

УДК 51.71

## РАЗРАБОТКА АЛГОРИТМА ГЕНЕРАЦИИ УСЛОВНО-РЕАЛЬНЫХ ДАННЫХ ДЛЯ АДАПТАЦИИ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ К МАЛОМУ НАБОРУ НАТУРНЫХ ДАННЫХ

Мусаев А.И. (СПбГЭТУ «ЛЭТИ»)

Научный руководитель – ассистент кафедры МО ЭВМ, Мандрикова Б.С.  
(СПбГЭТУ «ЛЭТИ»)

**Введение.** Современные исследования космической погоды и солнечной активности, ввиду их негативного воздействия на наземную и космическую технику и технологии, требуют наличия представительных выборок данных геофизических параметров, что соблюдается далеко не всегда [1]. Доступ к натурным данным может быть ограничен из-за высокой стоимости и трудности сбора [2]. Кроме того, такие данные часто содержат шумы и пропуски, что снижает их количество и качество для исследований [3]. В связи с этим, генерация условно-реальных данных становится важным инструментом для моделирования и прогнозирования влияния солнечной активности на земные и космические технологии. Основной задачей данной работы является создание алгоритма генерации условно-реальных временных рядов. Существуют решения, которые позволяют создавать синтетические данные временных рядов, однако, они не способны генерировать большой объем данных или являются недостаточно гибкими для генерации на основе имеющихся данных.

**Основная часть.** Для адаптации моделей машинного обучения, в частности нейронных сетей, с учетом ограниченности количества натуральных данных, и как следствие невозможности перекрытия пространства признаков, разработан алгоритм генерации условно-реальных данных.

На основе обзора существующих методов сформулированы ключевые требования к разрабатываемому алгоритму генерации условно-реальных данных:

- Поддержка генерации длинных временных рядов (свыше 10 000 измерений);
- Возможность задания характеристик сгенерированных данных на основе существующих временных рядов;
- Сохранение исходного тренда и возможность добавления аномалий в этот тренд;

На основе обзора и анализа существующих аналогов предлагается следующий алгоритм:

- 1) Создание базового временного ряда: формирование исходного ряда на основе медианных значений натуральных данных за периоды, соответствующие регулярным (спокойным) вариациям.
- 2) Выделение тренда: применение методов анализа временных для выделения основной тенденции. В работе на данном этапе предлагается применение кратномасштабного вейвлет-анализа путём последовательного исключения колебаний с меньшими периодами из исходного ряда.
- 3) Генерация аномалий: моделирование локальных особенностей и их аддитивное добавление в тренд. Локальные особенности должны быть подобны природным аномалиям. На данном этапе формируются форма (например, треугольный импульс), амплитуда, длительность и параметры (например, коэффициент асимметрии) локальных особенностей в зависимости от существующих аномалий.
- 4) Генерация шумов: применение различных типов шумов для имитации реальных условий измерений. На данном этапе задается тип шума (например, белый) и отношение сигнал/шум.

В качестве объекта исследования выбраны данные нейтронных мониторов наземных станций [4], отражающие интенсивность космических лучей. Однако, разработанный алгоритм универсален и не привязан к конкретному типу данных, поэтому может быть использован специалистами из других областей знаний. Массивы данных, сгенерированных на основе предложенного алгоритма, могут быть использованы как при адаптации моделей

машинного обучения, в метаобучении, так и для детального исследования и оценки чувствительности построенных моделей.

**Выводы.** Проведенный обзор существующих методов генерации условно-реальных данных показал, что на данный момент универсальных методов не существует, а проблема отсутствия представительных выборок данных геофизических параметров является актуальной. В работе предложен новый алгоритм генерации условно-реальных данных, позволяющий адаптировать модели машинного обучения, в частности нейронные сети (для задачи метаобучения), так и для детального исследования и оценки чувствительности построенных моделей. В дальнейшем планируется создание программы в среде программирования MatLab на основе разработанного алгоритма.

#### **Список использованных источников:**

1. Кузнецов В. Д. Космическая погода и риски космической деятельности // Космическая техника и технологии. 2014; 3(6).

2. Бухарицин А.П. Состояние и перспективы развития рынка услуг по сбору и обработке спутниковых данных дистанционного зондирования земли // Международный журнал прикладных и фундаментальных исследований. – 2021. – № 3. – С. 85-91;

3. Балабин Юрий Васильевич, Вашенюк Эдуард Владимирович, Гвоздевский Борис Борисович Шал и множественность на нейтронных мониторах // Вестник Кольского научного центра РАН. 2010. №2.

4. База данных сети нейтронных мониторов. Источник: <[www.nmdb.eu](http://www.nmdb.eu)>.

Автор \_\_\_\_\_ Мусаев А.И.

Научный руководитель \_\_\_\_\_ Мандрикова Б.С.