

ЄМЕЦЬ КИРИЛО

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-5157-9118>e-mail: kyrylo.v.yemets@lpnu.ua

МЕТОДИ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ З ВИРАЖЕНОЮ СЕЗОННІСТЮ НА ОСНОВІ ТРАНСФОРМЕРІВ

Прогнозування часових рядів є критично важливим завданням, що охоплює різні сфери, такі як фінанси та медицина, погода, тощо, вимагаючи точних прогнозів, отриманих на основі аналізу історичних даних. Складність цієї задачі посилюється при аналізі сезонних даних, які характеризуються непередбачуваними коливаннями та різноманітними впливами, тим самим зменшуючи ефективність існуючих інструментів штучного інтелекту. У цій роботі проведено огляд та експериментальне порівняння передових моделей прогнозування часових рядів, з вираженою сезонністю в даних, на основі трансформерів. У дослідження було включено три моделі: Temporal Fusion Transformer, PatchTST та DLinear. Подано особливості їх архітектури, принципи роботи, окреслено їхні переваги та обмеження. Результати моделювання на основі цих моделей оцінювалися із використанням трьох різних показників ефективності, під час розв'язання задачі аналізу даних про продажі продукції. Проведений аналіз показує, що моделі нейронних мереж, які використовують архітектури трансформера, демонструють високу ефективність під час розв'язання задачі прогнозування сезонних часових рядів. Однак ці моделі вимагають значних часових ресурсів для оптимізації десятків тисяч параметрів. Саме тому, вибір конкретної моделі прогнозування часових рядів повинен опиратися на конкретну задачу, доступні ресурси та пріоритети замовника.

Ключові слова: часові ряди, трансформери, сезонність.

YEMETS KYRYLO

Lviv Polytechnic National University

FORECASTING METHODS OF TIME SERIES WITH EXPRESSED SEASONALITY BASED ON TRANSFORMER

Forecasting time series data with seasonal patterns remains a significant challenge in various domains. The complex and non-stationary nature of such data requires advanced modeling techniques capable of capturing intricate temporal dependencies and periodic fluctuations. Traditional statistical methods like ARIMA models often struggle to adequately represent the underlying dynamics, motivating the exploration of more flexible and powerful approaches. In recent years, deep learning architectures, particularly transformer-based models, have demonstrated remarkable success in handling sequential data and capturing long-range dependencies. Their self-attention mechanisms and parallel processing capabilities make them well-suited for time series forecasting tasks involving seasonality. This study provides a comprehensive evaluation of three prominent transformer-based models: Temporal Fusion Transformer (TFT), PatchTST, and DLinear. Each model brings unique architectural innovations and training strategies to tackle the intricacies of seasonal time series forecasting effectively. The experimental results, evaluated on a real-world product sales dataset, reveal the potential of these models to outperform traditional methods and achieve superior forecasting accuracy. However, it is crucial to consider the trade-off between model complexity, computational resources, and the specific requirements of the forecasting task at hand. As the field of time series forecasting continues to evolve, transformer-based models offer a promising direction for handling complex seasonal patterns, paving the way for more accurate and reliable predictions across various industries. Nevertheless, further research is necessary to address the challenges of model interpretability, efficient parameter optimization, and the incorporation of domain-specific knowledge into these architectures.

Keywords: time series, transformers, seasonality.

Постановка проблеми

Проблема точного прогнозування є критично важливою у таких областях, як фінанси, медицина [1] та кліматологія, де важливо робити вірні передбачення, опираючись на історичні дані [2]. Специфічні виклики виникають під час роботи з сезонними даними, що характеризуються непередбачуваними змінами та різноманітними зовнішніми факторами, що ставить під сумнів ефективність наявних інструментів на основі штучного інтелекту. Відтак, необхідно застосувати детальний підхід, який би включав глибинний аналіз даних, акуратний вибір моделі та розробку дієвих стратегій. Саме тому актуальними є дослідження та порівняльний аналіз передових моделей для прогнозування часових рядів, особливо призначених для роботи з сезонними даними, з акцентом на подолання супутніх труднощів та підвищення точності прогнозів. Було обрано три моделі на основі архітектури трансформерів, зокрема Temporal Fusion Transformer [3], PatchTST [4] та DLinear [5], їх було оглянуто так, щоб виявити їх ключові переваги та потенційні слабкі місця. Ці моделі, нарівні з між собою, піддавались детальній оцінці трьома метриками, особливо у контексті аналізу даних продажів. Наш аналіз показав, що моделі на основі нейронних мереж з архітектурою трансформера забезпечують найвищу точність прогнозів для сезонних часових рядів. Втім, варто відмітити, що для досягнення оптимальних результатів такі моделі потребують великих ресурсів у налаштуванні параметрів, що може включати роботу з тисячами змінних.

Аналіз останніх джерел

Класичні методи прогнозування часових рядів, такі як ARIMA [6] (авторегресійне інтегроване ковзне середнє), VAR [7] (векторна авторегресія) та Prophet [8], доволі часто використовуються для прогнозування часових рядів. Вони забезпечують задовільні результати аналізу та прогнозування трендів та патернів для окремих задач. Однак ці методи зіштовхуються з обмеженнями, особливо коли йдеться про довгострокове прогнозування, через кілька причин. Методи на кшталт ARIMA вимагають, щоб дані часових рядів були

стаціонарними, тобто статистичні властивості серії не повинні змінюватися з часом. Ця вимога часто потребує попередньої обробки, такої як диференціювання, що може бути складним і може не бути ефективним для довгострокових трендів або нелінійних патернів. Як ARIMA, так і VAR моделі базуються на лінійних припущеннях, де майбутні значення передбачаються як лінійні комбінації минулих значень. Цей лінійний підхід може не вловлювати складні, нелінійні взаємозв'язки, властиві багатьом довгостроковим часовим рядам, таким як ті, що впливаються нерегулярними економічними циклами, технологічними досягненнями або змінами в навколишньому середовищі [9]. Моделі на кшталт ARIMA та Prophet можуть включати сезонні ефекти, але їхня здатність моделювати складні сезонні патерни або кілька сезонностей в одній серії є обмеженою. Довгострокові прогнози, які можуть охоплювати кілька сезонних циклів, вимагають більш тонкого розуміння сезонних впливів, яке ці моделі можуть не забезпечити належним чином. Довгострокове прогнозування часто вимагає врахування зовнішніх факторів або екзогенних змінних, які можуть істотно впливати на майбутні значення часового ряду. Класичні методи, як-от ARIMA та VAR, мають обмежені можливості інтегрувати ці фактори у модель прогнозування. Prophet дозволяє включати свята та особливі події, але його можливості щодо обробки безперервних зовнішніх змінних та раптових змін трендів можуть бути обмежені. Оскільки горизонт прогнозування розширюється, складність обчислень для моделей на кшталт VAR значно зростає, роблячи їх менш практичними для довгострокових прогнозів. Крім того, потреба у переоцінці моделей з новими даними може бути обчислювально інтенсивною та менш ефективною для великих наборів даних або коли потрібні часті оновлення. Довгострокові прогнози по суті супроводжуються вищою невизначеністю, і класичні методи можуть не завжди надавати міцний фреймворк для кількісної оцінки та управління цією невизначеністю. Хоча Prophet генерує інтервали прогнозування, точне вловлення невизначеності для довгострокових прогнозів, особливо в умовах структурних змін або чорних лебедів, залишається складним завданням. Враховуючи ці обмеження, дослідники та практики часто звертаються до більш гнучких, нелінійних моделей, таких як методи машинного навчання та глибокого навчання, для довгострокового прогнозування. Ці моделі можуть краще обробляти нелінійні взаємозв'язки, інтегрувати кілька джерел даних та адаптуватися більш динамічно до змін трендів та патернів з часом.

Метою роботи є дослідження ефективності використання моделей на основі архітектури трансформерів для передбачення часових рядів з вираженою сезонністю.

Виклад основного матеріалу

Temporal Fusion Transformer (TFT) [3], є високорозвиненою моделлю, призначеною для задач прогнозування, особливо ефективною у вирішенні складнощів сезонних часових рядів. Її архітектура ретельно розроблена для обробки та навчання на основі як статичної, так і динамічної інформації, що робить її надзвичайно вдалою для сценаріїв, де критично важливе розуміння тимчасових патернів і зв'язків. Однією з вирізняючих особливостей TFT є її механізми вибору змінних. Ці компоненти відіграють ключову роль у подаванні різних вхідних змінних для визначення тих, що найбільш актуальні для прогнозування в будь-який конкретний час. Ця здатність зосереджуватися на найінформативніших ознаках є особливо цінною в сезонному прогнозуванні, де значущість певних входів може змінюватися протягом року. Крім того, TFT використовує кодування для статичних коваріат для інтеграції інформації, що не змінюється в часі, у процес прогнозування. Включення статичних даних забезпечує стабільний контекст, який збагачує розуміння моделлю кожного часового ряду, покращуючи її здатність вловлювати унікальні характеристики, що впливають на сезонні патерни. Можливість опрацьовувати часові ряди з різними частотами TFT є іншим важливим аспектом її архітектури. Завдяки ефективному навчанню на основі як короткотермінових, так і довготермінових залежностей у даних, модель може точно передбачати майбутні тренди та патерни. Це розуміння динаміки часу є суттєвою перевагою для прогнозування сезонних коливань, що часто включають складні взаємодії на різних часових шкалах. Більше того, використання TFT механізму уваги з кількома головами дозволяє йому зосереджуватися на найбільш релевантній інформації з минулого під час виконання прогнозів. Цей цілеспрямований підхід покращує здатність моделі навчатися з довготермінових сезонних трендів, забезпечуючи, що прогнози базуються на глибокому розумінні історичних даних. Нарешті, здатність TFT генерувати інтервали прогнозування через квантильні виходи є особливо корисною для сезонного прогнозування. Надаючи діапазон можливих результатів замість одного прогнозу, він пропонує більш нюансований погляд на майбутні невизначеності, що є критично важливим для планування та управління ризиками в сезонних контекстах.

PatchTST, що розшифровується як Patch Transformer for Time Series Forecasting [4], представляє собою новаторську архітектуру, яка використовує можливості моделі Transformer, спеціально розроблену для вирішення викликів прогнозування часових рядів, включаючи ті, що пов'язані з сезонними патернами. Хоча конкретні деталі PatchTST можуть варіюватися в залежності від джерела або реалізації, основна ідея поєднує потужні механізми уваги Transformer з підходом на основі "патчів" для ефективної обробки даних часових рядів. Архітектура PatchTST побудована навколо концепції розбиття часового ряду на менші сегменти або "патчі". Цей метод дозволяє моделі зосередитися на конкретних частинах даних за раз, спрощуючи ідентифікацію та вивчення основних сезонних трендів та патернів. Обробляючи ці сегменти як окремі входи, PatchTST може застосовувати механізми самоуваги Transformer до відносин як всередині патчів, так і між патчами, дозволяючи їй вловлювати складні часові залежності на різних часових шкалах. Використання моделі Transformer у PatchTST надає перевагу у вирішенні задач з довготривалими залежностями. Це особливо корисно для сезонного прогнозування, де розуміння залежностей між віддаленими точками у часі

може бути вирішальним для передбачення майбутніх трендів. Механізм самоуваги динамічно визначає важливість різних часових кроків, дозволяючи PatchTST зосереджуватися на найактуальнішій інформації для точного прогнозування. Ще однією перевагою PatchTST є його масштабованість та гнучкість. Модель можна налаштувати для обробки часових рядів різної довжини та різного ступеня сезонності, що робить її придатною для широкого спектру завдань прогнозування. Незалежно від того, чи йдеться про короткострокові дані щоденних продажів з тижневими патернами або довгострокові дані клімату з річними циклами, адаптивна структура PatchTST може бути налаштована для оптимізації продуктивності з урахуванням конкретних характеристик даних.

DLinear [5], – це архітектури, розроблені для покращення лінійних моделей прогнозування часових рядів шляхом явного виділення компонент ряду, таких як тренд, сезонність та зміни розподілу в даних. DLinear прагне підвищити точність прогнозування, явно розкладаючи часовий ряд на компоненти тренду та сезонності (залишкові) та застосовуючи спеціалізовану обробку до кожної. Компоненти тренду отримують застосуванням ядра ковзного середнього до сирих даних, згладжуючи коливання для виявлення основного тренду. Сезонний компонент залишається після видалення компоненту тренду з сирих даних, захоплюючи циклічні патерни та шум.

Між 1989 і 1994 роками Школа бізнесу Університету Чикаго співпрацювала з нині закритим магазином Dominick's Finer Foods над серією випадкових експериментів на рівні магазинів, зосереджених на управлінні полицями та ціноутворенні в понад 25 різних категоріях у всій мережі з 100 магазинів. Це партнерство принесло близько дев'яти років детальних даних на рівні магазинів щодо продажів понад 3,500 унікальних кодів продуктів. Отриманий набір даних Dominick's охоплює дані сканерів на рівні магазинів, зібрані в Dominick's Finer Foods протягом більш ніж семи років, що включає 115704 тижневих часових ряди, що деталізують прибуток окремих одиниць зберігання товарів від рітейлера. Цей всеохоплюючий набір даних включає файли, специфічні для категорій, та загальні файли, що охоплюють усі категорії проекту, поряд з даними щоденних продажів понад 3500 продуктів, підрахунками клієнтів та демографічними даними та даними про продажі конкретних магазинів. Для моделювання Pytorch використовувався як фреймворк глибокого навчання для всіх методів. Обчислення виконувалися на графічному процесорі NVIDIA T4. Розмір лагу для даних часових рядів становив 10 тижнів. Прогнозування проводилося на 8 тижнів. Для методів глибокого навчання набір даних був переформатований як набір малих часових рядів з 10 вхідними ознаками та 8 вихідними ознаками.

Для оцінювання ефективності моделювання використано ряд показників, зокрема:

$$MSE = \frac{1}{n} \sum_{t=1}^n (F_t - A_t)^2, \quad (1)$$

$$SMAPE = \frac{100}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|F_t| + |A_t|) \times 0.5}, \quad (2)$$

$$COVERAGE = \frac{1}{n} \sum_{t=1}^n |F_t < A_t|, \quad (3)$$

де F_t - передбачене значення, A_t - істинне значення, t - поточне номер даних, n - кількість даних.

Результати моделювання на основі показників (1)-(3) зведено у таблиці 1.

Таблиця 1

Результати моделювання

Модель трансформерів/показник ефективності	MSE	SMAPE	COVERAGE	Параметри (в тисячах шт)
PatchTST	139,005	1,60	0,16	42,6
DLinear	156,759	1,62	0,18	3,6
Temporal Fusion Transformer	185,839	1,59	0,20	98,5

Аналізуючи наведені результати моделювання (табл. 1), можна зробити кілька ключових висновків щодо ефективності розглянутих моделей трансформерів - PatchTST, DLinear та Temporal Fusion Transformer, в контексті трьох показників: середньоквадратична помилка (MSE), симетрична середня абсолютна відсоткова помилка (SMAPE) та покриття (COVERAGE). PatchTST показує найкращі результати за середньоквадратичною помилкою (MSE) з показником 139,005, що свідчить про вищу точність прогнозів порівняно з іншими моделями. Ця модель також має найнижчу відсоткову помилку (SMAPE) 1,60 і показник покриття 0,16, вказуючи на досить високу надійність прогнозів. Однак, вона вимагає значної кількості параметрів (42,6 тис.), що може збільшити час та витрати на тренування моделі. DLinear має трохи гірші показники за MSE (156,759) та SMAPE (1,62) порівняно з PatchTST, але краще покриття (0,18). Важливою перевагою цієї моделі є значно менша кількість параметрів (3,6 тис.), що робить її більш ефективною з точки зору обчислень та швидшою для налаштування. Temporal Fusion Transformer демонструє найгірші результати за MSE (185,839) та має SMAPE (1,59), що є порівняно конкурентоспроможним з PatchTST, але вище покриття (0,20), що може вказувати на кращу здатність до загальної адаптації прогнозів до різних умов. Однак, ця модель має найбільшу кількість параметрів (98,5 тис.), що робить її найбільш ресурсомісткою з усіх розглянутих. Узагальнюючи, PatchTST видається найкращим варіантом для задач, де пріоритетом є точність прогнозування та ефективність використання ресурсів, незважаючи на вищі вимоги до кількості параметрів.

DLinear пропонує хороший баланс між точністю, ефективністю та низькою кількістю параметрів, що робить її оптимальним вибором для обмежених ресурсів або швидкого прототипування. Temporal Fusion Transformer, попри свою високу ресурсомісткість, може бути виправданим вибором для складних задач прогнозу.

Висновки

У цій роботі було показано важливість високоточного прогнозування в різних областях і викликів, пов'язаних з сезонними даними. Подано критичний аналіз класичних методів прогнозування, таких як ARIMA, VAR та Prophet, та описано їх обмеження у контексті довгострокового прогнозування та сезонності. Автором проведено огляд існуючих методів прогнозування часових рядів на основі трансформерів. Подано принципи їх роботи та обґрунтовано основні переваги. Проведено експериментальне моделювання ефективності роботи передових моделей на основі трансформерів (Temporal Fusion Transformer (TFT), PatchTST та DLinear) для підвищення точності прогнозів сезонних часових рядів. Порівняльний аналіз показує, що PatchTST демонструє найкращі результати за точністю (MSE та SMAPE) та надійністю (COVERAGE), але вимагає підбору значної кількості параметрів. DLinear пропонує хороший баланс між точністю та ефективністю за рахунок меншої кількості параметрів. Temporal Fusion Transformer, попри найбільшу кількість параметрів, демонструє конкурентні результати і може бути виправданим вибором для розв'язання складних задач. Саме тому, вибір конкретної моделі залежить від специфіки задачі, доступних ресурсів та пріоритетів щодо точності та швидкодії.

Acknowledgment

This work is supported by European Union's Horizon Europe research and innovation programme under grant agreement No 101138678, project ZEBAI (Innovative methodologies for the design of Zero-Emission and cost-effective Buildings enhanced by Artificial Intelligence).

Література

1. Ізонін І.В. Ансамблевий метод уточнення параметрів моделі регресії: прямий підхід // Математичне та комп'ютерне моделювання. Серія: Технічні науки, 2023, 35-44. DOI: <https://doi.org/10.32626/2308-5916.2023-24.35-44>
2. Casolaro A, Capone V, Iannuzzo G, Camastra F. Deep Learning for Time Series Forecasting: Advances and Open Problems. Information. 2023 Nov 4;14(11):598.
3. Lim B, Arik SO, Loeff N, Pfister T. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. 2019. Available from: <https://arxiv.org/abs/1912.09363>
4. Nie Y, Nguyen NH, Sinthong P, Kalagnanam J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. 2022. Available from: <https://arxiv.org/abs/2211.14730>
5. Wang G, Liao Y, Guo L, Geng J, Ma X. DLinear photovoltaic power generation forecasting based on reversible instance normalization. Proc. of 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS). Xiangtan, China: IEEE; 2023, p. 990–5.
6. George EP Box and Gwilym M Jenkins. Some recent advances in forecasting and control. Journal of the Royal Statistical Society. Series C (Applied Statistics), 17(2):91–109, 1968.
7. Athanasopoulos, G., Poskitt, D. S., & Vahid, F. (2012). Two canonical VARMA forms: Scalar component models vis-à-vis the echelon form. Econometric Reviews, 31(1), 60–83. DOI: <https://doi.org/10.1080/07474938.2011.607088>
8. Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2 <https://doi.org/10.7287/peerj.preprints.3190v2>
9. Ізонін І.В. Ансамблева технологія без вчителя - з вчителем з неітеративним алгоритмом навчання для аналізу коротких наборів біомедичних даних // Комп'ютерні системи та інформаційні технології, 2023, 67-74. DOI: <https://doi.org/10.31891/csit-2023-4-9>

References

1. Izonin I. An Ensemble Method for the Regression Model Parameter Adjustments: Direct Approach // Mathematical and computer modelling. Series: Technical sciences, 2023, 35-44. DOI: <https://doi.org/10.32626/2308-5916.2023-24.35-44>
2. Casolaro A, Capone V, Iannuzzo G, Camastra F. Deep Learning for Time Series Forecasting: Advances and Open Problems. Information. 2023 Nov 4;14(11):598.
3. Lim B, Arik SO, Loeff N, Pfister T. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. 2019. Available from: <https://arxiv.org/abs/1912.09363>
4. Nie Y, Nguyen NH, Sinthong P, Kalagnanam J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. 2022. Available from: <https://arxiv.org/abs/2211.14730>
5. Wang G, Liao Y, Guo L, Geng J, Ma X. DLinear photovoltaic power generation forecasting based on reversible instance normalization. Proc. of 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS). Xiangtan, China: IEEE; 2023, p. 990–5.
6. George EP Box and Gwilym M Jenkins. Some recent advances in forecasting and control. Journal of the Royal Statistical Society. Series C (Applied Statistics), 17(2):91–109, 1968.
7. Athanasopoulos, G., Poskitt, D. S., & Vahid, F. (2012). Two canonical VARMA forms: Scalar component models vis-à-vis the echelon form. Econometric Reviews, 31(1), 60–83. DOI: <https://doi.org/10.1080/07474938.2011.607088>
8. Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2 <https://doi.org/10.7287/peerj.preprints.3190v2>
9. Izonin I. An unsupervised-supervised ensemble technology with non-iterative training algorithm for small biomedical data analysis. // Computer systems and information technologies, 2023, 67-74. DOI: <https://doi.org/10.31891/csit-2023-4-9>