

## Введение

Временной ряд - это последовательность числовых значений, упорядоченных по времени, отражающих характер изменений в определенном процессе или явлении [5]. Он широко используется в различных областях, таких как медицинская диагностика [6] (например, данные ЭКГ и ЭЭГ пациента), финансовый анализ [7] (например, данные по курсам акций и валют) или изучение климатических изменений [8]. Для анализа таких временных рядов необходимо хранить данные. Один из способов хранения подобных данных — использование технологии OpenTSDB [9].

## Хранение временных рядов в OpenTSDB

При работе с динамическими временными рядами основной задачей является эффективная организация данных. Поскольку временные ряды обновляются и пополняются новыми значениями, требуются оптимизированные структуры данных для быстрого доступа к информации и эффективного выполнения запросов. Кроме того, база данных должна обеспечивать быструю и эффективную запись новых данных.

OpenTSDB (Open Time Series Database) обеспечивает хранение и агрегацию временных рядов для обеспечения масштабируемости и производительности при работе с большими объемами данных. В OpenTSDB элемент временного ряда состоит из вещественного значения, уникального идентификатора временного ряда, метки времени и набора тегов. Теги представляют собой символьные строки для хранения метаданных.

OpenTSDB предоставляет возможности для индексирования данных и агрегации по времени и меткам. Это позволяет эффективно выполнять запросы к большим наборам данных временных рядов, выполнять аналитику и получать агрегированные результаты. Данная база данных разработана с учетом горизонтального масштабирования, что позволяет легко добавлять новые узлы для увеличения общей емкости хранения и производительности.

В OpenTSDB запросы к базе данных формируются с использованием языка JSON [10]. Эти запросы поддерживают различные операции, такие как арифметические и логические выражения, фильтры, группировка, статистические и аналитические функции. Одной из таких функций является линейная интерполяция. Если недостаточно реальных данных для интерполяции, то результатом может быть значение «NaN». Также в OpenTSDB существуют запросы по временному диапазону, фильтрация по меткам, позволяющая выбирать временные ряды с определенными значениями метаданных, агрегация, с помощью которой есть возможность вычислить среднее значение, сумму или других статистических показателей для временных рядов.

OpenTSDB не предоставляет инструменты для интеллектуального анализа временных рядов, за исключением линейной интерполяции, но допускает расширения от сторонних разработчиков, добавляющие указанную функциональность (например, библиотека R2Time [11], реализованная на языке программирования R).

## Прогнозирование временных рядов

Одна из ключевых целей анализа и изучения временных рядов - это их прогнозирование. Прогноз представляет собой вероятностное предположение о состоянии определенного явления в будущем, основанное на специальном научном исследовании [12]. В данном разделе будет рассмотрено применение таких моделей прогнозирования, как регрессионная модель и модель экспоненциального сглаживания.

### Регрессионная модель

Регрессионная модель базируется на концепции регрессии, которая представляет собой статистическую зависимость математического ожидания одной случайной величины от значений другой или нескольких случайных величин. Основная цель при построении регрессионной модели заключается в обнаружении связи между исходной переменной и несколькими регрессорами (факторами). В данной работе будут рассмотрены следующие виды регрессионных моделей:

- простая линейная регрессия;
- множественная регрессия;
- нелинейная регрессия.

### **Простая линейная регрессия.**

Модель основана на предположении о существовании дискретного внешнего фактора  $X(t)$ , который влияет на рассматриваемый процесс  $Z(t)$ , и связь между этим фактором и процессом описывается в виде линейной функции. Уравнение для модели прогнозирования временного ряда на основе линейной регрессии представляет собой:

$$Z(t) = a_0 + a_1 \cdot X(t) + \varepsilon_t, \quad (1)$$

где  $a_0, a_1$  — коэффициенты регрессии;  $\varepsilon_t$  — ошибка модели.

Классический метод оценки коэффициентов линейной регрессии строится на методе наименьших квадратов. Этот метод позволяет найти значения коэффициентов  $a_0$  и  $a_1$ , при которых сумма квадратов разностей между фактическими значениями  $Z$  и предсказанными значениями  $Z_t$  будет минимальной, то есть:

$$\sum_{i=1}^n (Z_i - Z_{t_i})^2 = \sum_{i=1}^n \varepsilon_i^2 \rightarrow \min \quad (2)$$

**Множественная регрессия.** Для прогнозирования значений процесса  $Z(t)$  в момент времени  $t$  необходимо иметь значение  $X(t)$  для того же времени, что в реальности часто бывает сложно. На практике процесс  $Z(t)$  может зависеть от множества дискретных внешних факторов  $X_1(t), \dots, X_s(t)$ . Таким образом, модель прогнозирования принимает следующий вид:

$$Z(t) = a_0 + a_1 \cdot X_1(t) + a_2 \cdot X_2(t) + \dots + a_s \cdot X_s(t) + \varepsilon_t, \quad (3)$$

$a_0, a_1, \dots, a_s$  — коэффициенты регрессии;  $\varepsilon_t$  — ошибка модели.

Коэффициенты регрессии также определяются с использованием метода наименьших квадратов, но в этом случае решается система нормальных уравнений,

определитель которой выглядит как:

$$\begin{vmatrix} n & \sum X_1 & \dots & \sum X_s^2 \\ \sum X_1 & \sum X_1^2 & \dots & \sum X_1 X_s \\ \dots & \dots & \dots & \dots \\ \sum X_s & \sum X_1 X_s & \dots & \sum X_s^2 \end{vmatrix} \quad (4)$$

### **Нелинейная регрессия.**

Нелинейная модель предполагает существование определенной функции:

$$Z(t) = F(X(t), A), \quad (5)$$

где  $Z(t)$  — исходный процесс;  $X(t)$  — внешний фактор, от которого зависит процесс  $Z(t)$ ;  $A$  — функция, параметры которой необходимо определить в рамках построения модели прогнозирования.

На практике нелинейные регрессионные модели применяются редко, поскольку лишь небольшая часть процессов имеет известную заранее функциональную зависимость [3].

### **Модель экспоненциального сглаживания**

Для того, чтобы прогнозирующая система, включающая тот или иной математический аппарат, могла автоматически распознавать изменения параметров модели, используется метод экспоненциального сглаживания.

С этой целью используется сглаженная функция наблюдений:

$$\hat{y}_t = a \cdot y_t + (1 - a) \cdot \hat{y}_{t-1}. \quad (6)$$

Операция (6), выполняемая при каждом новом наблюдении, называется экспоненциальным сглаживанием. Значение  $a$ , которое примерно равно  $1/n$ , называется коэффициентом сглаживания. Поскольку операция (6) проводится одинаково для всех значений ряда динамики, мы можем выразить учет более ранних значений в

виде:

$$\begin{aligned}\hat{y}_t &= a \cdot y_t + (1 - a) \cdot (a \cdot y_{t-1} + (1 - a) \cdot \hat{y}_{t-1}) = \\ &= a \cdot y_t + (1 - a) \cdot (a \cdot y_{t-1} + (1 - a) \cdot (a \cdot y_{t-2} + (1 - a) \cdot \hat{y}_{t-3})) = \\ &= \dots = a \cdot \sum_{k=1}^{t-1} (1 - a)^k \cdot y_{t-k} + (1 - a)^t \cdot y_0. \quad (7)\end{aligned}$$

Таким образом, сглаженное значение  $\hat{y}_t$  представляет собой линейную комбинацию всех значений ряда динамики, при этом их веса уменьшаются по геометрической прогрессии со временем.

## Список литературы

- [1] Эйлин Нильсен - Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение. 2020. URL:<http://www.williamspublishing.com/PDF/978-5-907365-04-9/part.pdf>
- [2] Дейт К. - Введение в системы баз данных, Киев—Москва: Диалектика, 1998. С. 90
- [3] Аникин А. С., Говжеев Г. Д. — Применение регрессионного анализа для исследования временных рядов // Актуальные исследования. 2022. №6 (85). С. 18-22. URL:<https://apni.ru/article/3737-primenenie-regressionnogo-analiza-dlya-issled> (дата обращения: 08.10.2023)
- [4] С.Н. Спирков — АЛГОРИТМ ИСПОЛЬЗОВАНИЯ МЕТОДА ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ. 2018. URL:<https://elibrary.ru/item.asp?id=36969441>
- [5] Иванова Е.В., Цымблер М.Л. Обзор современных систем обработки временных рядов // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2020. Т. 9, № 4. С. 79–97.
- [6] Абрамов Михаил Владимирович АППРОКСИМАЦИИ ЭКСПОНЕНТАМИ ВРЕМЕННОГО КАРДИОЛОГИЧЕСКОГО РЯДА НА ОСНОВЕ ЭКГ // ВК. 2010. №9. URL: <https://cyberleninka.ru/article/n/approksimatsii-eksponentami-vremennogo-kardiologicheskogo-ryada-na-osnove-ekg> (дата обращения: 01.11.2023).
- [7] Передриенко А.И., Лютая Т.П., Харитонов И.М., Степанченко И.В. МЕТОДЫ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ С МАЛЫМИ ОБЪЁМАМИ ВЫБОРКИ // ИВД. 2020. №5 (65). URL: <https://cyberleninka.ru/article/n/metody-kratkosrochnogo-prognozirovaniya-finansovyh-vremennyh-ryadov-s-malymi-obyomami-vyborki> (дата обращения: 01.11.2023).

- [8] Михалаш Сергей Геннадьевич, Мингалёв Дмитрий Эдуардович, Евдокимов Сергей Игоревич Использование анализа временных рядов в изучении многолетних температурных изменений // Вестник Псковского государственного университета. Серия: Естественные и физико-математические науки. 2014. №4. URL: <https://cyberleninka.ru/article/n/ispolzovanie-analiza-vremennyh-ryadov-v-izuchenii-mnogoletnih-temperaturnyh-izmeneniy> (дата обращения: 01.11.2023).
- [9] OpenTSDB Documentation. URL: <http://opentsdb.net/docs/3x/build/html/> (дата обращения: 01.11.2023).
- [10] Json. URL: <https://www.json.org/json-en.html> (дата обращения: 01.11.2023).
- [11] R2Time: A Framework to Analyse Open TSDB Time-Series Data in HBase URL: <https://ieeexplore.ieee.org/document/7037792> (дата обращения: 01.11.2023).
- [12] Толковый словарь Ожегова. URL: <https://gufo.me/dict/ozhegov/прогноз> (дата обращения: 11.11.2023)