

УДК 69.003

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ДЕТРЕНДИРОВАНИЯ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ДАННЫХ МОНИТОРИНГА СТРОИТЕЛЬНЫХ ПРОЦЕССОВ

Павел Борисович Каган

Доцент, к.т.н., доцент кафедры информационных систем, технологий и автоматизации в строительстве Московского государственного строительного университета (НИУ МГСУ), г. Москва, Россия; e-mail: kagan@mgsu.ru

Паршин Дмитрий Андреевич

Аспирант кафедры информационных систем, технологий и автоматизации в строительстве Московского государственного строительного университета (НИУ МГСУ), г. Москва, Россия; e-mail: parshin.galaxy2@gmail.com

Аннотация: Обработка данных систем мониторинга различных процессов на этапе строительства и эксплуатации зданий требует разработки специальных средств, которые относятся к области искусственного интеллекта. Метод удаления тренда – один из способов предобработки данных, собранных с различных датчиков и IoT-устройств, контролирующих состояние зданий, сооружений и грунтовых массивов в период строительства и эксплуатации. В данной статье анализируется, как различные подходы к удалению тренда из временных рядов влияют на производительность и точность алгоритмов моделей вычислительного интеллекта CI. В ходе анализа сравниваются три подхода: линейное детрендирование, нелинейное детрендирование и дифференцирование первого порядка.

В качестве моделей CI используются пять репрезентативных методов: динамическая развивающаяся нечеткая нейронная сеть DENFIS, гауссовский процесс GP, многослойный перцептрон MLP, оптимально обрезанная экстремальная обучающая машина OP-ELM и машина опорных векторов SVM. Из экспериментов, проведенных на четырех наборах данных, сделаны три основных вывода: 1) удаление тренда не улучшает производительность в целом, 2) метод на основе декомпозиции на эмпирические моды обеспечивает лучшую производительность, чем линейное детрендирование, и 3) дифференцирование первого порядка в некоторых случаях может быть эффективным, а в некоторых контрпродуктивным для рядов с общими паттернами.

Ключевые слова: цифровые технологии, мониторинг строительных процессов, искусственный интеллект, предобработка данных, детрендирование

**USING DETRENDING METHODS FOR INTELLIGENT PROCESSING
OF CONSTRUCTION PROCESS MONITORING DATA**

Kagan Pavel Borisovich

Docent, Candidate of Technical Sciences, Docent Department of information systems technology and automation in construction of Moscow State University of Civil Engineering (MSUCE), Moscow, Russia; e-mail: kagan@mgsu.ru

Parshin Dmitry Andreevich

Postgraduate Student, Department of information systems technology and automation in construction of Moscow State University of Civil Engineering (MSUCE), Moscow, Russia; e-mail: parshin.galaxy2@gmail.com

Abstract: Data processing of monitoring systems of various processes at the stage of construction and operation of buildings requires the development of special tools that belong to the field of artificial intelligence. The trend removal method is one of the ways to preprocess data collected from various sensors and IoT devices that monitor the state of buildings, structures and soil masses during construction and operation. This article analyzes how different approaches to detrending time series affect the performance and accuracy of algorithms for CI computational intelligence models. The analysis compares three approaches: linear detrending, non-linear detrending, and first-order differentiation.

Five representative methods are used as CI models: DENFIS dynamic evolving fuzzy neural network, GP Gaussian process, MLP multilayer perceptron, OP-ELM optimally trimmed extremal learning machine, and SVM support vector machine. There are three main conclusions from the experiments performed on the four datasets: 1) detrending does not improve overall performance, 2) the empirical mode decomposition method provides better performance than linear detrending, and 3) first-order differentiation in some cases can be effective and in some cases counterproductive for series with common patterns.

Keywords: digital technologies, monitoring of construction processes, artificial intelligence, data preprocessing, detrending

Актуальность работы

Важным направлением развития мониторинга различных строительных процессов, как на этапе строительства, так и на этапе эксплуатации является цифровизация обработки данных [36], полученных в результате работы различных устройств и датчиков систем мониторинга.

Аналитика больших данных, собираемых с различных сенсоров и датчиков, требует больших вычислительных мощностей. Для повышения точности вычислительных моделей и уменьшения времени работы, которое

требуется для анализа получаемой информации с различных IoT-устройств, применяются методы предобработки данных.

Основа интеллектуальных систем, использующих данные устройств Интернета вещей, - анализ данных в режиме реального времени. Такие данные имеют временную метку и представляют собой временной ряд. Примерами таких данных могут служить информация, поступающая от различных датчиков (воды, дыма, температуры и т.д.).

Интеллектуальная система проводит мониторинг данных со всех устройств в режиме реального времени и моделирует поведение временных рядов в будущем. Например, оповещение о пожаре в помещении и его предотвращение или прогнозирование. Важное свойство временного ряда для анализа и моделирования его поведения интеллектуальной системой – стационарность. Большинство алгоритмов анализа работают для стационарных рядов.

Стационарность модулируемого набора данных предполагается для классических статистических методов, использующихся для исследования временных рядов, и для методов регрессионного анализа, основанных на нейронных сетях и вычислительном интеллекте [1].

Однако в настоящих временных рядах присутствует динамическая или статистическая нестационарность, переходный или прерывистый хаос, псевдослучайность или суперпозиция паттернов. Иными словами, временные ряды часто демонстрируют нелинейное поведение. При использовании классических статистических методов для анализа временных рядов нелинейность приводит к неточности результата. Как следствие, для повышения производительности методов моделирования и получения более точных результатов проводится удаление трендов (детрендинг) и сезонных отклонений (десезонализация) [2, 3].

На самом деле, обнаружение и моделирование трендов — одна из основных задач анализа и прогнозирования временных рядов. Основные причины, для которых применяется задача поиска трендов:

- превращение нестационарного временного ряда в стационарный;
- описание поведения нестационарного временного ряда путем разложения на такие компоненты, как тренд, циклы, флуктуации и шум.

Этапы предварительной обработки рядов (например, детрендинг) могут влиять на результаты методов анализа и прогнозирования, поэтому проблема поиска трендов во временных рядах по-прежнему содержит множество фундаментальных вопросов [4].

Анализ воздействия методов детрендинга на модели вычислительного интеллекта смог бы ускорить разработку надежных методологий, основанных на искусственном интеллекте, для временных рядов.

Несмотря на сильное влияние, которое может оказывать детрендинг на анализ временных рядов, в литературе описано мало экспериментальных и теоретических исследований на эту тему. Доступные результаты зачастую разрозненны или противоречивы, а иногда применимы лишь к узким предметным областям. Большая часть таких исследований основаны на синтетических или ограниченных данных реального мира [5, 6].

В данной статье анализируется влияние различных подходов детрендинга временных рядов на производительность моделей СИ. Статья организована следующим образом: с начала детрендинг рассматривается в общем контексте этапов предварительной обработки временных рядов. Затем дается определение тренда, а также описываются методы извлечения тренда, используемые в данном исследовании. В последних разделах описаны экспериментальные результаты.

Детрендинг — процесс, с помощью которого производится удаление тренда из временного ряда. Самый простой метод детрендинга заключается в подборе детерминированной компоненты (как правило, линейной) и вычитании этой компоненты из всех значений ряда. Остаток после этой операции является флуктуацией. Одно из главных ограничений для применения такого подхода заключается в том, что детрендинговый ряд

должен быть числовым рядом с нулевым средним значением для рассматриваемого периода времени [7].

В целом, детрендинг является лишь одним из возможных этапов предварительной обработки временного ряда перед построением модели СІ.

Другими этапами могут быть также:

- заполнение пропущенных значений,
- обработка выбросов,
- шумоподавление,
- фильтрация, нелинейные преобразования для улучшения гладкости,
- выбор переменных,
- вычисление динамических и статистических характеристик, таких как остаточная дисперсия и временные лаги.

Во многих случаях разделение этапов предварительной обработки и непосредственно анализа данных неясно. Например, можно совместно выполнять задачи, связанные как с выбором модели, так и с обработкой выбросов.

Более того, разница между фильтрацией, шумоподавлением и детрендингом может быть неочевидной. Один и тот же метод может использоваться для выполнения нескольких из перечисленных задач.

В этой статье подходы детрендинга являются единственным этапом предварительной обработки (помимо нормализации значений). При таком подходе временной ряд может быть разложен следующим образом:

$$y(t) = T(t) + r(t), \quad (1)$$

где $r(t)$ - остаточная величина или флуктуация вокруг тренда $T(t)$. Компонент флуктуации $r(t)$ может иметь как стохастические, так и детерминированные компоненты.

В некоторых областях очень распространено разделение временного ряда на три компонента: помимо тренда и флуктуации добавляется еще сезонный компонент. Для простоты в данной статье анализ ограничен

описанным выше подходом. Анализ влияния методов десезонализации оставлен для дальнейшей работы [8].

Методы извлечения тренда

На данный момент нет единого мнения о том, как следует моделировать тренды, хотя проводился ряд сравнительных исследований. Кроме того, несмотря на широкое распространение и использование термина, понятие тренда не имеет одного формального определения. В итоге, существует большое разнообразие методов извлечения тренда, и сложно сделать выводы, какие из этих методов являются наилучшими.

Согласно одному из определений, тренд — это гладкий аддитивный компонент, который содержит информацию о глобальных изменениях. В этом определении упор делается на такие ключевые аспекты, как гладкость, аддитивность и информация [9].

С точки зрения других ключевых аспектов: частоты и времени, тренд можно определить, как остаток после удаления высокочастотных компонентов выше определенного порога. Такое определение используется в методах, основанных на вейвлетах, кривых или анализе сингулярного спектра, и непараметрических методах фильтрации, таких как фильтр Ходрика-Прескотта или более поздних фильтрах для извлечения кусочно-линейных трендов на основе L1-минимизации [10, 11, 12].

Часть подходов по выявлению тренда основываются на свойстве регрессии. Это, например, модели ARIMA. Аналогичным образом, подходы к перемещению средних основаны на заранее заданной временной шкале для вычисления средних значений [13, 14].

Таким образом, процесс детрендрования зависит от используемых методов, их параметров и определения, лежащего в основе. Например, при использовании вейвлет-анализа выбор типа вейвлета может оказать определяющее влияние на результат процесса детрендрования [15].

В данной статье рассматриваются 3 наиболее распространенных беспараметрических автоматических метода общей применимости, которые находят четко определенный тренд.

а) Линейное детрендрование

Линейное детрендрование является наиболее распространенным методом поиска тренда. Этот способ заключается в построении прямой линии, соответствующей основному поведению временного ряда.

Процесс детрендрования представляет собой вычитание прямолинейного тренда $T(t) = a_0 + b_0 t$ из временного ряда и получение остатка с нулевым средним значением. Этот линейный подход к детрендрованию является частным случаем полиномиального детрендрования: полиномиальное детрендрование первого порядка [16].

Реализация, используемая в следующих разделах этой статьи, основана на подгонке по методу наименьших квадратов, выполняемой QR-разложением временной области. Для улучшения численных характеристик подбора оно применяется при следующем преобразовании временной области [17]:

$$\hat{t} = \frac{t - \mu_1}{\mu_2}, \quad (2)$$

где μ_1 и μ_2 - среднее значение и среднеквадратическое отклонение t соответственно. При детрендровании с полиномиальной подгонкой более высокого порядка могут возникать проблемы с переобучением.

б) Детрендрование методом EMD

Зачастую динамика временного ряда неизвестна, а его данные нестационарны. В таком случае можно использовать декомпозицию на эмпирические моды (EMD) [18].

Тренд в этом подходе — детерминированная монотонная функция или функция в пределах определенного временного интервала, где может быть не более одного экстремума [19].

EMD выполняется с помощью итеративного процесса, известного как просеивание. Алгоритм просеивания [20]:

1. определить все локальные экстремумы $y(t)$.
2. построить верхнюю и нижнюю огибающую, $y_{up}(t)$ и $y_{low}(t)$ соответственно, соединив все локальные экстремумы (максимумы и минимумы соответственно) с помощью кубического сплайна [21].
3. вычислить среднее значение огибающей $m(t)$ точка за точкой.
4. извлечь сигнал детализации, $d(t) = y(t) - m(t)$.
5. проверить свойства сигнала детализации: удовлетворяет этот сигнал условиям остановки или нет.
6. повторить шаги 1-5 до тех пор, пока остаток не удовлетворит определенному условию остановки.

В конце процесса просеивания исходный временной ряд $y(t)$ раскладывается следующим образом:

$$y(t) = \sum_{j=1}^p c_j(t) + r_p(t), \quad (3)$$

где $r_p(t)$ - конечный остаток, а тренд $c_j(t)$ - функции эмпирической моды, а p - количество мод.

Затем, применяя преобразование Гильберта к каждой моде, $y(t)$ может быть выражено в частотно-временных областях следующим образом [22]:

$$Y(t, \omega) = RP \left(\sum_{j=1}^n a_j(t) e^{i \int_0^t \omega_j(\tau) d\tau} \right), \quad (4)$$

где RP обозначает действительную часть преобразований, $a_j(t)$ - коэффициент расширения, зависящий от времени, а $\omega_j(\tau)$ - частота в момент времени τ . Таким образом, частота и амплитуда каждого компонента являются функциями времени. Это очень похоже на ряд Фурье, который будет иметь следующий вид [23]:

$$Y(t) = RP \sum_{j=1}^{\infty} a_j e^{i\omega_j t}, \quad (5)$$

где ω_j , и a_j являются константами.

В таком случае можно рассматривать эмпирические моды как обобщенное разложение Фурье. Такой подход имеет 2 преимущества:

1. более сжатое расширение;

2. возможность учитывать нелинейное и нестационарное поведение набора данных.

Таким образом, метод EMD позволяет разделить среднюю и флуктуирующую компоненты сигнала (устойчивые компоненты и возмущения, где среднее время возмущений равно нулю).

с) Дифференцирование первого порядка

Дифференцирование ряда — это переход к ряду, состоящему из попарных разностей исходного ряда:

$$y'_t = y_t - y_{t-1} . \quad (6)$$

Такая операция позволяет получить из нестационарного ряда стационарный, стабилизировать среднее значение ряда и избавиться от тренда, а иногда даже от сезонности [24].

Продифференцированный ряд имеет только $N - 1$ значений, так как невозможно вычислить разницу y'_1 для первого наблюдения.

Эксперименты

В последующих экспериментах используется моделирование на временных рядах в реальном времени (как для детрендирования, так и для моделирования флуктуаций). Эксперимент должен дать ответ на вопрос: какие подходы удаления тренда могут быть использованы в общем случае, а какие более эффективны в конкретных условиях.

В данной статье оцениваются три подхода к детрендированию: линейное, основанное на EMD и дифференцирование первого порядка.

CI методы

Для моделирования временных рядов были использованы следующие методы CI: DENFIS, GP, MLP, OP-ELM и SVM. Для DENFIS использовался оффлайн-режим обучения из набора инструментов DENFIS. Модели GP строятся с использованием настроек по умолчанию из набора инструментов GPML. В частности, используется ковариационная функция как сумма квадратичной экспоненциальной функции и независимого шума. Модели OP-ELM строятся с использованием набора инструментов OP-ELM. Для обучения

были выбраны настройки по умолчанию, а именно, все возможные ядра: линейное, сигмоидное и гауссово, и использовались 100 скрытых узлов [25, 26].

В случае отбора моделей MLP и SVM использовалась кросс-валидация с 10% тестовых данных от общей выборки. Для моделей SVM применялась стратегия поиска по сетке для изучения трех параметров в логарифмической шкале в диапазоне [-2, 10]. Регрессия вектора поддержки использовалась с ядрами радиальных базисных функций. Реализация этого метода была взята из библиотеки libSVM [27].

Метод MLP реализован с использованием стандартного набора инструментов нейронных сетей, включенного в библиотеку scikit-learn. Сети оптимизируются с использованием алгоритма Левенберга-Марквардта. Чтобы смягчить эффект локальных минимумов, для каждой подвыборки данных и для каждого размера сети (от 1 до 20 скрытых узлов) выполняется десять проходов, на усредненных результатах которых и выбирается модель, дающая наименьшую ошибку [28, 29].

Наборы данных

Для простоты в этом исследовании использовались одномерные временные ряды. Для исследования были выбраны наборы данных, удовлетворяющие следующим требованиям:

- данные сгенерированы настоящими устройствами, а не являются синтетическими;
- данные содержат как минимум несколько сотен тестовых значений, чтобы ошибки тестирования были статистически значимыми.

Эти наборы данных были выбраны для поиска компромисса между следующими целями:

1. упрощение сравнения с соответствующей литературой,
2. выбор наборов данных для широкого спектра характеристик (переменные, размер, динамическое поведение и т.д.). В частности, некоторые наборы данных представляют собой явно нестационарные процессы.

Все перечисленные ниже наборы данных взяты из открытых источников:

1. *Smart Home Dataset with weather Information*

Набор данных содержит записи показателей бытовой техники от интеллектуального счетчика в кВт за промежутки времени в 1 минуту и погодные условия конкретного региона [30].

2. *Energy consumption*

Ежегодная статистика по потреблению электричества и газа от частных энергетических компаний [31].

3. *Temperature Readings: IoT Devices*

Журнал записей температуры внутри и вне различных комнат жилого здания за один год [32].

4. *Solar Power Generation Data*

Набор данных с показателями собранной энергии от двух солнечных батарей за 34 дня [33].

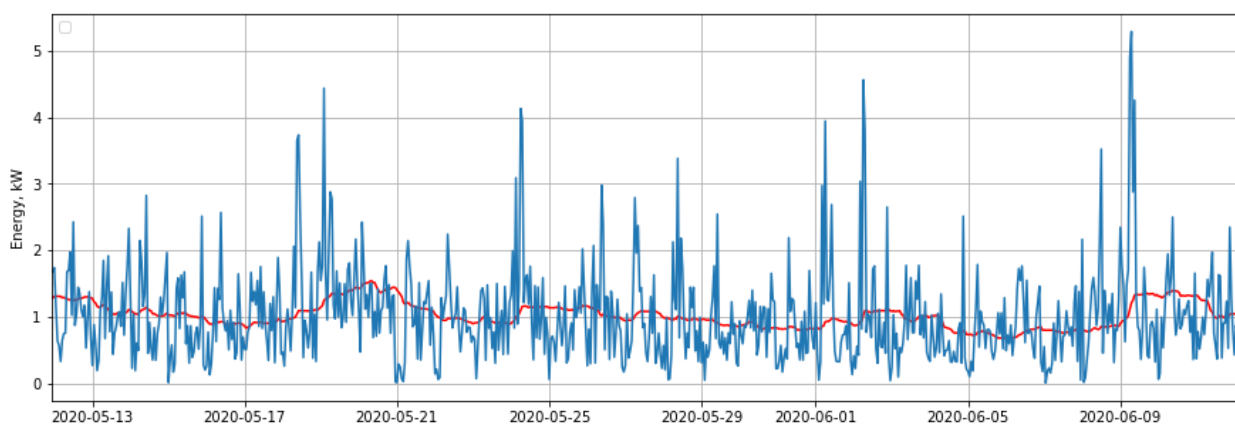


Рис. 1. Временной ряд потребления электроэнергии

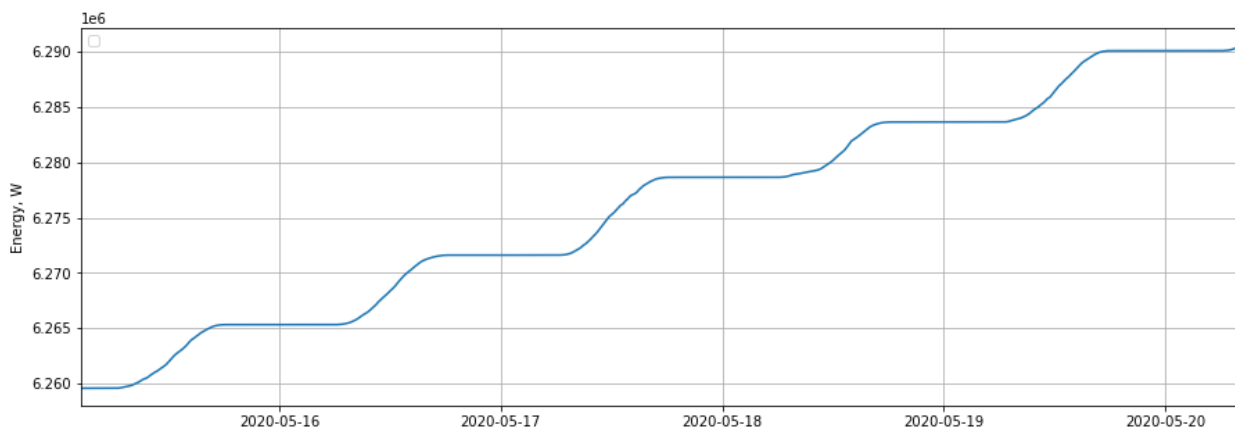


Рис. 2. Временной ряд и скользящая средняя собранной энергии с солнечной батареи

Моделирование и предсказание результатов

Предсказательные модели для детрендрированных временных рядов строятся в соответствии с традиционным подходом обучения моделей [34]:

1. Определяются обучающая и тестовая выборки данных;
2. Строится модель для обучающего набора;
3. Эффективность и точность модели оценивается на тестовом наборе данных.

Среднеквадратические ошибки моделей на тестовых данных, а также время обучения указаны в таблицах 1-4 [35].

Таблица 1. Результаты для набора данных №1

Метод	Детрендрирование	Среднеквадратическая ошибка	Время, с
DENFIS	отсутствует	0.697	1.25
	линейное	0.664	0.94
	emd	0.671	0.74
	дифференцирование	0.744	1.68
MLP	отсутствует	0.716	725
	линейное	0.739	712
	emd	0.723	981
	дифференцирование	0.724	671
OP-ELM	отсутствует	0.699	0.78
	линейное	0.709	0.63
	emd	0.712	0.97
	дифференцирование	0.720	0.71

SVM	отсутствует	0.724	37.4
	линейное	0.722	39.8
	emd	0.721	38.3
	дифференцирование	0.716	25.7
GP	отсутствует	0.701	34.6
	линейное	0.698	29.1
	emd	0.697	47.4
	дифференцирование	0.732	35.8

В таблице 1 представлены результаты вычисления среднеквадратичной ошибки для временных рядов показателей бытовой техники с разными способами детрендрования. Линейное детрендрование обеспечивает небольшое улучшение для DENFIS, SVM и GP. Однако результаты несколько хуже (ошибки на 2-3% выше) для MLP и OP-ELM. EMD показал в половине случаев улучшение, а в другой половине ухудшение качества моделей, если сравнивать с качеством без детрендрования. Дифференцирование приводит к худшим результатам, за небольшим исключением для SVM.

Таблица 2. Результаты для набора данных №2

Метод	Детрендрование	Среднеквадратическая ошибка	Время, с
DENFIS	отсутствует	0.335	6.47
	линейное	0.342	6.89
	emd	0.336	6.31
	дифференцирование	0.350	18.4
MLP	отсутствует	0.347	2788

	линейное	0.371	3089
	emd	0.377	2670
	дифференцирование	0.347	2145
OP-ELM	отсутствует	0.362	12.3
	линейное	0.337	12.8
	emd	0.335	11.5
	дифференцирование	0.342	11.7
SVM	отсутствует	0.356	2687
	линейное	0.349	2642
	emd	0.357	2376
	дифференцирование	0.353	2734
GP	отсутствует	0.334	2464
	линейное	0.335	3537
	emd	0.339	2662
	дифференцирование	0.372	2189

Аналогичные результаты для временных рядов потребления энергии приведены в таблице 2. Ошибки теста, как правило, немного хуже при линейном детрендировании, чем без детрендирования. EMD детрендирование дает результаты эквивалентные или немного лучше линейного, но все же хуже, чем при отсутствии детрендирования (за небольшим исключением для OP-ELM). Дифференцирование первого порядка дает явно лучшие результаты в случаях OP-ELM и SVM.

Таблица 3. Результаты для набора данных №3

Метод	Детрендерование	Среднеквадратическая ошибка	Время, с
DENFIS	отсутствует	0.438	2.46
	линейное	0.447	2.87
	emd	0.435	3.66
	дифференцирование	0.324	4.70
MLP	отсутствует	0.448	874
	линейное	0.457	882
	emd	0.451	1348
	дифференцирование	0.327	1037
OP-ELM	отсутствует	0.447	1.98
	линейное	0.456	1.64
	emd	0.453	2.96
	дифференцирование	0.344	2.25
SVM	отсутствует	0.439	448
	линейное	0.447	457
	emd	0.431	473
	дифференцирование	0.320	594
GP	отсутствует	0.436	163
	линейное	0.441	172
	emd	0.429	275
	дифференцирование	0.313	208

Результаты для временных рядов записей температуры приведены в таблице 3. Результаты для линейного детрендрования немного хуже для всех методов. Детрендрование, основанное на EMD, обеспечивает лучшие результаты, чем линейный, но все же дает очень незначительные улучшения или худшие результаты, если сравнивать с результатами без детрендрования. Тем не менее, дифференцирование первого порядка позволяет снизить показатель среднеквадратической ошибки по крайней мере на 25% для всех методов.

Таблица 4. Результаты для набора данных №4

Метод	Детрендрование	Среднеквадратическая ошибка	Время, с
DENFIS	отсутствует	0.237	1.85
	линейное	0.245	2.42
	emd	0.238	2.05
	дифференцирование	0.406	1.81
MLP	отсутствует	0.231	2047
	линейное	0.188	2176
	emd	0.143	2368
	дифференцирование	0.257	1987
OP-ELM	отсутствует	0.132	1.93
	линейное	0.216	1.61
	emd	0.169	2.47
	дифференцирование	0.243	1.72
SVM	отсутствует	0.166	159
	линейное	0.178	157

	emd	0.167	162
	дифференцирование	0.234	194
GP	отсутствует	0.118	160
	линейное	0.134	175
	emd	0.121	192
	дифференцирование	0.187	274

В таблице 4 представлены результаты для накопленной энергии с солнечных батарей. Ошибки теста, как правило, выше для линейного детрендрования, чем без детрендрования, за исключением MLP. EMD обеспечивает в целом значительно лучшие результаты, чем линейное детрендрование, но все же дает худший результат, если сравнивать с отсутствием детрендрования. Ошибки теста для дифференцирования явно хуже, без каких-либо исключений.

Дискуссия

Основываясь на приведенных выше результатах, можно сделать однозначный вывод, что в общем случае использование детрендрования не улучшает производительность методов CI ни в точности, ни в производительности. Более того, детрендрование не всегда давало положительный эффект — в части экспериментах результат применения модели к временному ряду без детрендрования был выше.

Таким образом, нельзя сделать вывод, является ли применение детрендрования желательным для конкретного случая. Возможно, именно этот факт и является причиной осторожного применения детрендрования при исследовании временных рядов.

Для конкретных методов детрендрования можно сделать следующие выводы:

- Линейный метод детрендрования и метод, основанный на EMD:

а. Извлечение тренда не дает увеличения точности результатов. Только для 1 из 4 наборов данных применение этих методов давало незначительный прирост точности вычислений;

б. Детрендинг на основе EMD дает лучшие результаты, чем линейное детрендинг. В части случаев разница была значительной, а в части — нет.

- Метод дифференцирования первого порядка:

а. В некоторых случаях значительно повышал точность моделей. Это может быть следствием его способности извлекать циклические компоненты.

б. Использование этого метода может привести к худшим результатам, чем результаты, полученные без детрендинга или с другими методами детрендинга.

Что касается вычислительного времени, то можно сделать вывод, что влияние линейного детрендинга зависит от метода и набора данных и не является значительным. Детрендинг на основе EMD приводит к значительному увеличению времени обучения. Дифференцирование первого порядка в целом не ускоряет вычисления, напротив, во многих случаях приводит к значительному замедлению.

Стоит иметь в виду, что результаты, представленные в данном исследовании, применимы к общим методам извлечения тренда, используемым в качестве единственного этапа предварительной обработки. Воздействие детрендинга в сочетании с другими методами предварительной обработки может существенно отличаться.

Заключение

В данной статье был проведен экспериментальный анализ влияния нескольких подходов детрендинга временных рядов на результаты применения моделей CI к этим временным рядам.

Целью исследования было сравнить несколько алгоритмов детрендинга с точки зрения эффективности предсказательных моделей

для временных рядов. Были рассмотрены 3 подхода детрендрования: линейное, основанное на EMD (или нелинейное, нестационарное) и дифференцирование первого порядка. В исследовании использовались следующие предсказательные методы: DENFIS, GP, MLP, OPELM и SVM.

На основе полученных результатов были сделаны выводы:

1. качественно эффект детрендрования однороден для всех рассмотренных методов CI;
2. в общем случае извлечение тренда не улучшает точность предсказательных моделей;
3. нестационарный метод детрендрования обеспечивает лучшую производительность, чем линейное детрендрование, хотя в большинстве случаев разница незначительна;
4. детрендрование с помощью метода дифференцирования первого порядка может отрицательно повлиять на результаты для временных рядов, которые показывают общие закономерности.

Таким образом, можно сделать вывод, что все рассмотренные методы детрендрования, как и любые интеллектуальные технологии обработки данных мониторинга в процессе строительства и эксплуатации, следует применять с особой осторожностью. Также необходимо дальнейшее исследование, чтобы проанализировать влияние детрендрования и десезонализации в совокупности на эффективность предсказательных методов для временных рядов.

Литература

- [1] Nie, S.Y.; Wu, X.Q. A historical study about the developing process of the classical linear time series models. // In Proceedings of The Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA); Springer: Berlin, Heidelberg, Germany, 2013; Volume 212, pp. 425–433.
- [2] Kenmei, B., Antoniol, G., di Penta, M. Trend analysis and issue prediction in large-scale open source systems. // In 2008 12th European Conference on Software Maintenance and Reengineering; 2008; pp. 73–82.
- [3] Gao, K., Khoshgoftaar, T.M.: A comprehensive empirical study of count models for software fault prediction. // IEEE Transactions on Reliability 56(2); 2007; pp. 223–236.
- [4] Agosto, A.; Cavaliere, G.; Kristensen, D.; Rahbek, A. Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). J. Empir. Financ.; 2016; pp. 640–663.

- [5] Karlis, D.; Meligkotsidou, L. Multivariate Poisson regression with covariance structure. // *Stat. Comput.*; 2005; pp. 255–265.
- [6] Lennon, H. Gaussian Copula Modelling for Integer-Valued Time Series. // Ph.D. Thesis, University of Manchester, Manchester, UK, 2016.
- [7] Fokianos, K. Count time series models. // *Time Ser. Appl. Handb. Stat.*; 2012; pp. 315–347.
- [8] Snyder, R. D., A. B. Koehler, and J. K. Ord. Forecasting for Inventory Control with Exponential Smoothing. // *International Journal of Forecasting*, 18 (1); 2002; pp. 5-18.
- [9] Xiong, H.; Shang P. Detrended fluctuation analysis of multivariate time series. // *Commun. Nonlinear Sci.*; 2017; pp. 12–21.
- [10] Flores-Marquez, E.L.; Ramírez-Rojas, A.; Telesca, L. Multifractal detrended fluctuation analysis of earthquake magnitude series of Mexican South Pacific Region. // *Appl. Math. Comput.*; 2015; pp. 1106–1114.
- [11] Cao, L.; Mees, A.; Judd, K. Dynamics from multivariate time series. // *Physica D* 1998; pp. 75–88.
- [12] Krzyszczak, J.; Baranowski, P.; Zubik, M.; Hoffmann, H. Temporal scale influence on multifractal properties of agro-meteorological time series. // *Agric. Forest Meteorol.*; 2017; pp. 223–235.
- [13] Zhang, G.,P. Time series forecasting using a hybrid ARIMA and neural network model. // *Neurocomputing* 50; 2003; pp. 159-175.
- [14] Conejo, A. J., Plazas, M. A., Espinola, R., & Molina, A. B. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. // *IEEE Transactions on Power Systems* 20 (2); 2005; pp. 1035-1042.
- [15] Aussem, A., Murtagh, F. Combining neural network forecasts on wavelet-transformed time series. // *Connection Science* 9 (1); 1997; pp. 113-122.
- [16] Msiza, I. S., F. V. Nelwamondo and T. Marwala. Artificial Neural Networks and Support Vector Machines for Water Demand Time Series Forecasting. // In: *IEEE International Conference on Systems, Man and Cybernetics*; Montreal; 2007; pp. 638–643.
- [17] Sifuzzaman, M., M. R. Islam, and M. Z. Ali. Application of Wavelet Transform and Its Advantages Compared to Fourier Transform. // *Journal of Physical Sciences* 13 (1); 2009; pp. 121–134.
- [18] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, Henry H Liu. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. // *Proc. R. Soc. Lond.*; 1998; pp. 903-995.
- [19] T. Tanaka, D. P. Mandic. Complex Empirical Mode Decomposition. // *IEEE Sig. Proc. Lett.*, vol. 14, no. 2; 2007; pp. 101–104.
- [20] P. Flandrin, G. Rilling, P. Gonçalvès. Empirical mode decomposition as a filter bank. // *IEEE Signal Process. Lett.*, vol. 11, no. 2; 2004; pp. 112–114.
- [21] Behforooz, H, Papamichael, N: Improved orders of approximation derived from interpolatory cubic splines. // *BIT*. 19; 1979; pp. 19–26.
- [22] J. C. Andre, L. M. Vincent, D. O'Connor, W. R. Ware. Applications of Fast Fourier transform to Deconvolution in Single Photon Counting. // *J. of Phys. Chem.* Vol. 83, No. 17; 1979; pp. 2285-2294.
- [23] N. Beaudoin. A high accuracy mathematical and numerical method for Fourier transform, integral, derivatives and polynomial splines of any order. // *Canadian Journal of Physics*. Vol. 76, No. 9; 1998; pp. 659-677.

- [24] S. M. Molaei and M. R. Keyvanpour. An analytical review for event prediction system on time series. // 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA); 2015; pp. 16.
- [25] E N. K. Kasabov, Q. Song. DENFIS: Dyanamic evolving neural fuzzy inference system and its application for time-series prediction. // IEEE Trans. Fuzzy Systems, vol.10, no.2; 2002; pp. 144-154.
- [26] Y. S. Abu-Mostafu, A. F. Atiya, M. magdon-Ismail, H. White. Introduction to the special issue on neural networks in financial engineering. // IEEE Transaction Neural Networks, vol.12, no.4; 2001; pp. 653-656.
- [27] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, A. Lendasse. OP-ELM: Optimally pruned extreme learning machine. // IEEE Trans.. Neural Netw., vol. 21, no. 1; 2010; pp. 158–162.
- [28] Y. Miche, A. Sorjamaa, A. Lendasse. OP-ELM: Theory, Experiments and a Toolbox. // In Proc. Int. Conf. Artif. Neural Netw., ser. Lect. Notes Comput. Sci, vol. 5163, Prague, Czech Republic; 2008; pp. 145–154.
- [29] Rahaman A, Islam M, Islam M, Sadi M, Nooruddin S. Developing IoT based smart health monitoring systems: a review. // Rev d’Intell Artif. 33(6); 2019; pp. 435–440.
- [30] <https://www.kaggle.com/taranvee/smart-home-dataset-with-weather-information>
- [31] <https://www.kaggle.com/lucabasa/dutch-energy>
- [32] <https://www.kaggle.com/atulanandjha/temperature-readings-iot-devices>
- [33] <https://www.kaggle.com/anikannal/solar-power-generation-data>
- [34] Guyon I, Elisseeff A. An introduction to variable and feature selection. // J Mach Learn Res 3; 2003; pp. 1157-1182.
- [35] Ansley, C. F., Kohn, R. Prediction mean squared error for state-space models with estimated parameters. // Biometrika 73; 1986; pp. 467–473.
- [36] Kagan, P. The use of digital technologies in building organizational and technological design / E3S Web Conf. Volume 263, 2021, XXIV International Scientific Conference “Construction the Formation of Living Environment” (FORM-2021) 263, 04040 (2021) doi: 10.1051/e3sconf/202126304040