисляют расстояние между исходными входными данными и восстановленными выходными данными, выбор средства оценки ошибок не влияет на конечный результат. Для расчета ошибки реконструкции для каждого сигнала был выбран метод MSE. Предположим, что отсчеты $X = \{x_1, x_2, \dots, x_n\}$ – входной сигнал, а $Y = \{y_1, y_2, \dots, y_n\}$ – соответствующий ему восстановленный сигнал. Тогда MSE как ошибка восстановления описывается уравнением

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2$$

Для определения порога обучающая выборка данных пропускается через обученную сеть. На выходе получается выборка восстановленных сигналов, по которой далее рассчитывается ошибка для каждого сигнала. Предположим, что $E = \{e_1, e_2, \dots, e_m\}$ – полученный набор ошибки восстановления, где m – количество сигналов в обучающей выборке. Таким образом, порог принятия решения Π определяется следующим выражением:

$$\Pi = \frac{1}{m} \sum_{i=1}^m e_i + \sqrt{\frac{1}{m} \sum_{i=1}^m \left(e_i - \frac{1}{m} \sum_{i=1}^m e_i \right)^2},$$

где $\frac{1}{m} \sum_{i=1}^m e_i$ – среднее значение ошибки восстановления; $\sqrt{\frac{1}{m} \sum_{i=1}^m \left(e_i - \frac{1}{m} \sum_{i=1}^m e_i \right)^2}$ – оценка стандартного отклонения.

Чтобы оценить модель нейронной сети, необходимо проанализировать ее способность делать прогнозы на новом наборе данных. Традиционные параметры, используемые для оценки модели нейронной сети: время обучения, точность (англ. accuracy), полнота (англ. recall), точность предсказания положительных случаев (англ. precision). На рис. 6 представлена матрица ошибок, с помощью которой можно определить эффективность прогноза. В данном контексте «да» означает, что модель считает сигнал аномальным (сигнал со вспышками) и классифицирует его как положительный результат, а «нет» означает, что модель считает сигнал нормальным (сигнал без вспышек) и классифицирует его как отрицательный результат.

		Реальность	
		Да	Нет
Прогноз	Да	TP	FP
	Нет	FN	TN

Рис. 4. Матрица ошибок

Точность (accuracy) – это метрика, которая показывает процент того, как часто модель делает правильную классификацию и нормальных, и аномальных сигналов по сравнению с общим числом примеров. Формула для расчета точности A выглядит следующим образом:

$$A = \frac{TP + TN}{TP + FP + TN + FN}.$$

Здесь TP (true positive) – количество верно выявленных аномальных сигналов;

TN (true negative) – количество верно выявленных нормальных сигналов;

FP (false positive) – количество нормальных сигналов, которые модель ошибочно классифицировала как аномальные;

FN (false negative) – количество аномальных сигналов, которые модель ошибочно классифицировала как нормальные;

Полнота (recall) – это метрика, которая измеряет способность модели выявлять положительные случаи в данных. В данной работе полнота показывает, какую долю всех сигналов со вспылками модель способна обнаружить. Высокая полнота означает, что модель эффективно выявляет большинство аномальных сигналов, и чем меньше данных со вспылками было пропущено, тем выше полнота. Формула для расчета полноты R выглядит следующим образом:

$$R = \frac{TP}{TP + FN}.$$

Точность предсказания положительных случаев – это метрика, показывающая долю сигналов, классифицированных как аномальные, действительно являющихся аномальными. По сравнению с предыдущей метрикой акцент в этом параметре делается на ошибки классификации. Чем меньше модель ошибается, помечая нормальные данные как аномальные, тем ближе этот параметр к единице. Формула для расчета точности P выглядит следующим образом:

$$P = \frac{TP}{TP + FP}.$$

3. Разработка и тестирование нейросетевой модели

3.1. Создание и обучение нейронной сети-автоэнкодера

После предварительной подготовки и обработки данных было выполнено проектирование сетей на основе автоэнкодера для обучения и создания моделей прогнозирования обучения. Размер входного слоя для всех сетей автоэнкодера равен длине

сегмента каждого набора данных. Поскольку выходные данные представляют собой реконструкцию входных данных, размер выходного слоя совпадает с размером входного.

На рисунке 7 показаны порядок слоев и структура сетей автоэнкодера, используемых в этом исследовании. Блок энкодер состоит из 5 слоев: первый имеет 240 нейронов во входном слое, в следующих 4 слоях количество нейронов уменьшается в порядке 128, 64, 32, 16 соответственно. Блок декодер аналогичен, но в обратном порядке. В данной работе рассматриваются несколько вариантов с разным количеством нейронов в скрытом слое (1, 2, 4, 8). На рисунке 7 отображен созданный автоэнкодер с восемью нейронами в скрытом слое.

Нейронные сети обучались с помощью алгоритма Adam (adaptive moment estimation) [18], проходя через набор обучающих данных на протяжении 500 циклов (эпох). На каждой эпохе сеть обучалась на подмножестве данных размером 10. Перед каждой эпохой обучающие данные перемешивались, чтобы предотвратить переобучение. Функцией потерь была задана средняя абсолютная ошибка (MAE).

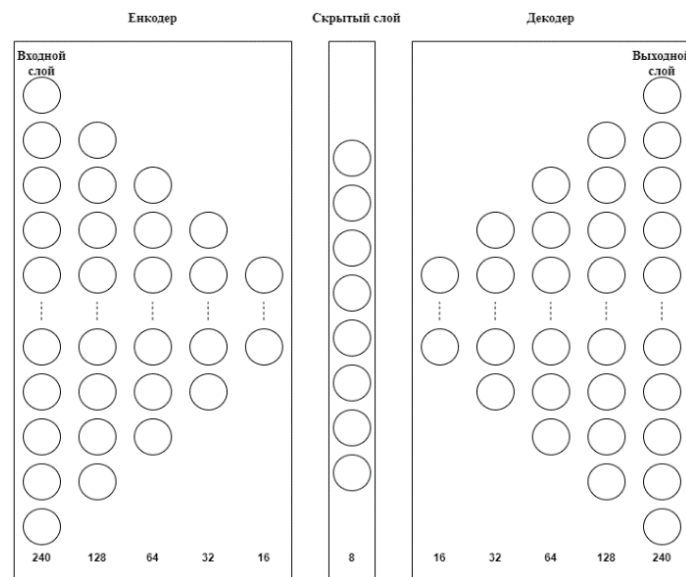


Рис. 5. Созданный автоэнкодер

Процесс обучения модели осуществлялся путем оптимизации функции потерь. После завершения прямого расчета модели с входными данными выходной прогноз сравнивается с входной меткой и рассчитывается разница с помощью метода средней абсолютной ошибки MAE. Потери позволяют модели посредством обратного распространения ошибки обновлять параметры модели так, что последующие прогнозы направляются в правильном направлении и достигают улучшенных результатов. Функция MAE вычисляется согласно уравнению:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|,$$

где MAE – сумма абсолютной разницы прогноза модели x_i и истинное значение y_i для всех точек данных n .

Как упоминалось, скрытое пространство действует как извлечения признаков входных данных. Затем модуль декодера использует только эти скрытые функции и переменные для восстановления выходных данных.

3.2. Результаты обучения и тестирования автоэнкодера

В ходе проведения экспериментов использовался набор экспериментальных данных (дневной ход амплитуды сигналов с 0 до 8 ч. UTC+8), связанных с распространением СДВ радиоволн по трассе от СДВ передатчика JJ1 (Япония) до Иркутска (пункт приема в

п. Теплоэнергетик). В этом наборе данных охвачено в общем 66 дней за первые 4 месяца 2017 года. Для обучения модели были выбраны 50 нормальных (невозмущенных) дней, в то время как для тестирования использовались 16 дней, включая 4 дня с наличием аномалий. В данном контексте аномалии представляют собой солнечные вспышки, отмеченные в данных.

Для четырех моделей нейронной сети с разным количеством нейронов в скрытом слое получены результаты обучения, приведенные на рисунке 8. Здесь показана ошибка MAE, которая рассчитывается после каждого цикла обучения. Так как исходные данные нормализованы и значения варьируются в пределах от 0 до 1, то MAE также приведено в относительных единицах. Верхний ряд графиков отображает процесс обучения нейронной сети с 1 и 2 нейронами в скрытом слое (панели 8а и 8б), а нижний ряд – с 4 и 8 нейронами в скрытом слое (панели 8в и 8г). Синяя линия обозначает ошибки модели на обучающей выборке данных, а оранжевая – на проверочной.

Как следует из графиков, во всех четырех случаях ошибка сокращается с каждой эпохой, однако для моделей с четырьмя и восемью нейронами в скрытом слое ошибка достигает своего минимума быстрее, чем в двух других моделях. Ошибка обучения для моделей с одним и двумя нейронами приближается к 0.02 (2 %), в то время как для четырех и восьми нейронов она приближается к нулю уже после 200 циклов обучения. Кроме того, значения потерь на обучающей выборке и проверочной выборке для моделей с четырьмя и восемью нейронами в скрытом слое практически одинаковы. Это подтверждает, что модель обучена очень хорошо и делает почти полную реконструкцию суточного хода амплитуды невозмущенного сигнала.

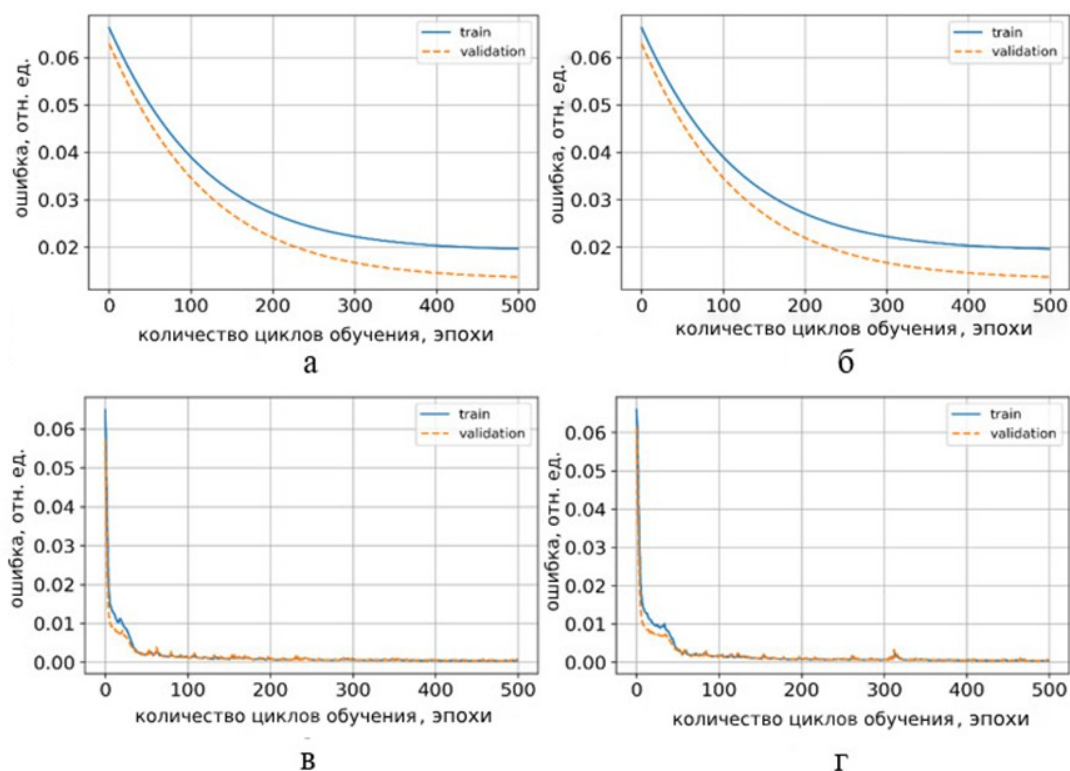


Рис. 6. Процесс обучения нейросетей с разным количеством нейронов в скрытом слое:

а – 1; б – 2; в – 4; г – 8

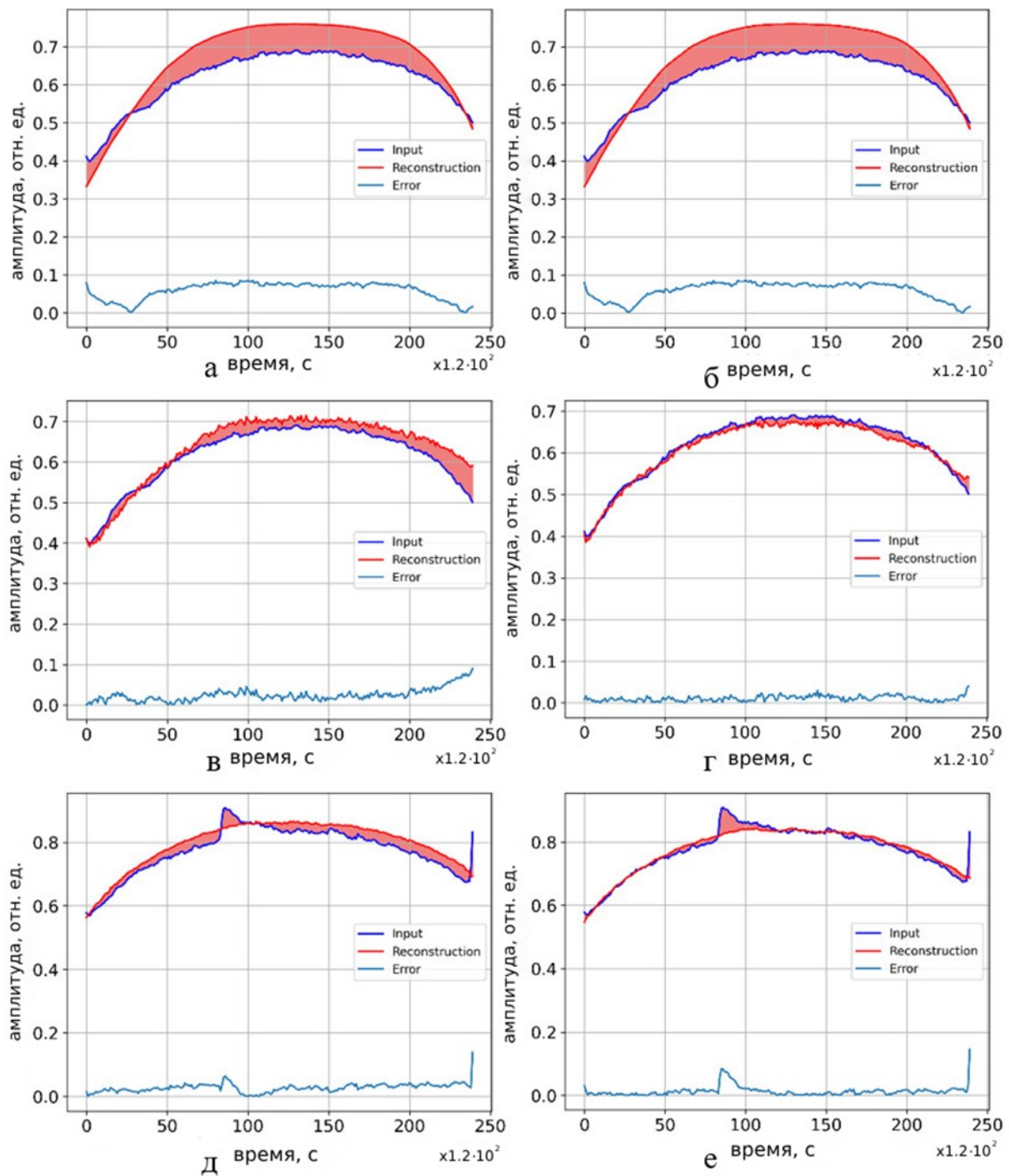


Рис. 7. Дневной ход амплитуды сигнала на входе и выходе обученной нейросети с разным количеством нейронов в скрытом слое: а – 1; б – 2; в, д – 4; г, е – 8

На рис. 9 представлены графики входных (синий цвет) и выходных (красный цвет) сигналов и их ошибки реконструкции (голубой цвет) для каждой обученной модели. По вертикали отображены значения амплитуды сигнала в относительных единицах, а по горизонтали – время в секундах (шкала имеет масштабирующий коэффициент $1.2 \cdot 10^2$). Как следует из графиков на панелях 9а и 9б, которые соответствуют моделям с одним и двумя нейронами в скрытом слое, входные и выходные сигналы имеют очень малое перекрытие. Это указывает на то, что модель дает неправильное восстановление суточного хода амплитуды по сравнению с исходными данными даже в спокойные дни. На графиках панелей 9в и 9г для моделей с большим количеством нейронов в скрытом слое (4 и

8 нейронов) наблюдается практически полное соответствие между входными и выходными сигналами для случая, когда солнечных вспышек не было.

На основе полученных функций выполняется обнаружение аномалий в СДВ сигналах. Основная идея обнаружения аномалий заключается в том, что реконструированный ход амплитуды и входные данные существенно отличаются для возмущенных и спокойных условий распространения радиосигнала. На панелях 9д и 9е показаны вариации амплитуды сигнала на входе и выходе обученной нейросети для дня с солнечной вспышкой. В модели с четырьмя нейронами в скрытом слое вспышка повлияла на восстановление дневного хода амплитуды, в то время как модель с восемью нейронами практически полностью убирает тренд суточного хода и позволяет выделить всплеск (отмечен стрелкой), вызванный солнечной вспышкой.

Таким образом, обученная сеть, имеющая восемь нейронов в скрытом слое, хорошо восстанавливает суточный ход без аномальных всплесков амплитуды, характерных для проявления эффектов от солнечных вспышек. Это означает, что ошибки восстановления (разностная функция) для нормальных и аномальных данных весьма различны, и для обнаружения аномалии на основе ошибки реконструкции далее вычисляется порог Π в качестве границы принятия решения.

В таблице 1 отражены результаты тестирования четырех моделей и их эффективность. Из таблицы следует, что время, требуемое для обучения модели, несущественно отличается друг от друга для выбранного набора данных. Модель с одним нейроном в скрытом слое требует больше времени и, также как и модель с двумя нейронами, показывает более низкие результаты. Это говорит о том, что количество нейронов в скрытом слое недостаточно, что приводит к тому, что модель не может достаточно хорошо запомнить характеристики входного сигнала, и, как результат, восстановленный выходной сигнал значительно отличается от входного. Модели с большим количеством нейронов в скрытом слое демонстрируют более высокие показатели. Наилучшие результаты получены для модели с восемью нейронами, общая точность оказалась близка к 90 %.

Таблица 1. Показатели производительности обученных нейросетей

Количество нейронов в скрытом слое	Время обучения, с	Accuracy, %	Recall, %	Precision, %
1	21.01	68.75	0.00	0.00
2	16.68	68.75	0.00	0.00
4	16.85	75.00	75.00	50.00
8	17.25	87.50	75.00	60.00

4. Заключение

Задача обнаружения аномалии в СДВ сигналах была успешно решена с применением нескольких моделей автоэнкодера. Такой подход позволил выполнить машинное обучение с использованием относительно небольшого объема данных, взятых за первые четыре месяца 2017 года. В результате тестирования на этом наборе данных выявлена высокая эффективность разработанных нейросетей для моделей с четырьмя и восемью нейронами в скрытом слое. Это свидетельствует о высоком потенциале модели для автоматизации обнаружения эффектов солнечных вспышек в больших массивах экспериментальных данных. Общая точность работы модели с 8 нейронами составила 87.50 %, что является высоким показателем.

Преимуществом предложенного способа является возможность реконструкции суточного хода амплитуды сигнала с устранением всплеска от вспышки, используя только данные текущего дня. Для сравнения, классический подход предполагает усреднение

суточного хода по нескольким дням до и после события. Для его корректной работы необходимо, чтобы усредняемые данные были без пропусков и без возмущений. Иначе на вектор ошибки оказывают влияние вспышки из всех дней, участвовавших в усреднении.

В дальнейшем планируется автоматизировать сканирование многолетних данных наблюдений и при помощи разработанного метода выполнить их разметку. Предложенный подход удобен тем, что позволяет не только обнаружить в экспериментальных данных наличие всплесков амплитуды, вызванные вспышками, но и идентифицировать их временные параметры (время максимума и величина всплеска, длительность отклика на вспышку), что будет использоваться для построения модели нижней ионосферы.

Литература

1. Савицкий Д. Е., Дунаев М. Е., Зайцев К. С. Выявление аномалий при обработке потоковых данных в реальном времени // International Journal of Open Information Technologies. 2022. Т. 14, № 6. С. 70–75.
2. Sarvani A., Venugopal B., Devarakonda N. Anomaly Detection Using K-means Approach and Outliers Detection Technique // Advances in Intelligent Systems and Computing: Soft computing: Theories and applications, Springer. 2019. Vol. 742. P. 375–385.
3. Крашенинников В. Р., Кувайскова Ю. Е., Маленова О. Е., Субботин А. Ю. Обнаружение аномалий на квазипериодических процессах, цилиндрических и круговых изображениях и их системах // Автоматизация процессов управления. 2022. № 4, С. 67–75.
4. Токарев Д. А., Погребан С.В. Аномалии трафика в ip сетях и их обнаружение // Известия института инженерной физики. 2013. № 4, С. 13–17.
5. Токмачев Д. А., Полетаев А. С., Безрукин А. Г., Ченский А. Г., Засенко В. Е., Губин Н. А. Система синхронизации сверхдлинноволнового интерферометра // Приборы и техника эксперимента. 2014. № 5. С. 77–84.
6. Гаврилов Б. Г., Ермак В. М., Ляхов А. Н., Поклад Ю. В., Рыбаков В. А., Ряховский И. А. Восстановление параметров среднеширотной нижней ионосферы при солнечных вспышках М- и Х-классов // Геомагнетизм и аэрономия. 2020. Т. 60, № 6. С. 762–768.
7. Полетаев А. С., Ченский А. Г., Токмачев Д. А. Регистрация вариаций амплитуды и фазы сигналов навигационных СДВ передатчиков // XIV конференция молодых ученых «Взаимодействие полей и излучения с веществом». Иркутск, 14–18 сентября. 2015. С. 252–255.
8. Arenicola O.V., Poletaev A.S., Chensky A.G. Influence of Solar Flares on VLF Radio Waves Propagation over JJI – Irkutsk Path // XIV International scientific-technical conference on Actual Problems of Electronic Instrument Engineering (APEIE – 2018). Novosibirsk, October 2 – 6, 2018. Vol. 1, Part 3, pp. 137 – 140.
9. Бельский М. И., Орлов А. Б., Пронин А. Е., Уваров А. Н. О зависимости величины ВФА СДВ от зенитного угла Солнца на протяженных трассах в различных широтах // Материалы VIII Регион. конф. по распространению радиоволн, СПб, С.-Петербург. гос. ун-т, 29–30 октября 2002 г.
10. Жанаева С. Б. К вопросу о подготовке данных при разработке модели нейронной сети // Вестник СибГУТИ. 2022. Т. 16, № 4. С. 69–79.
11. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer Series in Statistics / 2nd Edition, 2017. 764 p.
12. Грас Д. Data science. Наука о данных с нуля: пер. с англ. СПб: БХВ-Петербург, 2017. 336 с.
13. Львович И. Я., Семенов С. Н., Гладских Н. А., Пронин С. С. Методика прогнозирования медико-демографических показателей на базе методов экспоненциального сглаживания и множественной регрессии // Научно-медицинский вестник центрального Черноземья. 2007. № 27. С. 104–108.

14. STCE Newsletter for 4 Jan 2016 – 10 Jan 2016 [Электронный ресурс]. URL: <http://www.stce.be/newsletter/pdf/2016/STCEnews20160115.pdf> (дата обращения: 11.03.2024).
15. Chong Q., Zhiguo Z. et al. Predictive anomaly detection for marine diesel engine based on echo state network and autoencoder // Energy Reports. 2022. Vol. 8. P. 998–1003.
16. Juan Terven, Wang H. et al. Anomaly detection for hydropower turbine unit based on variational modal decomposition and deep autoencoder // Energy Reports. 2021. Vol. 7. P. 938–946.
17. Terven J., Cordova-Esparza D. M., Ramirez-Pedraza A., Chavez-Urbiola E. A., Romero-Gonzalez J. A. Loss Functions and Metrics in Deep Learning. [Электронный ресурс]. URL: <https://arxiv.org/abs/2307.02694>. (дата обращения: 28.09.2023).
18. Diederik P. Kingma, Jimmy Lei Ba. Adam: a method for stochastic optimization // International Conference on Learning Representations. Toulon, april 24–26, 2015.

Нгуен Хак Хоанг Зыонг

аспирант кафедры радиоэлектроники и телекоммуникационных систем, Иркутский национальный исследовательский технический университет (ИрНТУ, 664074, г. Иркутск, ул. Лермонтова, д. 83), тел. +7 961 1576 679, e-mail: hd04011996@gmail.com, ORCID ID: 0009-0000-1561-8741.

Полетаев Александр Сергеевич

к.ф.-м.н., доцент кафедры радиоэлектроники и телекоммуникационных систем, Иркутский национальный исследовательский технический университет (ИрНТУ, 664074, г. Иркутск, ул. Лермонтова, д. 83), тел. +7 395 2405 129, e-mail: poletaevas@ex.istu.edu, ORCID ID: 0000-0002-6448-0045.

Авторы прочитали и одобрили окончательный вариант рукописи.

Авторы заявляют об отсутствии конфликта интересов.

Вклад соавторов: Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.

Detection of solar flare effects in the amplitude variations of VLF radio signals using an autoencoder

Khac Hoang Duong Nguyen, Aleksandr S. Poletaev

Irkutsk National Research Technical University

Abstract: Detection of anomalies in changes in radio signal parameters is one of the tools for studying the dynamic properties of a radio communication channel. The development of machine learning technologies allows us to take a new approach to solving pressing problems in radio physics. The purpose of this work is to develop an algorithm for searching for the effects of solar flares in variations in the amplitude of VLF radio signals using machine learning methods. The developed models of a neural network based on an autoencoder are presented for automatic detection of anomalous amplitude changes based on the propagation data of VLF radio waves on the JJI (Japan, 22.2 kHz) – Irkutsk (receiving point in Teploenergetik) path. After training, the neural network reconstructs the daily variation of the signal amplitude so that the output data closely matches the propagation under quiet conditions. Anomalous signal changes associated with ionospheric disturbances during solar flares are detected by determining the reconstruction error and comparing it with a calculated threshold value. The threshold is determined by calculating the root mean square value of the reconstruction error of

the training data set. The results of using the autoencoders proposed by the authors to search for anomalies have shown their effectiveness (the overall accuracy was close to 90%). An important advantage of the method is the ability to perform an assessment using data only for the measurement day in question, which eliminates the influence of anomalies present on neighboring days.

Keywords: VLF signal, autoencoder, solar flares, anomaly, neural network, SID, ionosphere

For citation: Nguyen K H. D., Poletaev A. S. Detection of solar flare effects in the amplitude variations of VLF radio signals using an autoencoder. *Vestnik SibGUTI*, 2025, vol. 19, no. 1, pp. 28-44. <https://doi.org/10.55648/1998-6920-2025-19-1-28-44>.



Content is available under the
license
Creative Commons Attribution 4.0
License

© Nguyen K H. D., Poletaev A. S., 2025

The article was submitted: 13.01.2024;
revised version: 21.04.2024;
accepted for publication 18.06.2024.

References

1. Savitskiy D. E., Dunaev M. E., Zaitsev K. S. Vyyavlenie anomalii pri obrabotke potokovykh dannykh v realnom vremeni [Anomaly Detection in Real-Time Stream Data Processing]. *International Journal of Open Information Technologies*, 2022, vol. 14, no. 6, pp. 70–75.
2. Sarvani A., Venugopal B., Devarakonda N. Anomaly Detection Using K-means Approach and Outliers Detection Technique // *Advances in Intelligent Systems and Computing: Soft computing: Theories and applications*, Springer, vol. 742, 2019, pp. 375–385
3. Krascheninnikov V. R., Kuvaiskova Yu. E., Malenova O. E., Subbotin A. Yu. Obnaruzhenie anomalii na kvaziperiodicheskikh protsessakh, tsilindricheskikh i krugovykh izobrazheniyakh i ikh sistemakh [Detection of anomalies in quasiperiodic processes, cylindrical and circular Images, and their Systems]. *Avtomatizatsiya protsessov upravleniya*, 2022, no. 4, pp. 67–75.
4. Tokarev D. A., Pogreban S. V. Anomalii trafika v IP setyakh i ikh obnaruzhenie [Traffic anomalies in IP networks and their Detection]. *Izvestiya instituta inzhenernoy fiziki*, 2013, no. 4, pp. 13–37.
5. Tokmachev D. A., Poletaev A. S., Bezrukin A. G., Chensky A. G., Zasenkov V. E., Gubin N. A. Sistema sinkhronizatsii sverkhlinnovolnovogo interferometra [Synchronization system of an extremely long-wave interferometer]. *Pribory i tekhnika eksperimenta*, 2014, no. 5, pp. 77–84.
6. Gavrilov B. G., Ermak V. M., Lyakhov A. N., Poklad Yu. V., Rybakov V. A., Ryakhovsky I. A. Vosstanovlenie parametrov sredneshirotnoi nizhnei ionosfery pri solnechnykh vspyshkakh M- i X-klassov [Restoration of parameters of the mid-latitude lower ionosphere during M- and X-class solar flares]. *Geomagnetism and aeronomy*. 2020, vol. 60, no. 6, pp. 747–753.
7. Poletaev A. S., Chensky A. G., Tokmachev D. A. Registratsiya variatsii amplitudy i fazy signalov navigatsionnykh SDV peredatchikov [Registration of amplitude and phase variations of signals from navigational VLF transmitters]. *XIV konferentsiya molodykh uchenykh "Vzaimodeistvie polei i izlucheniya s veshchestvom"*, Irkutsk, September 14–18, 2015, pp. 252–255.
8. Yadrennikova O. V., Poletaev A. S., Chensky A. G. Influence of Solar Flares on VLF Radio Waves Propagation over JJI – Irkutsk Path. *XIV International scientific-technical conference on Actual Problems of Electronic Instrument Engineering (APEIE – 2018)*, Novosibirsk, October 2 – 6, 2018, vol. 1, part 3, pp. 137 – 140.
9. Belenkiy M. I., Orlov A. B., Pronin A. E., Uvarov A. N. O zavisimosti velichiny VFA SDV ot zenitnogo ugla Solntsa na protyazhennykh trasakh v razlichnykh shirotakh [On the dependence of VLF signal amplitude variations on the solar zenith angle over extended paths in different latitudes]. *Materials of the VIII Region. Conf. on Radio Wave Propagation*, St. Petersburg, St. Petersburg State University, October 29–30, 2002.
10. Zhanaeva S. B. K voprosu o podgotovke dannykh pri razrabotke modeli neyronnoy seti [Data Preparation for a Neural Network Model]. *Vestnik SibGUTI*, 2022, vol. 16, no. 4, pp. 69–79.

11. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics, 2nd Edition, 2017, 764 p.
12. *Gras D.* Data Science. Nauka o dannykh s nulya, SPb: BKhV-Peterburg, 2017, 336 p.
13. *Lvovich I. Ya., Semenov S. N., Gladskikh N.A., Pronin S. S.* Metodika prognozirovaniya mediko-demograficheskikh pokazateley na baze metodov ehksponentsial'nogo sglazhivaniya i mnogochlennoy regressii [Methodology for Forecasting Medico-Demographic Indicators Based on Exponential Smoothing Methods and Multiple Regression]. *Nauchno-meditsinskiy vestnik tsentral'nogo Chernozemya*, 2007, no. 27, pp. 104–108.
14. STCE Newsletter for 4 Jan 2016 – 10 Jan 2016, available at: <http://www.stce.be/newsletter/pdf/2016/STCEnews20160115.pdf> (accessed 11.03.2024).
15. *Chong Q., Zhiguo Z. et al.* Predictive anomaly detection for marine diesel engine based on echo state network and autoencoder. *Energy Reports*, 2022. vol. 8, pp. 998–1003.
16. *Juan Terven, Wang H. et al.* Anomaly detection for hydropower turbine unit based on variational modal decomposition and deep autoencoder. *Energy Reports*, 2021, vol. 7, pp. 938–946.
17. *Terven J., Cordova-Esparza D. M., Ramirez-Pedraza A., Chavez-Urbiola E. A., Romero-Gonzalez J. A.* Loss Functions and Metrics in Deep Learning. Available at: <https://arxiv.org/abs/2307.02694>. (accessed: 28.09.2023).
18. *Diederik P. Kingma, Jimmy Lei Ba.* Adam: a method for stochastic optimization. *International Conference on Learning Representations*, Toulon, April 24–26, 2015.

Khac Hoang Duong Nguyen

postgraduate student of the Department of Radio Electronics and Telecommunication Systems, Irkutsk National Research Technical University (INRTU, Russia, 664074, Irkutsk, Lermontov St., 83), phone: +7 961 1576 679, e-mail: hd04011996@gmail.com, ORCID ID: 0009-0000-1561-8741.

Aleksandr S. Poletaev

Cand. Sci. (Phys.–Math.), associate professor of the Department of Radio Electronics and Telecommunication Systems, Irkutsk National Research Technical University (INRTU, Russia, 664074, Irkutsk, Lermontov St., 83), phone: +7 395 2405 129, e-mail: poletaevas@ex.istu.edu, ORCID ID: 0000-0002-6448-0045.