

ОГЛЯД ПІДХОДІВ ДО ВИЗНАЧЕННЯ ПОРЯДКУ ФУР'Є У МОДЕЛІ FACEBOOK PROPHET ДЛЯ МОДЕЛЮВАННЯ СЕЗОННОЇ СКЛАДОВОЇ ЧАСОВОГО РЯДУ

Вінницький національний технічний університет

Анотація

Розглянуто важливість та основні проблеми моделювання часового ряду. Запропоновано новий підхід визначення параметрів для моделювання сезонності часового ряду за допомогою нейронної мережі. В ході дослідження була запропонована інформаційна технологія, яка генерує оптимальні значення для параметрів порядку ряду Фур'є, а також періоду сезонності, що використовуються при тренуванні моделі Facebook Prophet.

Ключові слова: інформаційна технологія, машинне навчання, нейронна мережа, ідентифікація параметрів моделі Prophet, системний аналіз.

Abstract

The thesis aims to overview the importance and know problems related to modeling time series. It is proposed to improve it by identifying time-series seasonality parameters using neural network. During research, an information technology was proposed. This system generates optimal values for parameters of the order of the Fourier series, as well as the seasonality period, used in training of the Facebook Prophet model.

Keywords: information technology, machine learning, neural network, identification of Prophet model parameters, system analysis.

Вступ

Опрацюванням та аналізом статистичних даних різного характеру та походження займається велика кількість спеціалістів. Часовими рядами може бути представлена фінансова інформація, рівень чи показники поширення інфекційного захворювання, дані моніторингу якості води та повітря тощо [1-5]. Основним завданням аналітиків є ретельне вивчення наявних даних та моделювання часових рядів для передбачення майбутніх даних, що у свою чергу, може допомогти попередити або уникнути певних подій. Однією з відомих моделей прогнозування часових рядів є модель Prophet, розроблена компанією Facebook. Ключовим у використанні моделі є оптимальне її налаштування, що дає можливість досягти якомога кращої точності передбачення. Відповідно, є потреба у визначенні оптимальних параметрів для моделювання сезонності часового ряду.

Проблемами аналізу часових рядів займаються вчені багатьох країн вже тривалий час. Як правило, використовуються такі підходи для пошуку порядку ряду Фур'є: частотний аналіз, аналіз спектра та перебір усіх можливих значень, наприклад від 3 до 12 чи 20, причому останній підхід є більш популярним. Однак, як правило, визначення порядку ряду Фур'є здійснюється не самостійно. Разом із ним ще визначаються й інші параметри моделі, а тоді кількість комбінацій для повного перебору зростає дуже сильно - часто їх взагалі не можна перебрати усі, коли інші параметри є дробовими числами, а тому потрібні більш точні методи, які, на відміну від інших, добре працюють з сильно зашумленими даними.

Метою даного дослідження було розробити і випробувати підхід до визначення порядку Фур'є у моделі Facebook Prophet для моделювання сезонної складової часового ряду за умов його значної зашумленості.

Алгоритм розв'язання поставленої задачі

Для вирішення поставленої задачі пропонується підхід з використанням методів машинного навчання за таким алгоритмом:

- 1) Синтезуємо ряд Фур'є

$$f(x) = \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

для заданого порядку n (решта змінних задається фіксованими);

- 2) На отриманий набір значень накладаємо шум із певною дисперсією d_1 ;
- 3) Використовуючи синтезовані на кроці 2 дані, будуємо модель Prophet. При цьому, ми відключаємо налаштування сезонності за замовчуванням та заміняємо на заданий порядок n ряду Фур'є;
- 4) Робимо прогноз для створеної моделі, а на отримані значення додатково накладаємо шум із заданою дисперсією d_2 .
- 5) Повторюємо кроки 1-4 для різних наборів вхідних параметрів (порядок ряду і дисперсія шумів) для генерування набору тренувальних даних.
- 6) Використовуємо тренувальні дані, отримані на кроці 5, для ідентифікації порядку ряду з використанням методів машинного навчання.

Програмна реалізація запропонованої інформаційної технології

Запропонований алгоритм авторами реалізовано на Python у вигляді програми-ноутбука на платформі Kaggle – найбільшій у світі платформі датасайнтистів від Google. Як було зазначено раніше, для тренування нейронної мережі необхідно синтезувати набір даних. Для цього було реалізовано функцію, яка, перебираючи в циклі різні значення порядку ряду Фур'є та дисперсії шуму, генерує дані моделлю Prophet. Приклад роботи даної функції у вигляді графіка зображено на рис. 1.

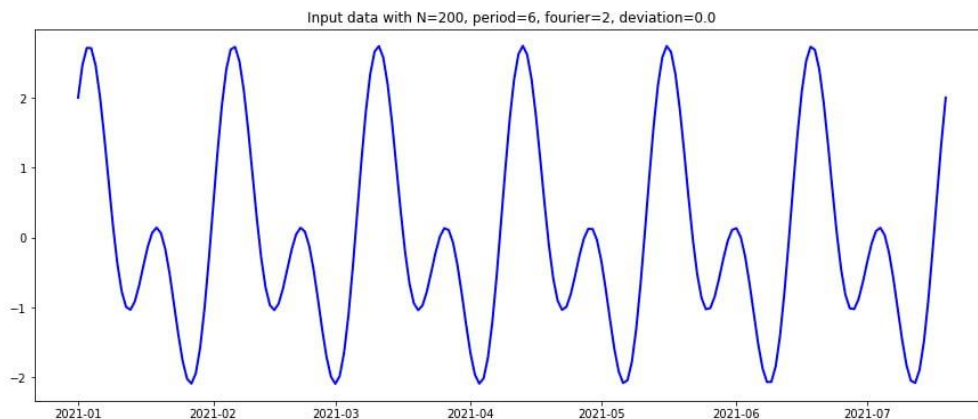


Рисунок 1 – Приклад роботи функції для генерування даних за допомогою моделі Prophet

Використовуючи описану функцію, було отримано набір тренувальних даних із більше 800 різних комбінацій порядку ряду Фур'є та дисперсії шуму (рис. 2).

value_1	value_2	value_3	value_4	value_5	value_6	...	value_1
1.212929	1.033774	1.490677	1.588684	1.327723	1.247872	...	-0.438
0.839207	1.214710	1.414985	1.412930	1.583367	1.651833	...	-0.542
0.881902	1.257847	1.499206	0.962251	1.067706	0.983035	...	-0.245
1.235705	0.767356	0.820297	1.253692	1.766596	1.389924	...	-0.892
0.903866	1.106160	1.049172	1.790335	1.210122	1.248544	...	-0.434
...
12.785682	5.139152	0.473676	2.031674	-0.133421	-0.055842	...	-1.828
13.039057	4.370853	-0.426115	1.719395	-0.604905	-0.593807	...	-0.983
12.363012	4.422465	0.239665	2.585668	0.533538	-0.348935	...	0.1629
13.145853	4.489238	-0.905436	0.917066	1.230329	-0.708511	...	-0.987
12.532139	4.957577	-1.328651	1.285472	-0.185101	-0.434669	...	-0.967

Рисунок 2 – Сформовані тренувальні дані

Тренувальні дані були використані ідентифікації порядку ряду Фур'є з використанням методу Random Forest Regressor. Для тестування було використано вибірково взяті 20% даних, які не використовувались для тренування моделі. Відносна похибка визначення порядку Фур'є склала 2-3%, що є дуже гарним показником. Зауважимо, що у разі збільшення дисперсії шуму, ця похибка суттєво зростає, тому цей метод має обмежене застосування.

Висновки

Запропоновано визначати оптимальний порядок ряду Фур'є для опису сезонності у моделі Facebook Prophet з використанням методів машинного навчання. Розроблено двічі зашумлений датасет, який, однак, дозволив побудувати доволі точну модель для ідентифікації цього порядку.

Експерименти показали, що за реальних умов варто ідентифікувати й інші параметри моделі Prophet, інакше, особливо, за високої зашумленості даних, визначення порядку ряду Фур'є може не допомогти побудувати дійсно адекватну модель.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Carmela Comito, Clara Pizzuti, Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review, Artificial Intelligence in Medicine, Volume 128, 2022, 102286, ISSN 0933-3657, <https://doi.org/10.1016/j.artmed.2022.102286>.
2. Santangelo, Omar Enzo, et al. "Machine Learning and Prediction of Infectious Diseases: A Systematic Review." *Machine Learning and Knowledge Extraction* 5.1 (2023): 175-198.
3. S. Jadon, A. Patankar and J. K. Milczek, "Challenges and Approaches to Time-Series Forecasting for Traffic Prediction at Data Centers," *2021 International Conference on Smart Applications, Communications and Networking (SmartNets)*, Glasgow, United Kingdom, 2021, pp. 1-8, doi: 10.1109/SmartNets50376.2021.9555422.
4. В. Б. Мокін, А. В. Лосенко, А. Р. Ящолт, «Інформаційна технологія аналізу та прогнозування кількості нових випадків хвороби на коронавірус SARS-COV-2 в Україні на основі моделі Prophet», Вісник ВПІ, вип. 5, с. 71–83, Лис. 2020.
5. Мокін В. Б. Інформаційна технологія аналізу та прогнозування багатохвильової кількості нових випадків захворювань на коронавірус COVID-19 на основі моделі Prophet / В. Б. Мокін, А. В. Лосенко, А. Р. Ящолт // Вісник Вінницького політехнічного інституту. - Вип. 6, С. 65–75, 2020.

Шмундяк Дмитро Олександрович – аспірант кафедри системного аналізу та інформаційних технологій.

Лосенко Арсен Володимирович – аспірант кафедри системного аналізу та інформаційних технологій.

Мокін Віталій Борисович – д. т. н., професор, завідувач кафедри системного аналізу та інформаційних технологій, e-mail vbmokin@vntu.edu.ua

Shmundiak Dmytro Oleksandrovych - graduate student of the Chair of System Analysis and Information Technologies.

Losenko Arsen Volodymyrovych - graduate student of the Chair of System Analysis and Information Technologies.

Mokin Vitaliy Borysovych - Doctor of Technical Sciences, Professor, Head of the Chair of System Analysis and Information Technologies, e-mail vbmokin@vntu.edu.ua