

Разработка метода и алгоритма генерации условно-реальных данных нейтронных мониторов.

Мусаев Артур Илдарович
каф. МО ЭВМ
СПбГЭТУ «ЛЭТИ»
Санкт-Петербург, Россия
musaew97@gmail.com

Аннотация — Современные исследования космической погоды и солнечной активности требуют доступа к данным, предоставляемым нейтронными мониторами. Однако доступ к реальным данным может быть ограничен из-за высокой стоимости и трудности сбора. В связи с этим генерация условно-реальных данных становится важным инструментом для моделирования и прогнозирования влияния солнечной активности на земные и космические технологии. В данной статье рассмотрен обзор существующих приложений для генерации условно-реальных данных временных рядов. В обзоре представлены методы и библиотеки, которые можно использовать для генерации данных. Аналогии сравнивались по трем критериям: гибкость настройки генерации, поддержка трендов и аномалий, масштабируемость. На основе сравнения аналогов было предложено собственное приложение для генерации условно-реальных временных рядов.

Ключевые слова — условно-реальные данные; синтетические данные; временные ряды.

ВВЕДЕНИЕ

Нейтронные мониторы используются для изучения космической погоды, солнечной активности, магнитных бурь и их воздействие на земные и космические технологические системы [1, 2]. Доступ к реальным данным нейтронных мониторов может быть ограничен из-за высокой стоимости сбора или недостаточного объема данных для конкретных исследований [3]. Кроме того, данные часто содержат шумы и сбои, что снижает их количество и качество для исследований [4]. Генерация условно-реальных данных нейтронных мониторов позволяет решить проблему сбора данных.

Целью работы является обзор существующих решений для генерации временных рядов и разработка алгоритма генерации временных рядов. Для достижения цели были поставлены следующие задачи:

- Провести обзор существующих методов и приложений для генерации синтетических временных рядов;
- Предложить на основе анализа алгоритм для генерации;

I. ОТБОР СУЩЕСТВУЮЩИХ ПРИЛОЖЕНИЙ ДЛЯ ГЕНЕРАЦИИ ВРЕМЕННЫХ РЯДОВ

Для анализа были отобраны приложения, реализующие методы генерации условно-реальных временных рядов. Для поиска аналогов был задействован ресурс Google Scholar с такими запросами, как “генерация временных рядов”, “моделирование временных рядов”,

“time series generation”. Таким образом среди отобранных аналогов были методы генерации временных рядов, методы предсказания временных рядов и библиотеки для генерации данных. Среди них методы: Visualising forecasting algorithm performance using time series[5], Correlated synthetic time series generation for energy system simulations using Fourier and ARMA signal processing[6], GRATIS[7], а также библиотеки: timeseries-generator[8] и mockseries[9].

A. Visualising forecasting algorithm performance using time series

В данной статье предлагается ввести для временного ряда некоторые метрики. После подсчета метрик для всех рядов создается датасет, где каждому временному ряду соответствует несколько столбцов его метрик. У данного датасета понижают размерность до 2, после чего у каждого ряда есть его “координаты”. Далее выбирается ряд, для которого нужно создать схожие временные ряды, создается большое количество случайных временных рядов и с помощью генетического алгоритма создаются схожие по метрикам временные ряды, в качестве метрики для генетического алгоритма используется евклидово расстояние до изначального ряда.

Недостаток данного алгоритма заключается в том, что он может генерировать лишь короткие временные ряды, так как при генерации больших временных рядов будет затратно считать для каждого метрики и, если генерировать их с одинаковым распределением, при увеличении рядов они будут становиться по метрикам всё более похожими друг на друга.

B. Correlated synthetic time series generation for energy system simulations using Fourier and Arma signal processing

В данной статье предлагается сначала выделить все сезонные составляющие с помощью преобразования Фурье, затем для оставшегося шума построить линейную модель, которая будет генерировать данные. После генерации предлагается вернуть все сезонные составляющие выделенные ранее.

Проблема данной модели в том, что она совсем не учитывает тренд. Кроме того, при генерации данных она будет рекурсивно возвращаться к созданными ею же данными. Ещё будет проблема в том, что шум - это плохие данные для предсказания с помощью регрессии, так как в них маловероятно будут присутствовать закономерности.

C. GRATIS

Данный генератор создаёт регрессионные модели, состоящие из нескольких компонент. Он похож на

предыдущий, но для каждой сезонности тоже присваивает свой вес, то есть после выделения сезонов каждому присваивается свой вес и на основе этого строится модель регрессии. Сначала внутри каждого сезона наблюдения присваивается вес, затем каждому сезону присваивается вес.

$$E(x_t|F_{-t}) = \sum_{k=1}^K \alpha_k (\phi_{k0} + \phi_{k1}x_{t-1} + \dots + \phi_{kp_k}x_{t-p_k})$$

где α_k – вес сезона k , ϕ_{ki} – веса каждого наблюдение для сезона k .

Этой модели свойственны те же проблемы, что и предыдущей, а именно не учитывается тренд и рекурсивный возврат к сгенерированным ею же данными.

D. *timeseries-generator*

В этом пакете предполагается, что временной ряд состоит из базового значения, умноженного на множество коэффициентов.

$$y_t = x_t * factor_1 * factor_2 * \dots * factor_n + Noiser$$

Этими факторами могут быть что угодно: случайный шум, линейные тенденции или сезонность. Факторы могут влиять на разные характеристики. Например, некоторые характеристики вашего временного ряда могут иметь сезонную составляющую, а другие — нет.

Встроенные факторы:

- **LinearTrend:** задает линейный тренд на основе входных данных
- **CountryYearlyTrend:** коэффициент рыночной капитализации на основе ВВП на душу населения за год.
- **EUEcoTrendComponents:** предоставляет ежемесячный коэффициент изменения на основе общедоступных данных о продукции промышленности ЕС
- **HolidayTrendComponents:** имитирует пик распродаж в праздничные дни. Он адаптирует праздничные дни по-разному в разных странах
- **BlackFridaySaleComponents:** имитация распродажи черная пятница
- **WeekendTrendComponents:** больше продаж в выходные, чем в будние дни
- **ProductSeasonTrendComponents:** имитирует продажи товаров, чувствительных к сезонности.

Недостатки данного пакета в том, что он больше настроен для задач маркетинга, так как создавался для продуктовой компанией. Кроме того, в нём нельзя построить зависимости от входного временного ряда, а только настроить все показатели самостоятельно.

E. *mockseries*

Это Python-пакет, который предлагает более широкие возможности для генерации временных рядов по сравнению с *timeseries-generator*. Этот инструмент позволяет создавать временные ряды как аддитивные и мультипликативные комбинации различных сигналов (таких как тренд, сезонность, шум и т.п.). Пример представления временного ряда:

$$y_t = (S_1(t)T_1(t) + S_2(t))u,$$

где $S_1(t)$ и $S_2(t)$ – сезонные компоненты, $T_1(t)$ – линейный тренд, u – шум.

Mockseries позволяет создавать временные ряды с динамическими характеристиками, которые изменяются со временем. Тем не менее, вопрос настройки параметров процессов на основе исходных данных всё ещё не решён.

II. КРИТЕРИИ СРАВНЕНИЯ АНАЛОГОВ

A. *Гибкость настройки генерации*

Данный критерий показывает, как мы можем настраивать параметры при генерации временных рядов.

B. *Поддержка трендов и аномалий*

Этот критерий отвечает, учитывает ли данный метод генерации тренд и можно, ли с помощью него добавить аномалии в генерируемый ряд.

C. *Масштабируемость*

С помощью этого критерия можно понять, насколько большие временные ряды мы можем генерировать и как будет падать качество генерации при увеличении длины ряды.

III. ТАБЛИЦА СРАВНЕНИЯ АНАЛОГОВ

TABLE I. TABLE TYPE STYLES

Метод	Критерии сравнения		
	<i>Гибкость настройки генерации</i>	<i>Поддержка трендов и аномалий</i>	<i>Масштабируемость</i>
A	Этот метод может генерировать временные ряды, на основе на загруженного файла. При генерации множества рядов можно менять параметры их генерации, меняя случайное распределение и длину полученных рядов. Кроме того, при понижении размерности можно выбирать какая метрика важнее, чтобы не терять корреляцию с ней.	Данный метод не поддерживает задание тренда, так как он генерирует случайные ряды. Кроме того, он генерирует лишь короткие ряды, на которых плохо будет проследиваться тренд. Аномалии же он генерирует лучше, потому что при случайной генерации аномалии в любом случае создадутся.	Данный метод не приспособлен к масштабируемости, потому что генерирует лишь короткие временные ряды в силу своего алгоритма.
B	Данный метод поддерживает генерацию временного ряда поданного из файла. Для ряда Фурье можно задать частоту и количество гармоник (еженедельные, ежемесячные и т.д.). Также	Данная модель никак не обрабатывает наличие трендов.	Данная модель может генерировать большие временные ряды, но при их генерации она будет рекурсивно возвращаться к созданным ею же данными, что может

Метод	Критерии сравнения		
	Гибкость настройки генерации	Поддержка трендов и аномалий	Масштабируемость
	самостоятельно можно выбрать модель регрессии, которая будет использоваться для предсказания шума.		повлечь проблемы.
C	Этот метод как и предыдущий поддерживает задание временного ряда из файла.	Наличие трендов не поддерживается.	Модель способна генерировать длинные ряды, но будет та же проблемы с рекурсивной генерацией.
D	У данного приложения большая гибкость параметризации, но она не поддерживает создание временного ряда, похожего на данный из файла.	Задание трендов присутствует. Аномалии – тоже, но по большей части это маркетинговые аномалии, как черная пятница.	Данное приложение способно генерировать большие временные ряды.
E	Тут так же присутствует большая гибкость настройки параметров, но отсутствует возможность подать собственный временной ряд.	Тренды и аномалии поддерживаются. Для задания аномалий имеются методы switch и transition. Метод switch позволяет мгновенно изменять значения временного ряда, моделируя сбои или аномалии. Метод transition предназначен для постепенного изменения значений временного ряда на заданном интервале времени в соответствии с определённой функцией.	Данное приложение способно генерировать большие временные ряды.

* так как методы имеют длинные названия данных их номера из п.1

Исходя из таблицы, можно сделать вывод, что данные приложения не подходят под генерацию больших (больше 10 000 измерений) синтетических временных рядов, заданных по существующим, так как методы А, В, С не приспособлены для задания параметров и генерации большого количества данных, а приложения D, E не приспособлены для генерации временных рядов на примере существующих.

IV. ОПИСАНИЕ МЕТОДА РЕШЕНИЯ

На основе анализа предлагается следующий метод генерации временных рядов.



1) Создаётся временной ряд, каждый отсчёт которого равен медианному значению природных данных, рассчитанному за спокойные периоды. Например, из матрицы радиационной активности за 1000 дней рассчитываются медианы для каждого часа, формируя массив из 24 значений.

2) На основе кратномасштабного анализа или другого метода[10] тренд извлекается из временного ряда, создавая сглаженную кривую.

3) В тренд аддитивно вносятся локальные особенности, такие как треугольные импульсы, импульсы по Гауссу или природные аномалии.



Рисунок 1 – треугольный импульс

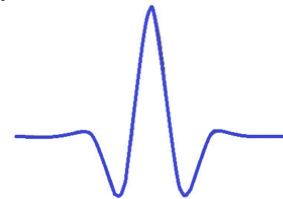


Рисунок 2- импульс, моделированный по Гауссу

4) В сигнал добавляется шум (например, белый или коррелированный), чтобы приблизить модель к реальным данным. Шум будет добавляться в специальном соотношении с сигналом по формуле:

$$SNR = \frac{A_{\text{аном}}^2}{A_{\text{шум}}^2} = \frac{P_{\text{аном}}}{P_{\text{шум}}},$$

где P – средняя мощность сигнала, A – среднеквадратичное значение амплитуды. Данное соотношение введено, чтобы показать отношения аномалии к шуму. Если SNR высокий, то аномалия заметно выделяется на фоне шума. Если SNR низкий,

аномалия "погружается" в шум, и её становится сложно распознать. Предполагается возможность выбора вида шума для добавления.

Предложенный метод генерации позволит создать большое количество данных временных рядов. При этом будут учтены тренд, создание аномалий и генерация временного ряда на основе существующего ряда.

V. ЗАКЛЮЧЕНИЕ

В результате выполнения работы был проведен обзор алгоритмов генерации условно-реальных временных рядов. Были найдены 3 метода генерации и 2 готовых библиотеки. Были сформированы критерии для алгоритмов и проведено сравнение.

В результате сравнения было выяснено, что существующие методы не приспособлены для генерации больших временных рядов и не поддерживают тренды и аномалии, а библиотеки не предполагают генерацию временного ряда с помощью существующего ряда.

Был предложен алгоритм генерации на основе выделения тренда, добавления аномалий и шума.

ЛИТЕРАТУРА

- [1] Пилипенко В.А. Воздействие космической погоды на наземные технологические системы. Солнечно-земная физика. 2021. Т. 7, № 3. С. 72–110. DOI: 10.12737/szf-73202106.
- [2] Кузнецов Владимир Дмитриевич КОСМИЧЕСКАЯ ПОГОДА И РИСКИ КОСМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ // Космическая техника и технологии. 2014. №3 (6).
- [3] 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;
- [4] Балабин Юрий Васильевич, Вашенок Эдуард Владимирович, Гвоздевский Борис Борисович Шал и множественность на нейронных мониторах // Вестник Кольского научного центра РАН. 2010. №2.
- [5] Kang, Yanfei \& Hyndman, Rob \& Smith-Miles, Kate. (2017). Visualising forecasting algorithm performance using time series instance spaces. International Journal of Forecasting. 33. 345-358. 10.1016/j.ijforecast.2016.09.004.
- [6] Talbot, Paul \& Rabiti, Cristian \& Alfonsi, Andrea \& Krome, Cameron \& Kunz, M. \& Epiney, Aaron \& Wang, Congjian \& Mandelli, Diego. (2020). Correlated synthetic time series generation for energy system simulations using Fourier and ARMA signal processing. International Journal of Energy Research. 44. 10.1002/er.5115.
- [7] Kang, Yanfei \& Hyndman, Rob \& Li, Feng. (2019). GRATIS: GeneRAting TIme Series with diverse and controllable characteristics. 10.48550/arXiv.1903.02787.
- [8] "timeseries-generator", Available at: <https://github.com/Nike-Inc/timeseries-generator>, accessed 08.01.2025.
- [9] "mockseries documentation.", Available at: <https://mockseries.catheu.tech/docs/tutorials/interactionscalaroperations>, accessed 08.01.2025.
- [10] Губанов Вячеслав Анатольевич Выделение тренда из временных рядов макроэкономических показателей // Научные труды: Институт народнохозяйственного прогнозирования РАН. 2005. №3. URL: <https://cyberleninka.ru/article/n/vydelenie-trenda-iz-vremennyh-ryadov-makroekonomicheskikh-pokazateley> (дата обращения: 16.01.2025).