

TRANSFORMERS IN MULTIVARIATE TIME SERIES FORECASTING: A REVIEW

ESMAEIL CHITGAR^{OR}, MAHDIEH HASHEMINEZHAD^{OR}* AND SEPEHR EBRAHIMI MOOD^{OR}

ABSTRACT. Long-term forecasting of multivariate time series is a fundamental challenge in the field of machine learning, with critical applications in numerous domains such as energy, transportation, and financial markets. The inherent complexity of this data, stemming from seasonal patterns, non-stationary trends, and interdependencies among variables, has constrained traditional forecasting methods. The importance of addressing this problem is evident in strategic decision-making, as the accuracy and efficiency of predictions directly impact energy resource management, traffic control, and financial risk analysis.

This review article provides a comprehensive framework for analyzing state-of-the-art Transformer-based architectures in the domain of multivariate time series forecasting. The proposed framework is structured around three main aspects: (1) quantitative performance evaluation of models on standard datasets, (2) analysis of computational considerations including time and memory complexity, and (3) examination of the models' capability in handling high-frequency time series. Accordingly, prominent models such as Autoformer, iTransformer, GTformer, FEDformer, ETSformer, and Pathformer are reviewed and comparatively analyzed.

The findings of this review indicate that recent innovations, including the application of frequency-domain processing, auto-correlation mechanisms, graph-based learning, and multi-scale architectures, have simultaneously led to improvements in forecasting accuracy and computational efficiency. These advancements outline a clear research path for the development of future architectures with a focus on greater interpretability, better scalability, and broader generalization capabilities.

Keywords: Transformer, Multivariate, Encoder, Decoder, Time Series.

Article Type: Review Paper.

Communicated by Khalil Shafie.

Received: 14-04-2025, Accepted: 17-11-2025, Published Online: 28-01-2026.

Cite this article: E. Chitgar, M. Hasheminezhad and S. Ebrahimi Mood, Transformers in multivariate time series forecasting: a review, *Mathematics and Society*, **11** no. 2 (2026) 97–128. <https://dx.doi.org/10.22108/msci.2025.144922.1740>.

1. Introduction

Multivariate time series forecasting represents a fundamental challenge in machine learning, with essential applications in fields such as energy systems, transportation networks, and financial markets. The complexity of such data arises from seasonal patterns, non-stationary behaviors, and the intricate relationships among variables, which often limit the performance of traditional forecasting methods. Understanding and effectively solving this challenge is crucial for strategic decision-making because improvements in prediction accuracy and computational efficiency directly influence energy management, traffic optimization, and financial risk assessment.

The transformer architecture has emerged as one of the most important deep learning models specifically designed for sequence processing. Initially developed for machine translation, transformers demonstrate high capability in learning long-term dependencies due to their use of attention mechanisms, and have been rapidly employed in various domains including natural language processing, computer vision, and more recently in time series forecasting. Key applications of transformers include machine translation, text summarization, computer vision, and time series forecasting, where they show high efficiency due to their ability to understand long-term and cross-variable dependencies.

However, the application of transformers to multivariate time series forecasting faces significant challenges such as sensitivity to noise and outliers, high computational costs for long sequences, limited interpretability, and the difficulty of modeling complex nonlinear relationships. To overcome these obstacles and enhance prediction accuracy and efficiency, numerous research studies have been conducted in recent years, leading to the emergence and development of diverse and innovative transformer architectures.

This review article provides a comprehensive framework for analyzing advanced transformer-based architectures in multivariate time series forecasting. The proposed framework is structured around three main axes: quantitative performance evaluation of models based on standard datasets, analysis of computational considerations including time complexity, and examination of model capabilities in handling high-frequency time series. Accordingly, prominent models such as Autoformer, iTransformer, GTformer, FEDformer, ETSformer, and Pathformer are examined and compared.

This review systematically examines transformer-based models for multivariate time series forecasting, covering architectural fundamentals, key innovations, and comprehensive performance evaluation. The analysis concludes with future research directions in this rapidly evolving field.

2. Main Results

Our analysis reveals significant advancements in transformer-based approaches for multivariate time series forecasting. This section presents a multi-dimensional evaluation framework designed

to rigorously assess these models beyond conventional accuracy metrics, addressing critical gaps in existing literature.

2.1. Comprehensive Evaluation Framework. Our evaluation framework systematically assesses transformer models across three critical dimensions: predictive accuracy on standard benchmarks, computational efficiency, and performance on high-frequency time series data. This multi-faceted approach provides a more holistic understanding of model capabilities and limitations than traditional evaluation methods that focus solely on prediction accuracy. In most research studies, transformer-based models are investigated and compared mainly by using error metrics on standard datasets and models are ranked based on these results. Despite the value of such comparisons, relying solely on these criteria is insufficient for practical model selection. Numerous factors including computational considerations, architectural features of models, and also their performance in specific scenarios such as processing high-frequency time series or different sampling patterns play a determining role in the final selection.

Three main criteria are considered for evaluating models: numerical performance on standard datasets, computational efficiency, and performance evaluation on high-frequency time series. Finally, by synthesizing these results, we provide a roadmap for selecting models appropriate to specific application conditions.

2.2. Quantitative Performance Analysis. Comparative evaluation of time series forecasting models based on standard and well-known datasets enables fair and systematic comparison between different methods. Benchmark datasets that have been widely used in previous studies including the Electricity Transformer Temperature (ETT) dataset (hourly and minute-level), the Electricity Consumption (ECL) dataset, the Traffic dataset, the Weather dataset, the Influenza-like Illness (ILI) dataset, the Exchange Rate dataset, and the Solar Energy dataset.

For evaluating model performance, two common metrics, Mean Squared Error (MSE) and Mean Absolute Error (MAE) are generally used. Lower values in both metrics indicate higher model accuracy. The performance of selected transformer models that have been trained and evaluated on a subset of standard datasets is summarized in the comparison table. These data have been extracted from reliable sources and used under multivariate forecasting scenarios with different time horizons. For each dataset, values of both metrics have been calculated and their averages have been reported as final performance indices.

Evaluations have been conducted for four different forecasting horizons. The best performance for each horizon is indicated by underlining. Based on the overall average errors reported in the table, GTformer and Pathformer models show the best performance among the examined models with the lowest squared error and absolute error values, respectively. In contrast, the Autoformer

model shows the weakest performance with the highest error values across both measurement criteria. Other models including iTransformer, FEDformer, and ETSformer are ranked in the middle positions. However, this overall ranking tells only part of the story and for a complete understanding of each model's capabilities, more precise examination based on different criteria is necessary.

A more complete analysis of the obtained results includes: Overall prediction accuracy, Long-term forecasting, Computational efficiency, Interpretability, High-dimensional data and interdependencies, and Learning multi-scale and seasonal patterns. GTformer and Pathformer have the lowest average error in total and show that they are capable of capturing temporal and cross-series correlations. iTransformer has good performance in short horizons, but has slight accuracy drop in longer horizons. FEDformer and two decomposition-based models, i.e., ETSformer and Autoformer, usually yield higher errors, although they remain stable on some datasets.

GTformer and Pathformer models still provide satisfactory results in long-term horizons. FEDformer is also suitable for very long forecasting horizons due to computational efficiency and frequency architecture. In contrast, Autoformer and ETSformer usually have more error in longer horizons. In terms of computational efficiency, FEDformer is the most efficient model with near-linear complexity, suitable for large data. iTransformer is designed for high-dimensional data and consumes less memory. GTformer has higher computational cost but provides very good accuracy if sufficient resources are available. Pathformer establishes a good balance between accuracy and speed.

For domains that require explainability, ETSformer and Autoformer are the best options. These two models provide the possibility of more precise analysis with decomposition of trend and seasonal components. FEDformer is also somewhat explainable from the frequency perspective. GTformer clarifies cross-series relationships with graphs, though its explanation is more difficult for the end user. For data with a high number of variables or complex relationships, iTransformer and GTformer are superior choices. Other models have relative weakness in these conditions. Due to multi-scale architecture, Pathformer is the best option for data with complex seasonal patterns. FEDformer also has appropriate performance in this domain with reliance on frequency analysis. Autoformer and ETSformer focus more on trend/seasonal separation.

2.3. Computational Efficiency Analysis. One of the main challenges in employing transformer models for time series forecasting is their computational load and infrastructure requirements. These considerations are particularly important in industrial applications or when dealing with large-scale data. A review of the literature shows that each architecture has adopted specific strategies to reduce time complexity and optimize resource consumption:

Autoformer reduces time complexity from $\mathcal{O}(L^2)$ to $\mathcal{O}(L \log L)$ by replacing the standard self-attention with an auto-correlation mechanism. This model requires moderate to high computational resources.

iTransformer by inverting the standard architecture, considers individual time series as tokens. This design transfers attention calculation from temporal dimension to variable dimension and transforms the dominant complexity of the model to $\mathcal{O}(N^2 \cdot D)$. Since in many practical problems $N \ll L$, this model significantly reduces memory consumption and training time especially when dealing with high-dimensional data.

FEDformer achieves linear complexity $\mathcal{O}(L)$ using frequency domain analysis and random selection of Fourier modes. This characteristic makes it highly efficient for long-term forecasting with large-scale data.

GTformer has a computational complexity of $\mathcal{O}(M \cdot (L/P)^2 + M^2)$ and uses graph-based representation and sequence division into patches, thus providing acceptable efficiency for datasets with a large number of variables.

ETSformer combines exponential-smoothing attention mechanisms with frequency attention, achieving $\mathcal{O}(L \log L)$ time complexity. In terms of resource consumption, this model falls in the intermediate level, faster than standard transformers but heavier than linear models. This design enables the model to capture level, growth, and seasonality of data more accurately.

Pathformer by introducing adaptive pathways, gives the model the ability to dynamically select and compute only K scales from available M scales for each input sample. This mechanism prevents computational waste on irrelevant scales and leads to significant reduction in computational load and training time compared to fixed multi-scale architectures.

In general, it can be said that Autoformer and ETSformer balance between accuracy and computational cost, FEDformer and iTransformer have the most improvement in scalability and memory efficiency, GTformer is designed for heavy multivariate data, and Pathformer is a practical option for environments with resource constraints.

2.4. Architectural Innovations and Design Principles. Each of the evaluated models introduces unique architectural innovations that address specific challenges in time series forecasting:

Autoformer introduces an innovative architecture in the domain of long-term time series forecasting that is designed based on encoder-decoder structure. The main innovation of this model is that instead of standard attention, it employs series decomposition blocks and an auto-correlation mechanism. In each encoder layer, first the signal is decomposed into two components: trend and seasonal. The smoothed trend is forwarded and the seasonal part is analyzed by the auto-correlation mechanism to identify periodic patterns in historical sub-sequences. Then the feedforward network and residual

connections cause stability and improvement of representation. In the decoder, the trend part is gradually reconstructed and finally combined with the seasonal part to produce the final prediction.

iTransformer provides an innovative approach to employing transformer architecture for multivariate time series forecasting by introducing the concept of inverted transformer. The core of this innovation is changing the common approach in data representation. Unlike conventional methods where values of all variables at one time step are considered as one token, iTransformer considers each time series as an independent token. This fundamental change involves swapping the temporal and variable dimensions. In this architecture, the attention mechanism is used to understand correlations between different variables instead of modeling temporal dependencies, and directly records dependencies between variables.

GTformer, a graph-based temporal order-aware transformer, is designed for long-term forecasting of multivariate time series and specifically focuses on dependencies between series and strict temporal order of data. For learning relationships between time series, GTformer introduces an adaptive graph learning method that identifies unidirectional and bidirectional relationships between series. In this model, inputs are divided into patches and a temporal-order-aware unit produces positional mappings so that the order of moments in data is not lost. Following this, an adaptive graph learning layer records unidirectional and bidirectional relationships among variables.

FEDformer is an innovative architecture that overcomes limitations of conventional models in processing long-term time series by combining time series decomposition and frequency domain analysis. This model uses a mixture-of-experts architecture decomposition and employs Fourier transform or wavelet transform to transfer data from time domain to frequency domain. In FEDformer architecture, first the input signal is optimally decomposed into trend and seasonal parts using mixture of experts approach. Then the seasonal part is transferred to frequency domain and attention mechanism is applied only to a subset of dominant frequency components.

ETSformer is an interpretable model inspired by classical exponential smoothing methods, decomposes time series data into three interpretable components: level, growth, and seasonal. The core of this model is based on two innovative attention mechanisms: exponential smoothing attention that effectively models temporal dependencies with a gradual emphasis on more recent data, and frequency attention that automatically identifies repetitive seasonal patterns using Fourier transform. In this model's architecture, layer-wise decomposition blocks separately extract and reconstruct each of the three parts.

Pathformer is an advanced multi-scale transformer architecture designed for understanding temporal patterns at different levels. This model precisely analyzes temporal features at various scales and combines two key aspects: temporal resolution and temporal distance, to efficiently model complex patterns. This model intelligently divides the time series into patches of different sizes and uses

two types of specialized attention: inter-patch attention that analyzes connections between different patches, and intra-patch attention that examines details within each patch.

2.5. Performance on High-Frequency Time Series. In the domain of time series forecasting, one of the less explored aspects is the impact of sampling patterns on the performance of transformer models. In many real applications including high-speed financial systems and IoT sensor networks, data is generated with very high frequency. This unique characteristic creates fundamental questions about the impact of sampling methods on the forecasting accuracy of models. Two main approaches exist in sampling high-frequency data: domain increase and infill.

This difference in sampling patterns significantly affects the statistical characteristics of data. For example, in some conditions high-frequency series suffer from induced negative correlations that can make forecasting more difficult. Studies have also shown that asymptotic characteristics of estimators can exhibit different behaviors in these two cases. It should be noted that most introduced standard datasets are at medium frequency level and do not reflect actual very high-frequency conditions. Therefore, the results of these studies cannot be fully generalized to high-frequency data and need for more theoretical and conceptual analyses exists.

Model performance in facing these challenges is as follows: Autoformer shows separation of trend and seasonal components and stable performance in datasets with strong autocorrelation, though higher error has been reported in some scenarios. iTransformer shows high scalability for data with large numbers of variables and suitable performance in short-term predictions. GTformer learns interdependencies with graph structure, strong performance in transformer and electricity datasets, and suitability for noisy conditions and complex cross-series dependencies. FEDformer shows high computational efficiency with near-linear complexity and suitability for large-scale data and long-term prediction lengths. ETSformer provides high interpretability through decomposition of level, growth, and seasonal components, suitable for domains such as finance and medicine that require explanation. Pathformer shows multi-scale modeling capability and complex seasonal patterns, with strong performance in weather datasets and some transformer datasets.

The comprehensive summary of capabilities and limitations of these models is presented in a comparison table. The results of this analysis show that optimal model selection should not only be based on intrinsic characteristics of data, but also should consider sampling structure and existence of negative correlations. Understanding these complex relationships can lead to more informed model selection and improvement of prediction accuracy in practical applications.

2.6. Interpretability and Practical Applications. Interpretability remains a critical consideration, particularly in sensitive domains such as healthcare, finance, and energy management where

understanding model decisions is as important as prediction accuracy. The evaluated models offer varying levels of interpretability through different architectural approaches.

ETSformer and Autoformer provide the highest level of interpretability through explicit decomposition of time series into meaningful components. This decomposition allows domain experts to understand which components are driving the forecasts and provides valuable insights for decision-making processes. The clear separation of trend and seasonal components in Autoformer, and level, growth, and seasonal components in ETSformer, enables straightforward analysis of temporal patterns and their contributions to future predictions.

FEDformer offers interpretability from a frequency perspective, enabling analysis of which frequency components contribute most significantly to the predictions. This frequency-based interpretability is particularly valuable for applications with strong seasonal characteristics and periodic patterns, allowing domain experts to identify dominant frequencies and their impact on forecasting outcomes.

GTformer provides insights through its learned graph structure, revealing relationships between different variables in multivariate time series. The adaptive graph learning mechanism identifies meaningful connections between variables, supporting domain knowledge validation and discovery of previously unknown relationships in complex systems.

Pathformer's multi-scale analysis reveals which temporal scales are most informative for different forecasting tasks. The adaptive pathway selection reveals the relevance of different time scales for specific prediction scenarios, providing insights into temporal pattern hierarchies and their relative importance.

iTransformer adds value by focusing on cross-variable relationships, which can be particularly useful in applications where understanding variable interactions is important. The attention mechanisms highlight dependencies between different time series.

The practical applications of these transformer models span numerous critical domains. In energy management, they enable accurate load forecasting, grid optimization, and resource allocation. In financial markets, they support risk assessment, trading strategies, and market analysis. In healthcare, they facilitate patient monitoring, disease progression prediction, and medical decision support. In transportation systems, they improve traffic flow optimization, demand forecasting, and infrastructure planning. In industrial applications, they enable predictive maintenance, process optimization, and quality control.

3. Conclusions

This comprehensive review demonstrates that transformer-based architectures have significantly advanced the state-of-the-art in multivariate time series forecasting, offering substantial improvements over traditional statistical methods and earlier deep learning approaches. The architectural innovations

introduced in models such as Autoformer, iTransformer, GTformer, FEDformer, ETSformer, and Pathformer have addressed key limitations of standard transformers while leveraging their strengths in capturing long-range dependencies.

The synthesis of this study indicates that recent innovations, including the application of frequency domain processing, auto-correlation mechanisms, graph-based learning, and multi-scale architectures, have simultaneously led to improvements in both prediction accuracy and computational efficiency. These advancements chart a clear research path for developing future architectures with focus on higher interpretability, better scalability, and broader generalization capabilities.

The key conclusions from our analysis can be summarized as follows: First, no single architecture emerges as universally superior across all evaluation dimensions. The optimal model choice depends on specific application requirements, including forecasting horizon, data characteristics, computational constraints, and interpretability needs. This highlights the importance of context-aware model selection and the value of having diverse architectural options for different scenarios.

Second, the most significant advances have come from incorporating domain-specific inductive biases into the transformer architecture. Models that explicitly model temporal patterns such as seasonality, trends, and periodicity generally outperform generic transformer architectures. These domain adaptations enable more efficient learning of temporal patterns and improve forecasting accuracy across various applications.

Third, computational efficiency remains a critical consideration, particularly for long sequences and real-time applications. The evolution from quadratic to linear or log-linear complexity represents a major advancement enabling practical deployment in resource-constrained environments. The various efficiency optimization strategies have made transformer-based forecasting feasible for large-scale applications.

Fourth, the handling of high-frequency time series presents unique challenges that require specialized approaches. The differences between sampling strategies and phenomena such as induced negative correlations necessitate careful model selection and potentially domain-specific adaptations. The varying model performance across different sampling scenarios underscores the need for continued research in high-frequency time series forecasting.

Several important challenges and research directions remain open, providing opportunities for future investigations: The integration of time-domain and frequency-domain approaches represents a promising direction for future research, potentially enabling more robust handling of both short-term patterns from dense sampling and long-term patterns from extended observation windows. Developing models with greater robustness to the statistical peculiarities of high-frequency data would significantly enhance practical applicability.

The creation of more adaptive architectures that can automatically adjust their processing strategies based on data characteristics, sampling patterns, and application requirements would reduce the need for manual model selection and hyperparameter tuning. Improving interpretability while maintaining competitive performance remains an important challenge, particularly for applications in regulated domains or scenarios requiring human oversight.

The development of more comprehensive benchmark datasets that better reflect real-world high-frequency conditions would enable more meaningful evaluation and comparison of forecasting models. Standardized evaluation protocols and diverse application scenarios would facilitate rigorous assessment of model capabilities and support reproducible research in this rapidly evolving field.

In conclusion, transformer-based architectures have fundamentally transformed the landscape of multivariate time series forecasting. The continued innovation in this field promises even more powerful, efficient, and interpretable forecasting models that will enable better decision-making across numerous application domains. The diversity of architectural approaches and the ongoing methodological advancements ensure that transformer-based models will remain at the forefront of time series forecasting research and applications, driving progress in energy management, financial forecasting, healthcare monitoring, intelligent transportation systems, and beyond.

Esmail Chitgar

Department of Computer Science, Faculty of Mathematical Sciences, Yazd University, Yazd, Iran

Email: ec.chitgar@gmail.com

Mahdieh Hasheminezhad

Department of Computer Science, Faculty of Mathematical Sciences, Yazd University, Yazd, Iran

Email: hasheminezhad@yazd.ac.ir

Sepehr Ebrahimi Mood

Department of Computer Science, Faculty of Mathematical Sciences, Yazd University, Yazd, Iran

Email: s.ebrahimi@yazd.ac.ir

مروری بر مدل‌های ترنسفورمر برای پیش‌بینی سری‌های زمانی چندمتغیره

اسمعیل چیت‌گر^۱، مهدیه هاشمی‌نژاد^{۲*} و سپهر ابراهیمی‌مود^۳

چکیده. پیش‌بینی بلندمدت سری‌های زمانی چندمتغیره، یک چالش بنیادین در حوزه یادگیری ماشین به شمار می‌رود که کاربردهای حیاتی در حوزه‌های متعددی همچون انرژی، حمل‌ونقل و بازارهای مالی دارد. پیچیدگی ذاتی این داده‌ها که ناشی از وجود الگوهای فصلی، روندهای ناپایدار و وابستگی‌های متقابل میان متغیرهاست، روش‌های سنتی پیش‌بینی را با محدودیت مواجه ساخته است. اهمیت حل این مساله در تصمیم‌گیری‌های راهبردی آشکار می‌شود، چرا که دقت و کارایی پیش‌بینی‌ها به طور مستقیم بر مدیریت منابع انرژی، کنترل ترافیک و تحلیل ریسک‌های مالی تاثیر می‌گذارد.

این مقاله مروری، چارچوبی جامع برای تحلیل معماری‌های پیشرفته مبتنی بر ترنسفورمر در حوزه پیش‌بینی سری‌های زمانی چندمتغیره ارائه می‌دهد. چارچوب پیشنهادی بر سه محور اصلی بنا شده است: (۱) ارزیابی کمی عملکرد مدل‌ها بر اساس مجموعه داده‌های استاندارد، (۲) تحلیل ملاحظات محاسباتی شامل پیچیدگی زمانی و (۳) واکاوی توانایی مدل‌ها در مواجهه با سری‌های زمانی با فراوانی بالا. بر این اساس، مدل‌های شاخصی از قبیل اتوفورمر، آی ترنسفورمر، جی تی فورمر، فدفورمر، ای تی اس فورمر و پث فورمر مورد بررسی و مقایسه تطبیقی قرار گرفته‌اند.

جمع‌بندی این مطالعه نشان می‌دهد که نوآوری‌های اخیر، از جمله به‌کارگیری پردازش در حوزه فرکانس، مکانیزم‌های خودهمبستگی، یادگیری مبتنی بر گراف و معماری‌های چندمقیاسی، توانسته‌اند به طور همزمان به بهبود دقت پیش‌بینی و کارایی محاسباتی منجر شوند. این پیشرفت‌ها، مسیر پژوهشی روشنی را برای توسعه معماری‌های آینده با تمرکز بر تفسیرپذیری بالاتر، مقیاس‌پذیری بهتر و قابلیت تعمیم گسترده‌تر ترسیم می‌کنند.

۱. مقدمه

ترنسفورمر^۱ یکی از مهم‌ترین مدل‌های یادگیری عمیق است [۱۵، ۲۸] که به‌طور خاص برای پردازش توالی‌ها، طراحی شده است. این مدل که ابتدا برای ترجمه ماشینی ارائه شد، به دلیل بهره‌گیری از مکانیزم توجه^۲، توانایی بالایی در یادگیری وابستگی‌های طولانی مدت دارد و به‌سرعت در حوزه‌های مختلف مانند پردازش زبان طبیعی^۳، بینایی کامپیوتر^۴، و اخیراً در پیش‌بینی سری‌های زمانی به‌کار گرفته شده است [۵، ۱۹، ۲۰]. برخی از کاربردهای اصلی ترنسفورمرها عبارت‌اند از:

- ترجمه ماشینی [۲۵، ۳۵]: یکی از اولین کاربردهای موفق ترنسفورمر، ترجمه متون از یک زبان به زبان دیگر بوده است.

عبارات و کلمات کلیدی: ترنسفورمر، چندمتغیره، رمزگذار، رمزگشا، سری‌های زمانی، معماری تجزیه، مکانیزم توجه، همبستگی خودکار نوع مقاله: مروری

دبیرتخصصی رابط: خلیل شفیعی

تاریخ دریافت: ۱۴۰۴/۰۱/۲۵ تاریخ پذیرش: ۱۴۰۴/۰۹/۲۳ تاریخ انتشار آنلاین: ۱۴۰۴/۱۱/۰۸

ارجاع به مقاله: ا. چیت‌گر، م. هاشمی‌نژاد و س. ابراهیمی‌مود، مروری بر مدل‌های ترنسفورمر برای پیش‌بینی سری‌های زمانی چندمتغیره، ریاضی و جامعه، ۱۱ شماره ۲ (۱۴۰۵) ۹۷-۱۲۸.

<http://dx.doi.org/10.22108/msci.2025.144922.1740>

¹transformer ²attention mechanism ³NLP ⁴computer vision

- خلاصه‌سازی متون [۷]: توانایی ترنسفورمر در یادگیری روابط بین کلمات در متون طولانی، آن‌را به ابزاری مناسب برای خلاصه‌سازی خودکار متون تبدیل کرده است.
- بینایی کامپیوتر [۸]: استفاده از ترنسفورمرها برای تحلیل و درک داده‌های تصویری در سال‌های اخیر رشد کرده است.
- پیش‌بینی سری‌های زمانی [۲۹]: در پیش‌بینی داده‌های پیوسته و وابسته به زمان، ترنسفورمرها به دلیل قابلیت درک وابستگی‌های بلندمدت و بین‌متغیری، کارایی بالایی دارند.

با وجود مزایای متعدد، به‌کارگیری ترنسفورمرها در پیش‌بینی سری‌های زمانی چندمتغیره با چالش‌های مهمی همراه است. از جمله این چالش‌ها می‌توان به مدل‌سازی همزمان وابستگی‌های کوتاه‌مدت و بلندمدت، حساسیت به اختلال^۵ و داده‌های پرت، هزینه محاسباتی بالا برای توالی‌های طولانی، مشکلات تفسیرپذیری نتایج و همچنین مدل‌سازی روابط پیچیده و غیرخطی میان متغیرهای مختلف اشاره کرد. برای غلبه بر این موانع و افزایش دقت و کارایی پیش‌بینی، در سال‌های اخیر پژوهش‌های متعددی انجام شده که منجر به ظهور و توسعه معماری‌های متنوع و نوآورانه‌ای از ترنسفورمرها شده است.

ساختار این مقاله به‌شرح زیر است: در بخش ۲، مبانی و معماری ترنسفورمر اولیه [۲۷] مرور می‌شود. بخش ۳ به بررسی مزایا و محدودیت‌های اصلی این مدل‌ها می‌پردازد. در بخش ۴، معماری‌های شاخص و نوآورانه‌ای که برای پیش‌بینی سری‌های زمانی چندمتغیره ارائه شده‌اند، معرفی می‌شوند. بخش ۵ به ارزیابی جامع این مدل‌ها اختصاص دارد و آن‌ها را از جنبه‌های مختلفی از جمله دقت پیش‌بینی (بر مبنای معیارهای عددی)، ملاحظات محاسباتی و نیز کارایی در مواجهه با سری‌های زمانی با فراوانی بالا^۶ مقایسه می‌کند. پیشینه تحقیقاتی این حوزه در بخش ۶ و در سه حوزه بهبود دقت، کاهش هزینه‌های محاسباتی و معرفی معماری‌های جدید مرور می‌شود. در نهایت، بخش ۷ به جمع‌بندی یافته‌ها و ارائه چشم‌اندازها و جهت‌گیری‌های تحقیقاتی برای آینده می‌پردازد.

۲. ترنسفورمر اولیه

ترنسفورمر اولیه [۲۷] از دو بخش اصلی رمزگذار و رمزگشا تشکیل شده است. رمزگذار ورودی‌ها را پردازش می‌کند و رمزگشا با بهره‌گیری از خروجی رمزگذار، توالی خروجی را تولید می‌کند. این دو بخش شامل چندین لایه متوالی هستند که هر یک از این لایه‌ها از اجزای زیر تشکیل شده است:

نگاشت کلمات. داده‌های ورودی به‌صورت بردارهای متنی (کلمات یا نشانه‌ها^۷) به ترنسفورمر اولیه وارد می‌شود. در موضوع پیش‌بینی سری‌های زمانی، این ورودی‌ها می‌توانند ماهیت‌های متفاوتی داشته باشند. برای نمونه، یک سری زمانی تک‌متغیره^۸ شامل مقادیر مشاهده‌شده از یک ویژگی منفرد مانند دما در طول زمان است، در حالی که در سری‌های چندمتغیره^۹ چند ویژگی مرتبط نظیر دما، رطوبت و فشار هوا به‌صورت هم‌زمان ثبت می‌شوند. از منظر آماری نیز برخی سری‌ها، ایستا^{۱۰} هستند، به این معنا که ویژگی‌هایی مانند میانگین و واریانس^{۱۱} آن‌ها در طول زمان تغییر نمی‌کند، اما در سری‌های نایستا^{۱۲} این ویژگی‌ها ثابت نبوده و اغلب روندها یا الگوهای فصلی مشاهده می‌شود. شناخت این خصوصیات نقش مهمی در تعیین سطح پیچیدگی مساله و همچنین انتخاب معماری مناسب برای مدل ایفا می‌کند.

با این حال، مدل ترنسفورمر نمی‌تواند به‌طور مستقیم با داده‌های متنی کار کند، از این رو از نگاشت کلمات^{۱۳} استفاده می‌شود که در آن هر کلمه یا نشانه به یک بردار عددی در فضای برداری متناظر تبدیل می‌شود. برای این تبدیل از فرمول (۱) استفاده می‌گردد:

$$(1) \quad E = W_e \cdot x$$

⁵noise ⁶high-frequency ⁷token ⁸uni-variate ⁹multi-variate ¹⁰stationary ¹¹variance ¹²non-stationary ¹³word embedding

رمزگذاری موقعیتی. ترنسفورمر به‌طور طبیعی ترتیب توالی ورودی را نمی‌داند. برای افزودن اطلاعات مربوط به موقعیت کلمات، رمزگذاری^{۱۴} موقعیتی^{۱۵} به نگاشت اضافه می‌شود. این رمزگذاری به‌صورت یک بردار با ابعاد مشابه نگاشت کلمات است که به هر ورودی اضافه می‌شود.

فرمول محاسبه‌ی رمزگذاری موقعیتی در ترنسفورمر اولیه به‌صورت (۲) است:

$$(2) \quad PE(pos, k) = \begin{cases} \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right), & k = 2i \\ \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right), & k = 2i + 1 \end{cases}$$

که در آن pos موقعیت کلمه در دنباله و d_{model} ابعاد بردار نگاشت است.

مکانیزم توجه چندسره. مکانیزم توجه، هسته‌ی اصلی ترنسفورمر است که به مدل اجازه می‌دهد به بخش‌های مختلف ورودی با وزن‌های متفاوت نگاه کند، به‌طوری‌که وابستگی‌های مهم‌تر تقویت و وابستگی‌های کم‌اهمیت تضعیف شوند. در ترنسفورمرها، معمولاً از خودتوجهی^{۱۶} استفاده می‌شود، به این معنی که هر عنصر ورودی به سایر عناصر ورودی توجه می‌کند تا روابط بین آنها را مدل‌سازی کند.

در توجه چندسره^{۱۷}، چندین مکانیزم خودتوجهی به‌صورت موازی اجرا می‌شوند تا اطلاعات از زوایای مختلف بررسی و ترکیب شوند، که باعث می‌شود مدل توانایی یادگیری روابط پیچیده‌تر بین عناصر ورودی را داشته باشد. سه بردار مهم در توجه چندسره وجود دارند:

- Q ^{۱۸}: نشان‌دهنده کلمه‌ای که به دنبال اطلاعاتی درباره سایر کلمات است.
 - K ^{۱۹}: نشان‌دهنده کلمه‌ای است که ممکن است اطلاعات مهمی داشته باشد.
 - V ^{۲۰}: نشان‌دهنده اطلاعات کلمه است که در نهایت در محاسبات استفاده می‌شود.
- این بردارها از طریق ضرب ماتریسی با ماتریس‌های وزنی از روش فرمول (۳) به‌دست می‌آیند:

$$(3) \quad Q = W_q \cdot x, \quad K = W_k \cdot x, \quad V = W_v \cdot x$$

عملیات توجه از طریق فرمول (۴) محاسبه می‌شود:

$$(4) \quad Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

در این فرمول d_k ابعاد بردارهای K است.

لایه‌ی نرمال‌سازی. بعد از هر مکانیزم توجه، یک لایه‌ی نرمال‌سازی^{۲۱} قرار می‌گیرد که خروجی را نرمال می‌کند. نرمال‌سازی کمک می‌کند تا مدل پایداری بیشتری داشته باشد و باعث همگرایی گرادینان^{۲۲} خواهد شد. فرمول نرمال‌سازی به‌صورت (۵) است

$$(5) \quad \hat{x} = \frac{x - \mu}{\sigma}$$

که در آن μ میانگین و σ انحراف معیار مقادیر ورودی است.

شبکه‌ی عصبی پیش‌خور. بعد از مکانیزم توجه، یک شبکه‌ی عصبی پیش‌خور^{۲۳} وجود دارد. این شبکه برای یادگیری وابستگی‌های غیرخطی استفاده می‌شود و به هر ورودی به‌صورت نقطه‌ای اعمال می‌شود.

¹⁴encoding ¹⁵positional ¹⁶self attention ¹⁷multi-head attention ¹⁸query ¹⁹key ²⁰value ²¹layer normalization ²²gradient
²³feedforward neural network

ارتباط رمزگذار و رمزگشا. رمزگذار^{۲۴} در ترنسفورمر، ورودی‌ها را به یک نمایش پنهان تبدیل می‌کند. این نمایش به رمزگشا^{۲۵} منتقل می‌شود که با استفاده از مکانیزم توجه، وابستگی بین نمایش پنهان ورودی و خروجی فعلی را برقرار می‌کند. در رمزگشا، توجه به دو نوع تقسیم می‌شود:

- **خودتوجهی:** عملکردی مشابه خودتوجهی در رمزگذار داشته و برای پیدا کردن وابستگی‌های بین کلمات خروجی به کار می‌رود.
- **توجه رمزگذار به رمزگشا:** توجه به خروجی رمزگذار داشته و ارتباط بین اطلاعات ورودی و خروجی را برای مدل‌سازی روابط پیچیده برقرار می‌کند.

خروجی نهایی. در نهایت، خروجی رمزگشا از طریق یک لایه‌ی بیشینه‌ی هموار^{۲۶} به احتمال پیش‌بینی‌های نهایی تبدیل می‌شود. این احتمال‌ها نشان می‌دهند که چه کلمه‌ای با بالاترین احتمال به عنوان خروجی در هر مرحله تولید شود.

۳. مزایا و معایب مدل‌های ترنسفورمر

ترنسفورمرها، مانند سایر مدل‌ها، دارای مزایا و معایبی [۲۹] هستند که در ادامه آمده‌اند:

۱.۳. مزایای مدل‌های ترنسفورمر.

- **توانایی درک وابستگی‌ها و تعاملات بلندمدت:** این ویژگی به‌ویژه برای مدل‌سازی سری‌های زمانی اهمیت دارد و به ترنسفورمرها اجازه می‌دهد الگوهای پیچیده داده‌های ترتیبی را شناسایی کنند.
- **موفقیت در مدل‌سازی ویژگی‌های روند و فصلی:** ترنسفورمرها در مدیریت ویژگی‌های کلیدی سری‌های زمانی، مانند روند^{۲۷} و فصلی^{۲۸} بودن عملکرد خوبی داشته‌اند، این ویژگی برای تحلیل دقیق سری‌های زمانی بسیار اهمیت دارد.
- **تطبیق‌پذیری و تغییرات سفارشی:** تغییرات و تطبیق‌های متعددی در ترنسفورمرها اعمال شده است تا با چالش‌های خاص مدل‌سازی سری‌های زمانی مقابله کرده و آن‌ها را برای وظایف مختلف منعطف سازد.

۲.۳. معایب مدل‌های ترنسفورمر.

- **چالش در مدل‌سازی هم‌زمان وابستگی‌های کوتاه‌مدت و بلندمدت:** مدیریت هم‌زمان وابستگی‌های کوتاه‌مدت و بلندمدت و شناسایی ویژگی‌های فصلی همچنان یکی از چالش‌های اصلی مدل‌سازی سری‌های زمانی است.
- **پیچیدگی در مدل‌سازی سری‌های زمانی:** با وجود موفقیت ترنسفورمرها، کاربرد آن‌ها در سری‌های زمانی هنوز در حال توسعه است و بهینه‌سازی عملکرد آن‌ها برای این حوزه با چالش‌هایی مواجه است.

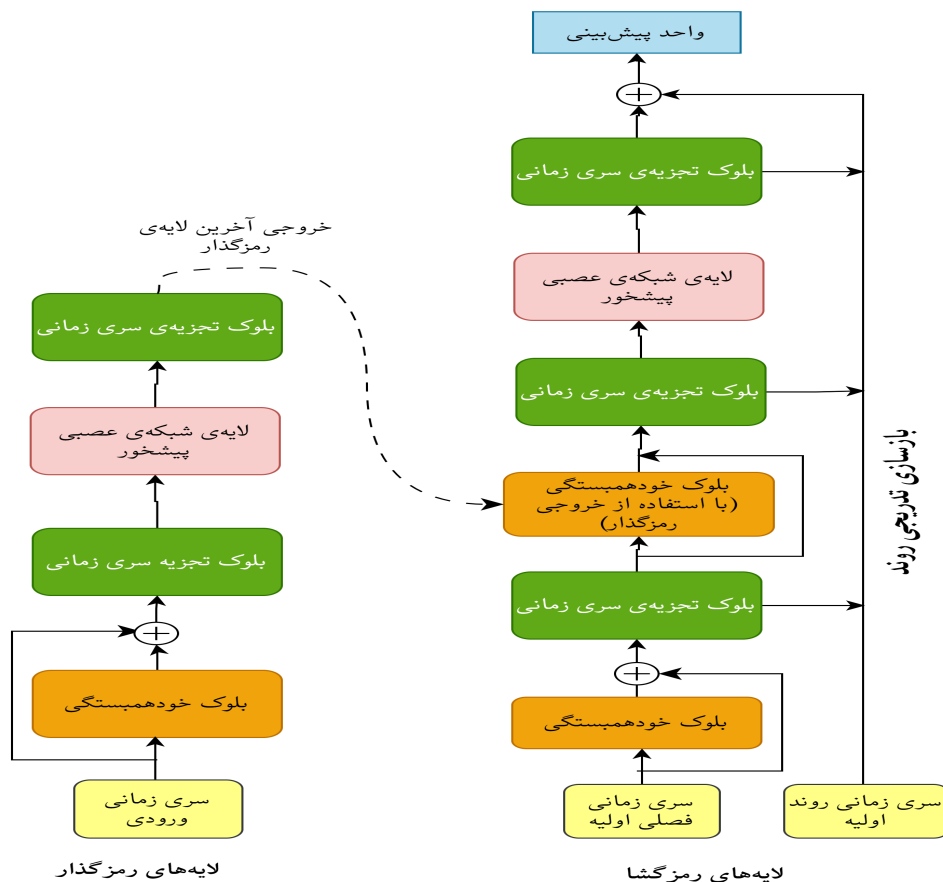
۴. ترنسفورمرهای سری زمانی

در این بخش، چند مدل ترنسفورمر شاخص در حوزه پیش‌بینی سری‌های زمانی، با تمرکز بر رویکردها و نوآوری‌هایشان، معرفی می‌شوند.

۱.۴. اتوفورمر. اتوفورمر^{۲۹} [۱] یک معماری نوآورانه در حوزه‌ی پیش‌بینی بلندمدت سری‌های زمانی است که بر پایه‌ی ساختار رمزگذار-رمزگشا طراحی شده است. نوآوری اصلی این مدل آن است که به‌جای استفاده از توجه استاندارد، از بلوک تجزیه‌ی^{۳۰} سری‌ها و مکانیزم خودهمبستگی^{۳۱} استفاده کرده است. در هر لایه‌ی رمزگذار ابتدا سیگنال به دو بخش روند و فصلی تجزیه می‌شود. روند هموار شده، و به جلو منتقل می‌شود و بخش فصلی توسط مکانیزم خودهمبستگی تحلیل می‌شود تا الگوهای تناوبی در زیردنباله‌های تاریخی شناسایی شوند. سپس شبکه‌ی پیشخور و اتصالات باقیمانده باعث پایداری و بهبود نمایش می‌شوند. در

²⁴encoder ²⁵decoder ²⁶softmax ²⁷trend ²⁸seasonal ²⁹Autoformer ³⁰decomposition ³¹auto-correlation

رمزگشا، بخش روند به تدریج بازسازی شده و در نهایت با بخش فصلی ترکیب می‌شود تا پیش‌بینی نهایی تولید گردد. این طراحی هم قابلیت تفسیرپذیری مدل را افزایش می‌دهد و هم پیچیدگی محاسباتی را کاهش می‌دهد. در شکل ۱ معماری این مدل نشان داده شده است.



شکل ۱. معماری کلی اتوفرمر: ورودی ابتدا در بلوک تجزیه سری‌ها به روند و فصلی تقسیم می‌شود. بخش فصلی توسط مکانیزم خودهمبستگی پردازش شده و بخش روند به‌طور تدریجی در رمزگشا و رمزگذار می‌شود. خروجی نهایی حاصل جمع دو بخش روند و فصلی است.

Figure 1: Overall architecture of autoformer: Input is first decomposed into trend and seasonal components in the series decomposition block. The seasonal component is processed by the auto-correlation mechanism, and the trend component is gradually reconstructed in the encoder and decoder. The final output is the sum of trend and seasonal components.

۲.۴. آی ترنسفورمر. آی ترنسفورمر^{۳۲} [۱۶] با معرفی مفهوم ترنسفورمر معکوس^{۳۳}، رویکردی نوآورانه در به‌کارگیری معماری ترنسفورمر برای پیش‌بینی سری‌های زمانی چندمتغیره ارائه می‌دهد. هسته‌ی این نوآوری، تغییر رویکرد رایج در نمایش داده‌ها است. برخلاف روش‌های مرسوم که در آن‌ها مقادیر همه متغیرها در یک گام زمانی به‌عنوان یک نشانه در نظر گرفته می‌شوند،

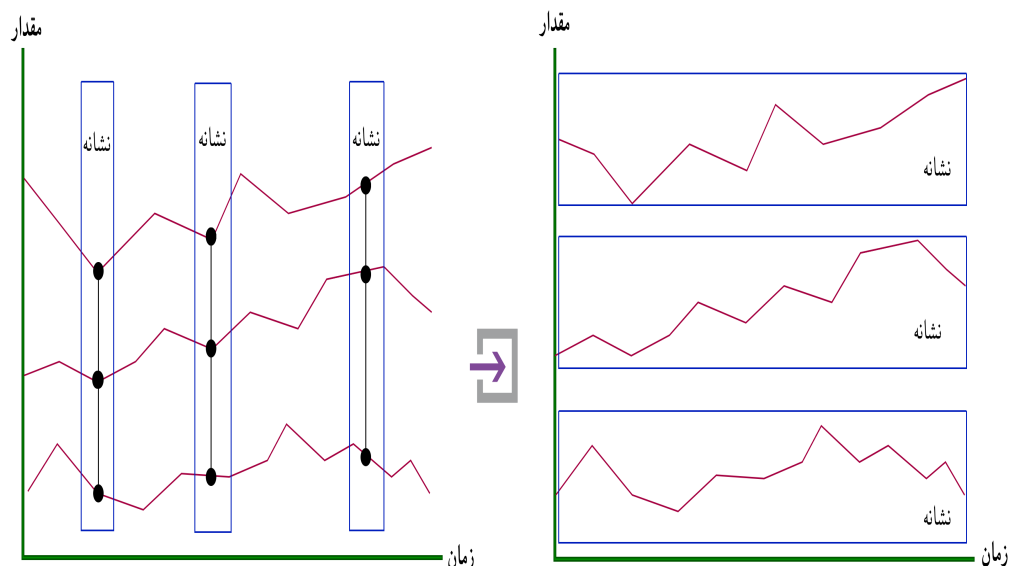
³²iTransformer ³³inverted transformer

آی‌ترنسفورمر هر سری زمانی (هر متغیر) را به‌عنوان یک نشانه‌ی مستقل در نظر می‌گیرد. این تغییر اساسی به معنای جایگزینی ابعاد زمانی و متغیر است.

در این معماری، مکانیزم توجه به‌جای مدل‌سازی وابستگی‌های زمانی، برای درک هم‌بستگی‌های بین متغیرهای مختلف به‌کار می‌رود و وابستگی‌های میان متغیرها را مستقیماً ثبت می‌کند. سپس یک شبکه‌ی پیشخور مستقل روی نشانه هر متغیر اعمال می‌شود تا نمایش‌های غیرخطی مخصوص به هر سری زمانی را یاد بگیرد. این رویکرد متغیرمحور که تمرکز آن بر نمایش متغیرهاست، حتی برای پنجره‌های زمانی بسیار بلند نیز اثربخش است.

معماری آی‌ترنسفورمر منحصراً از یک رمزگذار تشکیل شده که از مکانیزم خودتوجهی و شبکه‌های پیشخور استاندارد بهره می‌برد. نداشت به‌صورت متغیرمحور انجام می‌شود و به دلیل ماهیت طراحی، نیاز به کدگذاری موقعیت مرسوم وجود ندارد، زیرا ترتیب و ساختار متغیرها به‌طور ضمنی در معماری مدل گنجانده شده است.

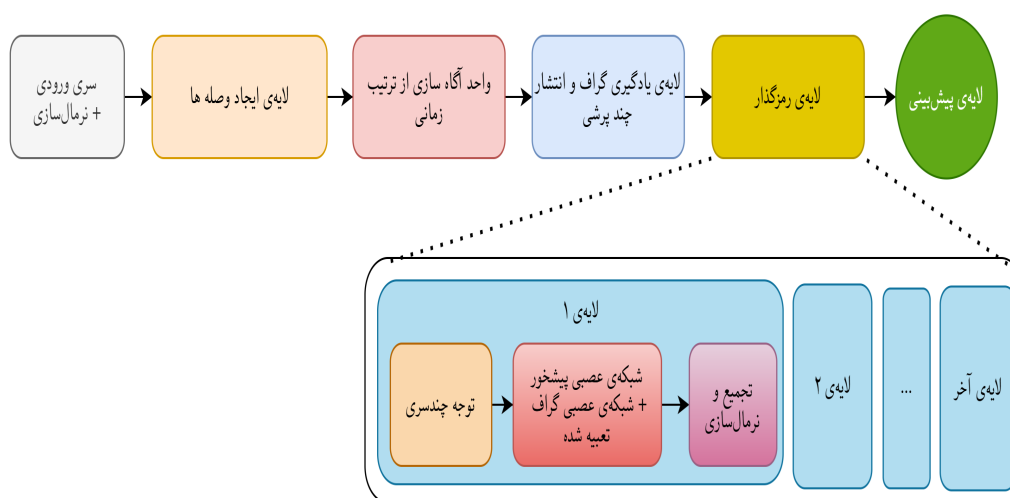
این طراحی انعطاف‌پذیر، موجب بهبود عملکرد مدل در پیش‌بینی‌های چندمتغیره، به‌ویژه در بازه‌های زمانی طولانی می‌شود. آی‌ترنسفورمر عملکردی رقابتی بر روی مجموعه‌داده‌های واقعی نشان داده و ثابت کرده که می‌توان با دیدگاهی متفاوت و با استفاده از اجزای پایه‌ای ترنسفورمر، به پیش‌بینی موثر سری‌های زمانی پرداخت. همچنین، این مدل در برابر تغییر در تعداد متغیرهای ورودی انعطاف‌پذیری بالایی دارد. تفاوت رویکرد مدل در نحوه‌ی انتخاب نشانه در شکل ۲ نشان داده شده است.



شکل ۲. مقایسه رویکردهای نشانه‌گذاری در پیش‌بینی سری‌های زمانی: در روش مرسوم (تصویر چپ) هر نشانه نماینده یک مقطع زمانی است، در حالی که در آی‌ترنسفورمر (تصویر راست) هر نشانه متناظر با یک سری زمانی کامل است. این تغییر رویکرد امکان مدل‌سازی مستقیم وابستگی‌های بین متغیرها را فراهم می‌کند.

Figure 2: Comparison of tokenization approaches in time series forecasting: In the conventional method (left image) each token represents a time point, while in iTransformer (right image) each token corresponds to a complete time series. This approach change enables direct modeling of dependencies between variables.

۳.۴. جی‌تی‌فورمر. جی‌تی‌فورمر^{۳۴} [۱۱]، ترنسفورمری آگاه از ترتیب زمانی مبتنی بر گراف^{۳۵}، برای پیش‌بینی بلندمدت سری‌های زمانی چندمتغیره طراحی شده و به‌طور ویژه بر وابستگی میان سری‌ها و ترتیب زمانی دقیق^{۳۶} داده‌ها تمرکز دارد. برای یادگیری روابط بین سری‌های زمانی، جی‌تی‌فورمر یک روش یادگیری گراف تطبیقی^{۳۷} معرفی می‌کند که روابط یک‌سویه و دوسویه را بین سری‌ها شناسایی می‌کند. به‌عنوان مثال، در یک سناریوی پیش‌بینی وضعیت هوا، این مدل می‌تواند گرافی را یاد بگیرد که روابط بین دما، رطوبت و سرعت باد را به تصویر بکشد. در این مدل، ورودی‌ها به وصله‌هایی^{۳۸} تقسیم می‌شوند و یک واحد آگاه از ترتیب زمانی، نگاشت‌های موقعیتی را تولید می‌کند تا ترتیب لحظات در داده‌ها از دست نرود. در ادامه، یک لایه یادگیری گراف تطبیقی روابط یک‌سویه و دوسویه میان متغیرها را ثبت می‌کند. رمزگذار ترکیبی از لایه‌های ترنسفورمر و انتشار گرافی^{۳۹} است که امکان بهره‌گیری از همسایگی‌های چندپرسی^{۴۰} را فراهم می‌آورد. نتیجه‌ی این طراحی، توانایی بالاتر در پیش‌بینی دقیق سری‌های چندمتغیره است. با ترکیب دو بخش یادگیری گراف و کدگذاری موقعیتی ترتیبی، مدل قادر است تا وابستگی‌های بین‌سری و درون‌سری را بهتر درک کرده و دقت پیش‌بینی را افزایش دهد. در شکل ۳ می‌توان معماری کلی این مدل را ملاحظه کرد.



شکل ۳. معماری جی‌تی‌فورمر: ابتدا داده‌ها به وصله‌هایی تقسیم می‌شوند، سپس واحد آگاه از ترتیب زمانی نگاشت‌های موقعیتی را ایجاد می‌کند. لایه‌ی گراف تطبیقی روابط متغیرها را یاد می‌گیرد و رمزگذار (شامل ترنسفورمر و انتشار گرافی)، اطلاعات همسایگان چندپرسی را ترکیب می‌کند.

Figure 3: Architecture of GTformer: First, data is divided into patches, then the temporal-order-aware unit creates positional embeddings. The adaptive graph learning layer learns variable relationships, and the encoder (including transformer and graph propagation) combines multi-hop neighbor information.

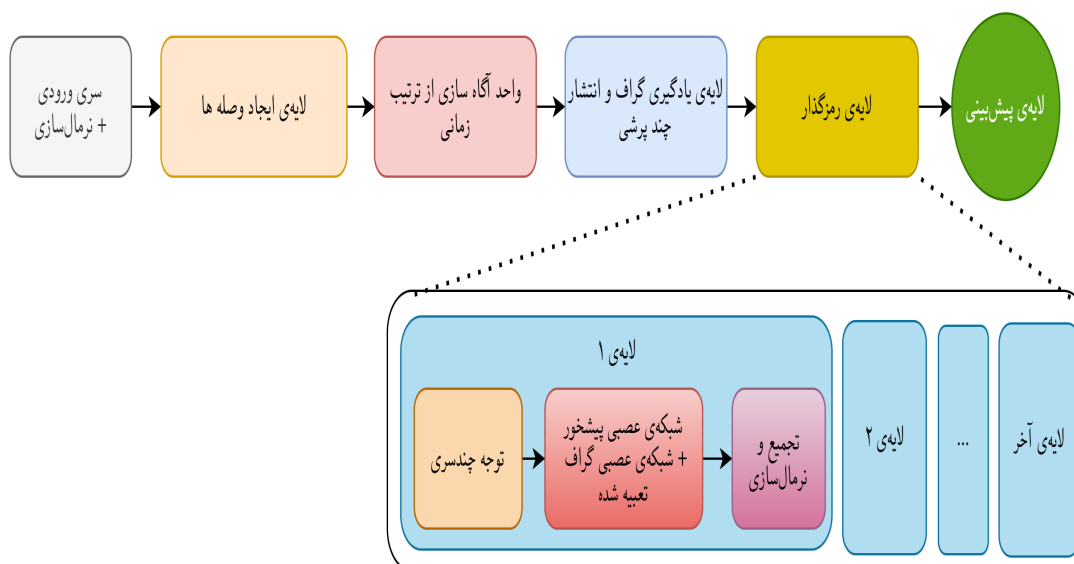
۴.۴. فدفورمر. فدفورمر^{۴۱} [۲۳] یک معماری نوآورانه است که با ترکیب تجزیه‌ی سری‌های زمانی و تحلیل در حوزه‌ی فرکانس^{۴۲}، محدودیت‌های مدل‌های مرسوم در پردازش سری‌های زمانی بلندمدت را برطرف می‌کند. این مدل از معماری مبتنی بر

^{۳۴}GTformer ^{۳۵}graph-based temporal order aware ^{۳۶}strict temporal order ^{۳۷}adaptive graph learning ^{۳۸}patches ^{۳۹}graph prppagation ^{۴۰}multi-hp neighbors ^{۴۱}FEDformer ^{۴۲}frequency

ترکیب متخصصان^{۴۳} برای تجزیه‌ی موثر سری‌های زمانی استفاده می‌کند و از تبدیل فوریه^{۴۴} یا تبدیل موجک^{۴۵} برای انتقال داده‌ها از حوزه‌ی زمان به حوزه‌ی فرکانس بهره می‌برد.

در معماری فدفورمر، ابتدا سیگنال ورودی با استفاده از رویکرد ترکیب متخصصان به بخش‌های روند و فصلی بهینه تجزیه می‌شود. سپس بخش فصلی به حوزه‌ی فرکانس منتقل شده و مکانیزم توجه بر روی زیرمجموعه‌ای از اجزای فرکانسی غالب اعمال می‌گردد. این روش، امکان شناسایی دقیق ویژگی‌های فصلی و روندهای بلندمدت را فراهم می‌کند و همزمان پیچیدگی محاسباتی را تا مرتبه خطی کاهش می‌دهد.

ترکیب معماری ترکیب متخصصان با تحلیل فرکانسی به گونه‌ای طراحی شده است که توانایی مدل در شناسایی الگوهای پیچیده و روابط غیرخطی را به‌طور قابل توجهی افزایش می‌دهد. شکل ۴ طرح کلی این معماری را نشان می‌دهد.



شکل ۴. معماری فدفورمر: سیگنال ورودی با استفاده از معماری ترکیب متخصصان به بخش‌های روند و فصلی تفکیک شده، بخش فصلی به حوزه‌ی فرکانس منتقل می‌شود. مکانیزم توجه بر روی فرکانس‌های منتخب اعمال شده و نتایج با بخش روند ترکیب می‌شوند تا پیش‌بینی نهایی تولید شود.

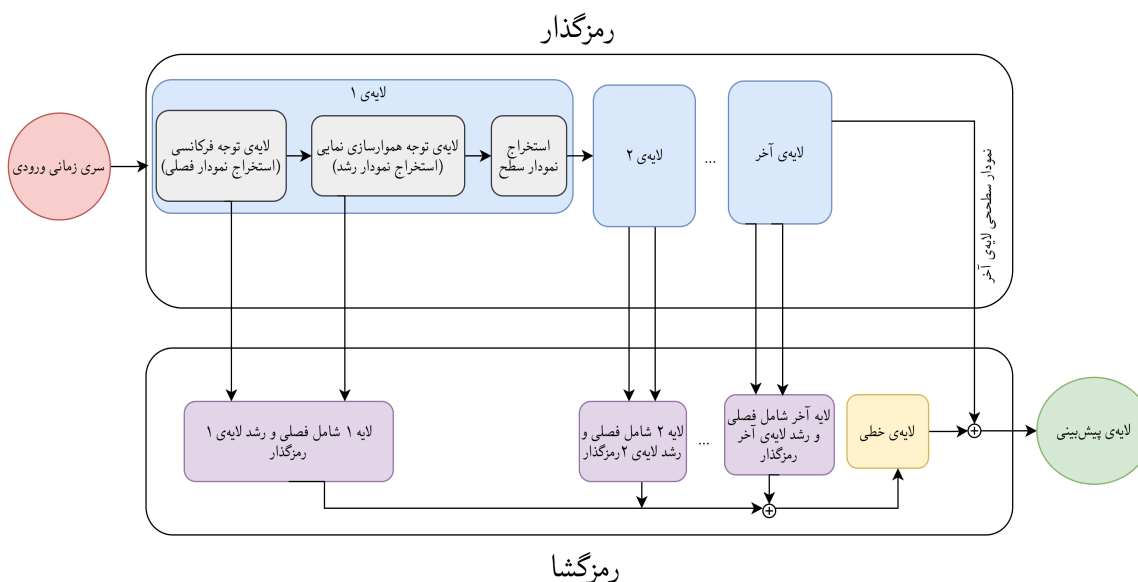
Figure 4: Architecture of FEDformer: The input signal is decomposed into trend and seasonal components using mixture of experts architecture. The seasonal component is transferred to the frequency domain, attention mechanism is applied to selected frequencies, and results are combined with the trend component to generate the final prediction.

۵.۴. ای‌تی‌اس‌فورمر. ای‌تی‌اس‌فورمر^{۴۶} [۳۱] یک مدل تفسیرپذیر است که با الهام از روش‌های کلاسیک هموارسازی^{۴۷} طراحی شده و داده‌های سری زمانی را به سه بخش اصلی و قابل فهم شامل سطح^{۴۸}، رشد^{۴۹} و فصلی تقسیم می‌کند. هسته‌ی اصلی این مدل بر دو مکانیزم توجه نوآورانه استوار شده است:

- توجه هموارسازی نمایی: که با تمرکز تدریجی بر داده‌های جدیدتر، وابستگی‌های زمانی را به‌طور مؤثر مدل‌سازی می‌کند.

⁴³mixture of experts ⁴⁴fourier transform ⁴⁵wavelet transform ⁴⁶ETSformer ⁴⁷exponential smoothing ⁴⁸level ⁴⁹growth

• **توجه فرکانسی:** که با بهره‌گیری از تبدیل فوریه، الگوهای تکراری فصلی را به‌صورت خودکار شناسایی می‌نماید. در معماری این مدل، بلوک‌های تجزیه‌ی لایه‌به‌لایه به‌طور جداگانه هر یک از سه بخش را استخراج و بازسازی می‌کنند. توجه هموارسازی نمایی وظیفه‌ی مدل‌سازی بخش رشد و توجه فرکانسی مسئول شناسایی الگوهای فصلی است. در نهایت، پیش‌بینی نهایی از ترکیب هوشمندانه‌ی این سه بخش حاصل می‌شود. این رویکرد نه تنها دقت پیش‌بینی را نسبت به ترنسفورمرهای استاندارد افزایش می‌دهد، بلکه با ارائه‌ی تجزیه‌ی واضح هر بخش، تفسیرپذیری نتایج را نیز به‌طور چشمگیری بهبود می‌بخشد. شکل ۵ طرح کلی این معماری را نشان می‌دهد.



شکل ۵. معماری ای‌تی‌اس‌فورمر: ورودی به سه بخش سطح، رشد و فصلی تجزیه می‌شود. بلوک‌های تجزیه لایه‌به‌لایه هر بخش را پردازش کرده و توجه هموارسازی نمایی و توجه فرکانسی به‌طور ویژه برای رشد و فصلی استفاده می‌شوند. خروجی نهایی از ترکیب سه بخش ساخته می‌شود.

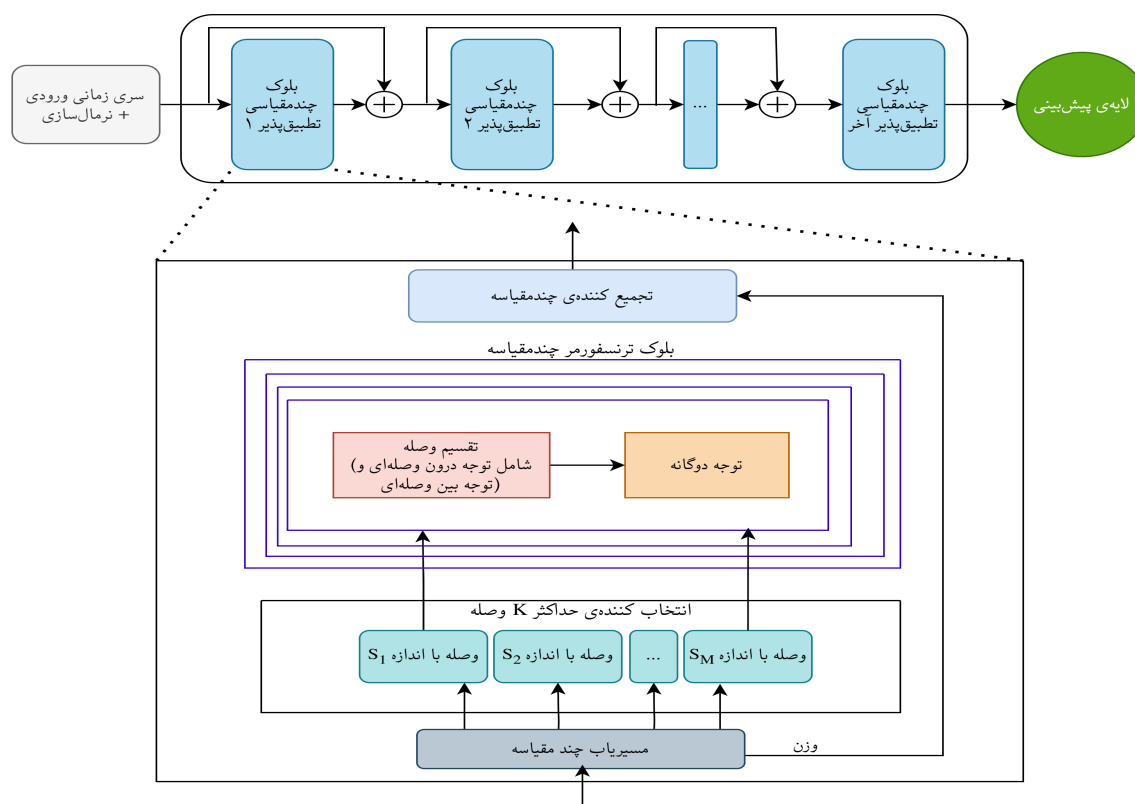
Figure 5: Architecture of ETSformer: Input is decomposed into three main components: level, growth, and seasonal. Layer-wise decomposition blocks process each component separately, with exponential smoothing attention and frequency attention specifically used for growth and seasonal components. The final prediction results from the intelligent combination of these three components.

۶.۴. پث‌فورمر. پث‌فورمر^{۵۰} [۲] یک معماری ترنسفورمر پیشرفته و چندمقیاسی^{۵۱} است که برای درک الگوهای زمانی در سطوح مختلف طراحی شده است. این مدل با تحلیل دقیق ویژگی‌های زمانی در مقیاس‌های گوناگون و ترکیب دو جنبه‌ی کلیدی، یعنی وضوح زمانی^{۵۲} و فاصله زمانی^{۵۳}، الگوهای پیچیده را به‌طور کارآمد مدل‌سازی می‌کند. این مدل با استفاده از یک روش تقسیم‌بندی هوشمند، سری زمانی را به وصله‌هایی با اندازه‌های مختلف (مانند بازه‌های روزانه، هفتگی و ماهانه) تقسیم می‌کند. این رویکرد به مدل اجازه می‌دهد تا هم‌زمان الگوهای کوتاه‌مدت و بلندمدت را شناسایی کند. برای پردازش این وصله‌ها، پث‌فورمر از دو نوع توجه تخصصی استفاده می‌کند:

⁵⁰Pathformer ⁵¹multi scale ⁵²temporal resolution ⁵³temporal distance

- توجه بین‌وصله‌ای^{۵۴}: ارتباطات میان وصله‌های مختلف را تحلیل می‌کند
- توجه درون‌وصله‌ای^{۵۵}: به بررسی جزئیات درون هر وصله می‌پردازد

ویژگی برجسته‌ی این مدل، انتخاب پویای اندازه‌های بهینه‌ی وصله‌ها توسط یک مسیریاب^{۵۶} چندمقیاسی است. این مسیریاب با تحلیل ویژگی‌هایی مانند روندها و الگوهای فصلی، بهترین مقیاس‌ها را مشخص می‌کند. در نهایت، اطلاعات پردازش‌شده از طریق یک تجمیع‌کننده چندمقیاسی^{۵۷} با هم ترکیب می‌شوند. این معماری پیشرفته توانسته است محدودیت‌های مدل‌های تک‌مقیاسی را برطرف کند و دقت پیش‌بینی را به میزان قابل توجهی بهبود بخشد. شکل ۶ ساختار کلی این مدل را نمایش می‌دهد.



شکل ۶. معماری پث‌فورمر: ورودی به وصله‌های چندمقیاسی تقسیم شده و توسط مکانیزم توجه دوگانه پردازش می‌شود. مسیریاب چندمقیاسی مسیرهای بهینه را انتخاب و تجمیع‌کننده، خروجی نهایی را تولید می‌کند.

Figure 6: Architecture of Pathformer: Input is divided into multi-scale patches of different sizes (such as daily, weekly, and monthly intervals). Dual attention mechanisms (inter-patch and intra-patch) process these patches. The multi-scale router dynamically selects optimal patch sizes, and the multi-scale aggregator combines the processed information to produce the final output.

⁵⁴inter patch ⁵⁵intra patch ⁵⁶router ⁵⁷multi scale aggregator

۵. ارزیابی جامع مدل‌های منتخب

در بیشتر پژوهش‌ها، برای بررسی و مقایسه‌ی مدل‌های مبتنی بر معماری ترنسفورمر، صرفاً از معیارهای محاسبه‌ی خطا روی مجموعه داده‌های استاندارد استفاده شده و مدل‌ها بر اساس همین نتایج رتبه‌بندی می‌شوند. با وجود ارزشمندی این گونه مقایسه‌ها، باید توجه داشت که در عمل، انتخاب یک مدل مناسب تنها بر اساس چنین معیارهایی کافی نیست. عوامل متعددی از قبیل ملاحظات محاسباتی، ویژگی‌های معماری مدل‌ها و همچنین عملکرد آن‌ها در سناریوهای خاص نظیر پردازش سری‌های زمانی با فراوانی بالا یا الگوهای نمونه‌گیری متفاوت، نقش تعیین‌کننده‌ای در انتخاب نهایی ایفا می‌کنند. به‌عنوان مثال ممکن است مدلی در رتبه‌ی نخست بر مبنای معیارهای عمومی عملکرد قرار گیرد، ولی برای کاربردی خاص نامناسب باشد. برای نمونه از قابلیت تفسیرپذیری پایینی برخوردار باشد یا به دلیل هزینه‌ی محاسباتی بسیار بالا، مستلزم استفاده از منابع سخت‌افزاری پرهزینه باشد. در این فصل سه معیار اصلی برای ارزیابی مدل‌ها در نظر گرفته می‌شود:

(۱) عملکرد عددی بر روی مجموعه داده‌های استاندارد

(۲) کارایی محاسباتی

(۳) ارزیابی عملکرد بر روی سری‌های زمانی با فراوانی بالا

در پایان، با جمع‌بندی این نتایج، یک نقشه‌ی راه برای انتخاب مدل متناسب با شرایط کاربردی ارائه خواهد شد.

۱.۵. ارزیابی عددی مدل‌ها بر پایه مجموعه داده‌های استاندارد. ارزیابی مقایسه‌ای مدل‌های پیش‌بینی سری‌زمانی بر پایه مجموعه داده‌های استاندارد و شناخته‌شده، امکان مقایسه‌ی عادلانه و نظام‌مند بین روش‌های مختلف را فراهم می‌کند. مجموعه داده‌های معتبری که به‌طور گسترده در مطالعات پیشین مورد استفاده قرار گرفته‌اند عبارت‌اند از:

- دمای ترانسفورماتور برق^{۵۸} در دو واحد زمانی ساعتی^{۵۹} و دقیقه‌ای^{۶۰}
- مصرف الکتریسیته^{۶۱}
- داده‌های ترافیکی^{۶۲}
- اطلاعات آب‌وهوایی^{۶۳}
- آمار بیماری‌های شبه آنفولانزا^{۶۴}
- نرخ تبادل ارز^{۶۵}
- تولید انرژی خورشیدی^{۶۶}

برای ارزیابی عملکرد مدل‌ها، عموماً از دو معیار رایج میانگین مربعات خطا^{۶۷} و میانگین قدر مطلق خطا^{۶۸} استفاده می‌گردد. مقادیر پایین‌تر در هر دو معیار، دلالت بر دقت بالاتر مدل دارد. عملکرد مدل‌های ترنسفورمر منتخب که بر روی یک زیرمجموعه از مجموعه داده‌های استاندارد آموزش دیده و ارزیابی شده‌اند، در جدول ۱ خلاصه شده است. این داده‌ها از منابع معتبر استخراج و در چارچوب سناریوهای پیش‌بینی چندمتغیره با افق‌های^{۶۹} زمانی مختلف مورد استفاده قرار گرفته‌اند. برای هر مجموعه داده، مقادیر هر دو معیار محاسبه و میانگین آن‌ها به‌عنوان شاخص نهایی عملکرد گزارش شده است. ارزیابی‌ها برای چهار افق پیش‌بینی مختلف (۹۶، ۱۹۲، ۳۳۶ و ۷۲۰ گام زمانی) انجام شده است. بهترین عملکرد برای هر افق (کمترین مقدار خطا) با خط زیرین^{۷۰} در هر ردیف مشخص گردیده است.

⁵⁸electricity transformer temperature (ETT) dataset ⁵⁹ETTh1, ETTh2 ⁶⁰ETTm1, ETTm2 ⁶¹electricity (ECL) dataset

⁶²traffic dataset ⁶³weather dataset ⁶⁴influenza-like illness (ILI) dataset ⁶⁵exchange rate dataset ⁶⁶solar energy dataset

⁶⁷mean squared error (MSE) ⁶⁸mean absolute error (MAE) ⁶⁹horizons ⁷⁰underline

جدول ۱. ارزیابی مدل‌های ترنسفورمر بر روی مجموعه‌داده‌های استاندارد با دو معیار میانگین مربعات و قدرمطلق خطا

Table 1: Evaluation of transformer models on standard datasets with two metrics: mean squared error and mean absolute error

۱. اتوفرمر	۲. آی ترنسفورمر	۳. جی تی فورمر	۴. فدفورمر	۵. ای تی اس فورمر	۶. پت فورمر	خطا بر حسب میانگین
مربع قدرمطلق	مربع قدرمطلق	مربع قدرمطلق	مربع قدرمطلق	مربع قدرمطلق	مربع قدرمطلق	
ترانسفورماتور (دقیقه نوع ۱)						
۰.۴۷۵	۰.۳۳۴	۰.۳۶۸	۰.۳۱۱	۰.۳۷۹	۰.۴۱۹	۹۶
۰.۵۰۵	۰.۳۳۴	۰.۳۶۸	۰.۳۱۱	۰.۳۷۹	۰.۴۱۹	۱۹۲
۰.۵۵۳	۰.۳۷۷	۰.۳۹۱	۰.۳۶۳	۰.۴۲۶	۰.۴۴۱	۳۳۶
۰.۶۲۱	۰.۴۲۶	۰.۴۲۰	۰.۳۹۳	۰.۴۴۵	۰.۴۵۹	۷۲۰
۰.۶۷۱	۰.۴۹۱	۰.۴۵۹	۰.۴۵۱	۰.۵۴۳	۰.۴۹۰	
۰.۴۶۲	۰.۴۹۹	۰.۴۶۲	۰.۴۹۹	۰.۴۶۲	۰.۴۶۲	
ترانسفورماتور (دقیقه نوع ۲)						
۰.۲۵۵	۰.۳۳۹	۰.۱۸۰	۰.۱۷۳	۰.۲۵۲	۰.۲۰۳	۹۶
۰.۲۸۱	۰.۳۴۰	۰.۲۵۰	۰.۲۳۹	۰.۲۶۹	۰.۲۶۹	۱۹۲
۰.۳۳۹	۰.۳۷۲	۰.۳۱۱	۰.۳۴۸	۰.۳۳۶	۰.۳۲۵	۳۳۶
۰.۴۳۳	۰.۴۳۲	۰.۴۱۲	۰.۴۰۷	۰.۳۹۹	۰.۴۲۱	۷۲۰
۰.۴۱۴	۰.۴۱۲	۰.۴۱۲	۰.۴۱۲	۰.۴۱۴	۰.۴۱۴	
۰.۴۱۴	۰.۴۱۴	۰.۴۱۴	۰.۴۱۴	۰.۴۱۴	۰.۴۱۴	
ترانسفورماتور (ساعتی نوع ۱)						
۰.۴۴۹	۰.۴۵۹	۰.۳۸۶	۰.۳۷۵	۰.۳۷۶	۰.۴۱۹	۹۶
۰.۵۰۰	۰.۴۸۲	۰.۴۴۱	۰.۴۲۴	۰.۴۲۰	۰.۴۲۰	۱۹۲
۰.۵۲۱	۰.۴۹۶	۰.۴۸۷	۰.۴۶۵	۰.۴۵۹	۰.۴۶۵	۳۳۶
۰.۵۱۴	۰.۵۱۲	۰.۵۰۳	۰.۴۷۵	۰.۵۰۶	۰.۵۰۶	۷۲۰
۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	
۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	۰.۴۷۹	
ترانسفورماتور (ساعتی نوع ۲)						
۰.۳۴۶	۰.۳۸۸	۰.۲۹۷	۰.۲۸۲	۰.۳۵۸	۰.۳۹۷	۹۶
۰.۴۵۶	۰.۴۵۲	۰.۳۸۰	۰.۳۵۷	۰.۴۲۹	۰.۴۳۹	۱۹۲
۰.۴۸۲	۰.۴۸۶	۰.۴۲۸	۰.۴۳۲	۰.۴۹۶	۰.۴۸۷	۳۳۶
۰.۵۱۵	۰.۵۱۱	۰.۴۲۷	۰.۴۴۵	۰.۴۶۳	۰.۴۷۴	۷۲۰
۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	
۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	۰.۴۹۷	
الکتریسیته						
۰.۲۰۱	۰.۳۱۷	۰.۱۴۸	۰.۲۴۰	۰.۲۳۹	۰.۱۹۳	۹۶
۰.۲۲۲	۰.۳۳۴	۰.۱۶۲	۰.۲۵۳	۰.۲۵۰	۰.۲۰۱	۱۹۲
۰.۲۳۱	۰.۳۳۸	۰.۱۷۸	۰.۲۶۹	۰.۲۷۰	۰.۲۱۴	۳۳۶
۰.۲۵۴	۰.۳۶۱	۰.۲۲۵	۰.۳۱۷	۰.۲۸۹	۰.۲۴۶	۷۲۰
۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	
۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	۰.۳۴۵	
ترافیک						
۰.۶۱۳	۰.۳۸۸	۰.۳۹۵	۰.۲۶۸	۰.۳۷۰	۰.۵۸۷	۹۶
۰.۶۱۶	۰.۳۸۲	۰.۴۱۷	۰.۲۷۶	۰.۳۹۴	۰.۶۰۴	۱۹۲
۰.۶۲۲	۰.۳۳۷	۰.۴۳۳	۰.۲۸۳	۰.۴۱۲	۰.۶۲۱	۳۳۶
۰.۶۶۰	۰.۴۰۸	۰.۴۶۷	۰.۳۰۲	۰.۴۵۱	۰.۶۲۶	۷۲۰
۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	
۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	۰.۶۳۲	
آب و هوا						
۰.۲۶۶	۰.۳۳۶	۰.۱۷۴	۰.۲۱۴	۰.۱۶۰	۰.۲۱۷	۹۶
۰.۳۰۷	۰.۳۶۷	۰.۲۲۱	۰.۲۵۴	۰.۲۰۶	۰.۲۷۶	۱۹۲
۰.۳۵۹	۰.۳۹۵	۰.۲۷۸	۰.۲۹۶	۰.۲۶۳	۰.۳۳۹	۳۳۶
۰.۴۱۹	۰.۴۲۸	۰.۳۵۸	۰.۳۴۷	۰.۳۴۲	۰.۴۰۳	۷۲۰
۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	
۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	۰.۳۴۷	
میانگین کلی						
۰.۴۲۴	۰.۴۱۴	۰.۳۳۲	۰.۳۴۷	۰.۳۲۴	۰.۳۹۱	
۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	
۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	۰.۳۲۹	

بر اساس مقادیر میانگین کلی خطاهای محاسبه‌شده در جدول، مدل‌های جی‌تی‌فورمر و پث‌فورمر به‌ترتیب با کمترین مقادیر خطای مربعات و خطای قدرمطلق، بهترین عملکرد را در بین مدل‌های مورد بررسی نشان می‌دهند. در مقابل، مدل اتوفورمر با بیشترین مقادیر خطا در هر دو معیار سنجش، ضعیف‌ترین عملکرد را به خود اختصاص داده است. سایر مدل‌ها شامل آی‌ترنسفورمر، فدفورمر و ای‌تی‌اس‌فورمر در رتبه‌های میانی این رده‌بندی قرار گرفته‌اند. با این حال، این رتبه‌بندی کلی تنها بخشی از داستان را روایت می‌کند و برای درک کامل توانایی‌های هر مدل، لازم است بررسی دقیق‌تری براساس معیارهای مختلف صورت گیرد. در ادامه تحلیل کامل‌تری براساس نتایج به‌دست آمده ارائه می‌شود که عبارت‌اند از:

- **دقت پیش‌بینی کلی:** جی‌تی‌فورمر و پث‌فورمر در مجموع کمترین میانگین خطا را دارند و نشان می‌دهند که در ضبط همبستگی‌های زمانی و بین‌سری توانمند هستند. آی‌ترنسفورمر در افق‌های کوتاه عملکرد خوبی دارد، اما در افق‌های بلندتر کمی افت دقت دارد. فدفورمر و دو مدل مبتنی بر تجزیه یعنی ای‌تی‌اس‌فورمر و اتوفورمر معمولاً خطای بالاتری دارند، اگرچه در برخی مجموعه‌داده‌ها پایدارند.
- **پیش‌بینی بلندمدت:** مدل‌های جی‌تی‌فورمر و پث‌فورمر در افق‌های بلندمدت (بیش از ۳۳۶ گام) همچنان نتایج مطلوبی ارائه می‌دهند. فدفورمر نیز به دلیل کارایی محاسباتی و معماری فرکانسی برای طول‌های خیلی بلند مناسب است. در مقابل، اتوفورمر و ای‌تی‌اس‌فورمر معمولاً در افق‌های بلندتر خطای بیشتری دارند.
- **کارایی محاسباتی:**
 - فدفورمر: کارآمدترین مدل با پیچیدگی نزدیک به خطی، مناسب برای داده‌های حجیم.
 - آی‌ترنسفورمر: برای داده‌های با ابعاد بالا طراحی شده و حافظه‌ی کمتری مصرف می‌کند.
 - جی‌تی‌فورمر: هزینه‌ی محاسباتی بالاتری دارد ولی در صورت وجود منابع کافی، دقت بسیار خوبی ارائه می‌دهد.
 - پث‌فورمر: تعادل خوبی بین دقت و سرعت برقرار می‌کند.
- **تفسیرپذیری:** ای‌تی‌اس‌فورمر و اتوفورمر بهترین گزینه‌ها برای حوزه‌هایی هستند که نیازمند توضیح‌پذیری هستند. این دو مدل با تجزیه‌ی مؤلفه‌های روند و فصلی امکان تحلیل دقیق‌تری فراهم می‌کنند. فدفورمر نیز از منظر فرکانسی تا حدی تبیین‌پذیر است. جی‌تی‌فورمر با گراف‌ها روابط بین‌سری را روشن می‌کند، هرچند تبیین آن برای کاربر نهایی دشوارتر است.
- **داده‌های با ابعاد بالا و وابستگی‌های متقابل:** برای داده‌هایی با تعداد متغیر زیاد یا روابط پیچیده، آی‌ترنسفورمر و جی‌تی‌فورمر انتخاب‌های برتر هستند. سایر مدل‌ها در این شرایط ضعف نسبی دارند.
- **یادگیری الگوهای چندمقیاسی و فصلی:** پث‌فورمر به دلیل معماری چندمقیاسی، بهترین گزینه برای داده‌های دارای الگوهای فصلی پیچیده است. فدفورمر نیز با تکیه بر تحلیل فرکانسی در این حوزه عملکرد مناسبی دارد. اتوفورمر و ای‌تی‌اس‌فورمر بیشتر روی جداسازی روند/فصلی تمرکز دارند.

۲.۵. ارزیابی مدل‌ها بر اساس ملاحظات محاسباتی. یکی از چالش‌های اصلی در به‌کارگیری مدل‌های ترنسفورمر برای پیش‌بینی سری‌های زمانی، بار محاسباتی و نیازهای زیرساختی آن‌هاست. این ملاحظات به‌ویژه در کاربردهای صنعتی یا داده‌های با ابعاد بزرگ اهمیت بیشتری دارند. مرور مقالات نشان می‌دهد که هر معماری راهبرد خاصی برای کاهش پیچیدگی زمانی و بهینه‌سازی مصرف منابع اتخاذ کرده است:

- **اتوفورمر:** با جایگزینی مکانیزم خودهمبستگی به‌جای خودتوجهی استاندارد، پیچیدگی زمانی را از $O(L^2)$ به $O(L \log L)$ کاهش می‌دهد. این مدل نیازمند منابع محاسباتی در سطح متوسط تا بالا است.
- **آی‌ترنسفورمر:** با وارونه کردن معماری استاندارد، تک‌تک سری‌های زمانی (متغیرها) به‌عنوان توکن در نظر گرفته می‌شوند. این طراحی، محاسبه توجه را از بعد زمانی (L) به بعد متغیرها (N) منتقل می‌کند و پیچیدگی غالب مدل

را به $O(N^2 \cdot D)$ تبدیل می‌نماید. از آنجایی که در بسیاری از مسائل عملی $N \ll L$ است، این مدل مصرف حافظه و زمان آموزش را به‌ویژه در داده‌های با ابعاد بالا به‌طور چشمگیری کاهش می‌دهد.

- **فدفورمر:** با استفاده از تحلیل در حوزه فرکانس و انتخاب تصادفی مدهای فوریه، به پیچیدگی خطی $O(L)$ دست می‌یابد. این ویژگی آن را برای پیش‌بینی‌های بلندمدت با داده‌های حجیم بسیار کارآمد می‌سازد.
- **جی‌تی‌فورمر:** دارای پیچیدگی محاسباتی $O(M \cdot (L/P)^2 + M^2)$ است و در آن از نمایش مبتنی بر گراف و تقسیم توالی به وصله‌ها استفاده می‌گردد، بنابراین برای داده‌های با تعداد متغیر زیاد کارایی قابل‌قبولی دارد.
- **ای‌تی‌اس‌فورمر:** با ترکیب مکانیزم‌های توجه هموارسازی نمایی و توجه فرکانسی، پیچیدگی زمانی $O(L \log L)$ دارد. از نظر مصرف منابع، این مدل در سطح میانی قرار می‌گیرد، سریع‌تر از ترنسفورمرهای استاندارد اما سنگین‌تر از مدل‌های خطی. این طراحی به مدل امکان می‌دهد سطح، رشد و فصلی بودن داده‌ها را با دقت بیشتری مدل کند.
- **پث‌فورمر:** با معرفی مسیرهای انطباقی، به مدل این توانایی را می‌دهد که برای هر نمونه ورودی، تنها K مقیاس از بین M مقیاس موجود را به‌صورت پویا انتخاب و محاسبه کند. این مکانیزم از هدررفت محاسبات روی مقیاس‌های نامرتب جلوگیری کرده و در مقایسه با معماری‌های چندمقیاسی ثابت، منجر به کاهش محسوس بار محاسباتی و زمان آموزش می‌شود.

به‌طور کلی می‌توان گفت اتوفورمر و ای‌تی‌اس‌فورمر توازن میان دقت و هزینه محاسباتی برقرار می‌کنند، فدفورمر و آی‌ترنسفورمر بیشترین بهبود در مقیاس‌پذیری و کارایی حافظه را دارند، جی‌تی‌فورمر برای داده‌های چندمتغیره سنگین طراحی شده، و پث‌فورمر گزینه‌ای عملی برای محیط‌های با محدودیت منابع است.

۳.۵. ارزیابی مدل‌ها در مواجهه با سری‌های زمانی با فراوانی بالا. در حوزه پیش‌بینی سری‌زمانی، یکی از جنبه‌های کمتر مورد توجه قرار گرفته، تأثیر الگوهای نمونه‌گیری بر عملکرد مدل‌های ترنسفورمر است. در بسیاری از کاربردهای واقعی از جمله سیستم‌های مالی پرسرعت و شبکه‌های حسگر^{۷۱} اینترنت اشیاء^{۷۲}، داده‌ها با فرکانس بسیار بالا تولید می‌شوند. این ویژگی منحصر به فرد، پرسش‌های بنیادینی درباره‌ی تأثیر روش‌های نمونه‌گیری بر دقت پیش‌بینی مدل‌ها ایجاد می‌کند. دو رویکرد اصلی در نمونه‌گیری داده‌های با فراوانی بالا وجود دارد:

- **افزایش دامنه^{۷۳}:** گسترش تدریجی بازه‌ی زمانی مشاهده و جمع‌آوری داده‌ها از دامنه‌ی وسیع‌تر
- **پرش‌دگی^{۷۴}:** افزایش تراکم نمونه‌ها در بازه‌ی زمانی ثابت

این تفاوت در الگوی نمونه‌گیری، تأثیر معناداری بر ویژگی‌های آماری داده‌ها دارد. برای مثال، در برخی شرایط سری‌های با فراوانی بالا دچار همبستگی منفی القایی^{۷۵} می‌شوند [۴۱] که می‌تواند پیش‌بینی را دشوارتر سازد. همچنین مطالعاتی همچون [۴۰] نشان داده‌اند که ویژگی‌های مجانبی برآوردها در این دو حالت می‌تواند رفتار متفاوتی داشته باشد.

لازم است توجه شود که بیشتر مجموعه داده‌های استاندارد معرفی شده از نظر فراوانی در سطح متوسط قرار دارند و بازتاب‌دهنده‌ی شرایط واقعی فراوانی بسیار بالا نیستند. بنابراین نتایج این مطالعات را نمی‌توان به‌طور کامل به داده‌های با فراوانی بالا تعمیم داد و نیاز به تحلیل‌های نظری و مفهومی بیشتر وجود دارد. عملکرد مدل‌های ترنسفورمر در مواجهه با این چالش‌ها به شرح زیر است:

- **اتوفورمر:** جداسازی روند و فصلی و عملکرد پایدار در مجموعه داده‌هایی با خودهمبستگی قوی، هرچند در برخی سناریوها خطای بیشتری گزارش شده است
- **آی‌ترنسفورمر:** مقیاس‌پذیری بالا برای داده‌های با تعداد زیاد متغیر و عملکرد مناسب در پیش‌بینی‌های کوتاه‌مدت
- **جی‌تی‌فورمر:** یادگیری وابستگی‌های متقابل با ساختار گرافی، عملکرد قوی در مجموعه داده‌های ترانسفورماتور و برق و مناسب برای شرایط پر از اختلال و وابستگی‌های پیچیده بین سری

⁷¹sensor ⁷²internet of things ⁷³domain increase ⁷⁴infill ⁷⁵induced negative correlation

- **فدفورمر:** کارایی محاسباتی بالا با پیچیدگی نزدیک به خطی و مناسب برای داده‌های حجیم و طول‌های پیش‌بینی بلندمدت
 - **ای‌تی‌اس‌فورمر:** تفسیرپذیری بالا از طریق تجزیه سطح، رشد و فصلی و مناسب برای حوزه‌هایی مانند مالی و پزشکی که نیاز به توضیح دارند
 - **پثفورمر:** توانایی مدل‌سازی چندمقیاسی و الگوهای فصلی پیچیده و عملکرد قوی در مجموعه داده‌های آب‌وهوا و برخی مجموعه داده‌های ترانسفورماتور
- جدول ۲ خلاصه‌ای جامع از قابلیت‌ها و محدودیت‌های این مدل‌ها ارائه می‌دهد.

جدول ۲. ارزیابی عملکرد مدل‌های ترنسفورمر در داده‌های با فراوانی بالا

Table 2: Performance evaluation of transformer models on high-frequency time series

مدل	قابلیت‌های کلیدی	محدودیت‌ها
اتوفورمر	شناسایی تناوب‌های غالب با خودهمبستگی	آسیب‌پذیری در برابر تغییرات طیفی
آی‌ترنسفورمر	تمرکز بر روابط بین متغیرها	ناپایداری در داده‌های پرتراکم
جی‌تی‌فورمر	مدل‌سازی گرافی و حساسیت زمانی	افت جزئیات در وصله‌های بزرگ
فدفورمر	استخراج مؤلفه‌های تناوبی و فرکانس بالا	حساسیت به وابستگی‌های پیچیده در پرشدگی
ای‌تی‌اس‌فورمر	مقاومت در برابر اختلال با هموارسازی نمایی	محدودیت در شرایط پرشدگی
پثفورمر	رویکرد چندمقیاسی و توجه دوگانه	چالش در تبدیل فوریه تحت پرشدگی

نتایج این تحلیل نشان می‌دهد که انتخاب مدل بهینه نه تنها باید مبتنی بر ویژگی‌های ذاتی داده باشد، بلکه باید ساختار نمونه‌گیری و وجود همبستگی‌های منفی را نیز در نظر گیرد. درک این روابط پیچیده می‌تواند به انتخاب آگاهانه‌تر مدل‌ها و بهبود دقت پیش‌بینی در کاربردهای عملی منجر شود.

۴.۵. جمع‌بندی و نقشه راه انتخاب مدل. بررسی سه معیار اصلی (ارزیابی عددی، محاسباتی و سری‌های از منظر فراوانی بالا) نشان می‌دهد که معماری هیچ کدام از مدل‌ها بهترین مطلق نیست و هر یک نقاط قوت خاص خود را دارند. با توجه به مشاهدات و تحلیل‌های انجام‌شده، نقاط قوت کلیدی هر یک از مدل‌های ترنسفورمر به شرح زیر می‌باشد:

- **اتوفورمر:** جداسازی روند و فصلی و عملکرد پایدار در مجموعه داده‌هایی با خودهمبستگی قوی، هرچند در برخی سناریوها خطای بیشتری گزارش شده است.
- **آی‌ترنسفورمر:** مقیاس‌پذیری بالا برای داده‌های با متغیرهای بسیار زیاد، عملکرد مناسب در پیش‌بینی‌های کوتاه‌مدت.
- **جی‌تی‌فورمر:** یادگیری وابستگی‌های متقابل با ساختار گرافی، عملکرد قوی در مجموعه داده‌های ترانسفورماتور و برق، مناسب برای شرایط پر از اختلال و وابستگی‌های پیچیده بین‌سری.
- **فدفورمر:** کارایی محاسباتی بالا با پیچیدگی نزدیک به خطی، مناسب برای داده‌های حجیم و طول‌های پیش‌بینی بلندمدت.
- **ای‌تی‌اس‌فورمر:** تفسیرپذیری بالا از طریق تجزیه سطح، رشد و فصلی، مناسب برای حوزه‌هایی مانند مالی یا پزشکی که نیاز به تبیین دارند.
- **پثفورمر:** توانایی مدل‌سازی چندمقیاسی و الگوهای فصلی پیچیده، عملکرد قوی در مجموعه داده‌های آب‌وهوا و برخی از مجموعه داده‌های ترانسفورماتور.

جدول ۳ مقایسه‌ای از روش نشانه‌گذاری و ساختار معماری (رمزگذار/رمزگشا) این مدل‌ها ارائه می‌دهد. اهمیت این جدول در آن است که نشان می‌دهد تفاوت عملکرد مدل‌ها ریشه در طراحی معماری آن‌ها دارد. به‌عنوان مثال، آی‌ترنسفورمر تنها رمزگذار دارد و نشانه‌گذاری را بر اساس متغیر انجام می‌دهد، همین ویژگی دلیل برتری آن در داده‌های با ابعاد بالا است. در مقابل، پث‌فورمر با معماری چندمقیاسی رمزگذار-رمزگشا، برای داده‌های دارای الگوهای فصلی پیچیده مناسب‌تر است.

جدول ۳. مقایسه‌ی روش نشانه‌گذاری و نوع معماری در مدل‌های ترنسفورمر

Table 3: Comparison of tokenization methods and architecture types in transformer models

رمزگذار/رمزگشا	روش نشانه‌گذاری	مدل
رمزگذار-رمزگشا	براساس دنباله (خودهمبستگی)	اتوفورمر
فقط رمزگذار	بر اساس متغیر	آی‌ترنسفورمر
فقط رمزگذار	بخش‌بندی + گراف	جی‌تی‌فورمر
رمزگذار-رمزگشا	بخش‌های زمان-فرکانس	فدفورمر
رمزگذار-رمزگشا	سطح/رشد/فصلی	ای‌تی‌اس‌فورمر
رمزگذار-رمزگشا	بخش‌بندی چندمقیاسی	پث‌فورمر

بر این اساس، نقشه‌ی راه انتخاب مدل به شرح زیر است:

- (۱) نیاز به بیشترین دقت و منابع محاسباتی کافی: جی‌تی‌فورمر یا پث‌فورمر.
- (۲) داده‌های حجیم یا منابع محدود: فدفورمر یا پث‌فورمر.
- (۳) کاربردهای حساس به تبیین (مالی/پزشکی): ای‌تی‌اس‌فورمر یا اتوفورمر.
- (۴) داده‌های با ابعاد بالا با روابط پیچیده: آی‌ترنسفورمر یا جی‌تی‌فورمر.
- (۵) الگوهای چندمقیاسی و فصلی: پث‌فورمر یا فدفورمر.

این نقشه‌ی راه نشان می‌دهد که در نهایت، هیچ معماری به‌تنهایی بهترین مطلق نیست. انتخاب مناسب باید بر اساس ویژگی‌های داده، محدودیت منابع و هدف نهایی پژوهش صورت گیرد.

۶. پیشینه‌ی تحقیق

توسعه‌ی ترنسفورمرها در پیش‌بینی سری‌های زمانی را می‌توان به‌طور کلی براساس اهدافی که دنبال می‌کنند در ۳ بخش بهبود دقت، کاهش هزینه‌ی محاسباتی و معرفی معماری‌ها یا مکانیزم‌های جدید دسته‌بندی کرد.

۱.۶. بهبود دقت. تی‌اف‌تی^{۷۶} [۱۲] یک مدل پیش‌بینی چندافاق است که برای پردازش همزمان انواع مختلف داده‌های ورودی شامل متغیرهای ثابت (مانند شناسه یک فروشگاه)، داده‌های زمانی گذشته و داده‌های آینده از پیش مشخص طراحی شده است. این مدل با به‌کارگیری مکانیزم‌های انتخاب متغیر^{۷۷} و دروازه‌بندی نمونه‌آگاه^{۷۸}، ویژگی‌های مرتبط را در هر گام زمانی انتخاب می‌کند. همچنین با استفاده از یک رمزگشای مبتنی بر خودتوجهی تفسیرپذیر^{۷۹}، وابستگی‌های پیچیده زمانی کوتاه‌مدت و بلندمدت را مدل‌سازی می‌کند تا هم به دقت بالا و هم به قابلیت تفسیر برسد. اتوفورمر [۲۲] یک معماری مبتنی بر تجزیه ارائه می‌کند که اطلاعات روند بلندمدت را از متغیرهای پیش‌بینی‌شده جدا می‌سازد تا دقت پیش‌بینی سری‌های زمانی بلندمدت بهبود یابد. این مدل به‌جای خودتوجهی، از مکانیزم خودهمبستگی برای کشف شباهت زیرسری‌ها بر اساس تناوب استفاده می‌کند. فدفورمر [۳۹]

⁷⁶temporal fusion transformer ⁷⁷variable selection ⁷⁸instance-wise gating ⁷⁹interpretable self-attention decoder

بر تحلیل حوزه فرکانس تمرکز دارد و با ترکیب تجزیه روند- فصلی با ساختارهای تقویت شده‌ی فرکانسی، وابستگی‌های عمومی و محلی را به خوبی درک می‌کند. اس‌اس‌دی‌نت^{۸۰} [۱۳] با ترکیب ترنسفورمر با مدل‌های فضای حالت^{۸۱}، الگوهای زمانی را یاد می‌گیرد و پارامترهای آن را تخمین می‌زند. سپس، برای تجزیه‌ی روند- فصلی از مدل‌های فضای حالت استفاده کرده و قابلیت تفسیر مدل را افزایش می‌دهد. ترنسفورمر ناایستا [۱۷] موضوع بیش از حد ایستا کردن را در پیش‌بینی سری‌های زمانی حل می‌کند. همچنین با معرفی ماژول‌های ایستا و ناایستا، عملکرد بلوک‌های توجه را بهبود می‌بخشد. کاست^{۸۲} [۳۰] با بهره‌گیری از ثبات‌های زمانی در حوزه زمان، روند تمایز یک دنباله را یاد می‌گیرد و داده‌ها را به حوزه فرکانس برای کشف بازنمایی‌های فصلی منتقل می‌کند. اسکیل‌فورمر^{۸۳} [۲۴] چارچوبی چند مقیاسی که مدل‌های پیش‌بینی سری‌های زمانی مبتنی بر ترنسفورمر مانند فدفورمر و اتوفورمر را بهبود می‌بخشد. این مدل به‌طور مکرر پیش‌بینی‌ها را در مقیاس‌های مختلف با استفاده از وزن‌های مشترک اصلاح می‌کند. ای‌تی‌اس‌فورمر [۳۱] با استفاده از هموارسازی نمایی روندها را استخراج کرده و ضریب کاهش‌دهنده‌ای را برای ایجاد روند پایدار ترکیب می‌کند. این رویکرد دید با قابلیت تفسیر بهتری درباره نتایج پیش‌بینی ارائه داده و از یادگیری باقیمانده برای مدل‌سازی وابستگی‌های پیچیده بهره می‌برد. آی‌ترنسفورمر [۱۶] ترتیب پردازش اطلاعات را معکوس می‌کند. ابتدا اطلاعات سری زمانی هر متغیر را تحلیل کرده و سپس داده‌های تمامی متغیرها را ادغام می‌کند. این رویکرد با تجمیع تغییرات سری‌ها، بازنمایی سری‌های زمانی را بهبود می‌بخشد و مدیریت سری‌های نوسانی را تقویت می‌کند. ام‌تی‌اس‌تی^{۸۴} [۳۶] از یک رویکرد با تحلیل وضوح چندگانه^{۸۵} استفاده می‌کند که از ماهیت چندمقیاسی داده‌های سری زمانی واقعی بهره می‌گیرد. پث‌فورمر [۲] با ادغام مسیرهای تطبیقی^{۸۶} یادگیری بازنمایی چندمقیاسی را تقویت می‌کند. این مدل به‌صورت پویا مقیاس‌های مرتبط را بر اساس داده‌های ورودی انتخاب کرده و بر الگوهای زمانی آگاه‌کننده^{۸۷} تمرکز می‌کند. جی‌تی‌فورمر [۱۱] یک مکانیزم مبتنی بر گراف با آگاهی از ترتیب زمانی^{۸۸} را ادغام می‌کند تا هم وابستگی‌های زمانی و هم روابط زیرین بین متغیرها را ثبت کند. این مدل با ایجاد یک نمایش گراف، وابستگی‌های پیچیده بین متغیرها را درک کرده و دقت پیش‌بینی را افزایش می‌دهد. جی‌تی‌اف‌تی^{۸۹} [۴] با استفاده از یک ترنسفورمر مشترک در حوزه‌های زمانی و فرکانسی طراحی شده است. این مدل به‌طور همزمان وابستگی‌های زمانی و فرکانسی را درک کرده و پیش‌بینی دقیق‌تری از سری‌های زمانی چندمتغیره ارائه می‌دهد. ای‌دی‌اتوفورمر^{۹۰} [۱] که از اتوفورمر مشتق شده، با ترکیب روش‌های تجزیه و تقطیر توجه^{۹۱}، پیش‌بینی توالی‌های بلند را بهبود می‌بخشد. این رویکرد ترکیبی، دقت پیش‌بینی را افزایش داده و کارایی محاسباتی را حفظ می‌کند. فردفورمر^{۹۲} [۲۳] با ادغام تجزیه‌ی فرکانس، ارب‌های^{۹۳} موجود در ترنسفورمرها را رفع می‌کند. این مدل با حذف ارب‌های فرکانسی، دقت پیش‌بینی را به ویژه برای سری‌های زمانی پیچیده و دارای الگوهای متنوع افزایش می‌دهد.

کاهش هزینه‌ی محاسباتی. بسیاری از مدل‌های ترنسفورمر از رویکردهای مختلفی برای کاهش هزینه‌ی محاسباتی استفاده می‌کنند که اغلب با معرفی مکانیزم‌های توجه پراکنده است: لاگ‌ترنس^{۹۴} [۱۰] لایه‌های توجه پراکنده‌ی لگاریتمی^{۹۵} را برای دستیابی به پیچیدگی لگاریتمی نسبت به طول ورودی معرفی می‌کند. اینفورمر^{۹۶} [۳۸] پرس‌وجوهای غالب را براساس شباهت‌های کلید پرس‌وجو انتخاب کرده و به کارایی محاسباتی مشابه لاگ‌ترنس دست می‌یابد. همچنین دارای یک رمزگشای مولد برای پیش‌بینی مستقیم بلندمدت است که از انباشته شدن خطا از پیش‌بینی‌های یک مرحله‌ای جلوگیری می‌کند. ریفورمر^{۹۷} [۹] با استفاده از توجه مبتنی بر هش محلی حساس^{۹۸} و لایه‌های باقیمانده‌ی معکوس، پیچیدگی محاسباتی را کاهش می‌دهد. پیرافورمر^{۹۹} [۱۴] یک ماژول توجه هرمی سلسله‌مراتبی^{۱۰۰} با ساختار درخت دودویی معرفی می‌کند که وابستگی‌های زمانی را در بازه‌های مختلف

⁸⁰SSDNET ⁸¹state space models ⁸²CoST ⁸³Scaleformer ⁸⁴MTST ⁸⁵multi-resolution ⁸⁶adaptive pathways ⁸⁷informative temporal patterns ⁸⁸graph based temporal order aware mechanism ⁸⁹joint time frequency transformer (JTFT) ⁹⁰AD-Autoformer ⁹¹attention distilling ⁹²Fredformer ⁹³biases ⁹⁴LogTrans ⁹⁵LogSparse ⁹⁶Informer ⁹⁷Reformer ⁹⁸local sensitive hashing ⁹⁹Pyraformer ¹⁰⁰hierarchical pyramidal attention module

با پیچیدگی زمانی و حافظه‌ی خطی به تصویر می‌کشد. کواتفورمر^{۱۰۱} [۳] مکانیزم توجه چرخشی^{۱۰۲} براساس کواترنیون‌ها^{۱۰۳} را پیشنهاد می‌کند و اطلاعات دوره و فاز قابل یادگیری را برای الگوهای دوره‌ای پیچیده به‌کار می‌گیرد. اتفورمر [۳۲] و فدفورمر [۳۹] نیز به دنبال دستیابی به پیچیدگی محاسباتی کمتر نسبت به ترنسفورمر اولیه هستند.

معماری‌ها یا مکانیزم‌های جدید. راب‌فورمر^{۱۰۴} [۳۳] با استفاده از یک ماژول تجزیه‌ی تطبیقی^{۱۰۵} برای مدیریت الگوهای پیچیده‌ی سری زمانی و اختلال، پیش‌بینی را بهبود می‌بخشد. این ماژول نقاط قوت تجزیه‌ی مد تجربی ترکیبی^{۱۰۶} و تحلیل فوریه^{۱۰۷} را ترکیب می‌کند و منجر به افزایش دقت و پایداری پیش‌بینی‌ها می‌شود. تایمر^{۱۰۸} [۱۸] رویکرد جدیدی را با بهره‌گیری از ترنسفورمرهای پیش‌آموزش‌یافته مولد به‌عنوان مدل‌های بزرگ سری زمانی ارائه می‌دهد. این رویکرد به‌ویژه در وظایفی با دسترسی محدود به داده‌ها نتایج امیدوارکننده‌ای نشان می‌دهد. این مدل در وظایفی مانند تکمیل داده و تشخیص ناهنجاری، عملکرد بهتری نسبت به مدل‌های یادگیری عمیق کوچک‌تر دارد. تری‌فورمر^{۱۰۹} [۶] یک مکانیزم توجه وصله‌ای^{۱۱۰} متغیر خاص مثلثی^{۱۱۱} با ساختار درختی ارائه می‌کند که اندازه‌ی ورودی را به‌صورت نمایی کاهش می‌دهد. پیچ‌تی‌اس‌تی^{۱۱۲} [۲۱] از یک طراحی مستقل از کانال استفاده می‌کند که در آن هر کانال با یک سری زمانی تک متغیره، نگاشت یکسانی را در همه‌ی سری‌ها به اشتراک می‌گذارد. کراس‌فورمر^{۱۱۳} [۳۷] از وابستگی متقاطع^{۱۱۴} برای پیش‌بینی سری‌های زمانی چندمتغیره استفاده می‌کند. در [۲۶] با استفاده از ساختار سلسله‌مراتبی و تجزیه احتمالاتی، اجزای فصلی و روند را به‌طور مؤثری جدا کرده و دقت پیش‌بینی را با توجه به الگوهای قطعی و احتمالاتی بهبود می‌بخشد. همچنین [۲۲] رویکردی مبتنی بر مجموعه‌ای از ترنسفورمرهای زمانی برای پیش‌بینی مالی معرفی می‌کند که با ترکیب چند ترنسفورمر، مدل به‌طور جامع‌تری وابستگی‌های بلندمدت و پویای بازار را درک می‌کند. علاوه بر این مدل‌های ترنسفورمر، برخی تحقیقات نشان می‌دهند که مدل‌های خطی مانند دی‌لینیر^{۱۱۵} [۳۴] می‌توانند در شرایط خاصی عملکرد بهتری نسبت به معماری‌های پیچیده ترنسفورمر داشته باشند. این یافته اهمیت بررسی دقیق پیچیدگی مدل و ویژگی‌های داده را هنگام انتخاب رویکرد پیش‌بینی برجسته می‌کند. حوزه‌ی پیش‌بینی سری‌های زمانی با ترنسفورمرها پویا و در حال تکامل سریع است. تحقیقات بیشتر احتمالاً منجر به مدل‌های پیچیده‌تر و کارآمدتر خواهد شد که از نقاط قوت رویکردهای مختلف بهره می‌گیرند و ایده‌های جدیدی را برای مقابله با چالش‌های تحلیل سری‌های زمانی دقیق و کارآمد معرفی می‌کنند.

۷. نتیجه‌گیری و جهت‌های تحقیقاتی آینده

پیش‌بینی سری‌های زمانی چندمتغیره به‌عنوان یکی از چالش‌برانگیزترین حوزه‌های یادگیری ماشین، نیازمند رویکردهایی است که بتوانند وابستگی‌های پیچیده‌ی زمانی و بین‌متغیری را به‌طور همزمان مدل‌سازی کنند. معماری‌های مبتنی بر مکانیزم توجه نشان داده‌اند که می‌توانند به‌طور مؤثر الگوهای بلندمدت و کوتاه‌مدت را یاد بگیرند. با این حال، نتایج این مرور نشان داد که هنوز مسائل مهمی وجود دارد که عبارت‌اند از:

- **مدل‌سازی در شرایط داده‌های با فراوانی بالا:** پدیده‌هایی همچون همبستگی منفی القایی و تفاوت در ساختار نمونه‌گیری (افزایش دامنه در برابر پرشدگی در بازه‌ی ثابت) می‌توانند بر دقت و پایداری مدل‌ها اثرگذار باشند. بررسی ما نشان داد که برخی معماری‌ها مانند آی‌ترنسفورمر نسبت به اختلال‌های فرکانسی مقاوم‌ترند، در حالی که معماری‌هایی مانند فدفورمر یا ای‌تی‌اس‌فورمر به دلیل تکیه بر تبدیل فوریه در شرایط پرشدگی آسیب‌پذیرند.
- **چالش تعادل بین دقت و کارایی محاسباتی:** با افزایش نرخ نمونه‌برداری یا طول دنباله، پیچیدگی محاسباتی معماری‌های موجود رشد چشمگیری پیدا می‌کند و نیازمند طراحی‌های بهینه‌تر است.

¹⁰¹Quatformer ¹⁰²learning to rotate attention ¹⁰³quaternions ¹⁰⁴Robformer ¹⁰⁵adaptive decomposition module ¹⁰⁶ensemble empirical mode decomposition ¹⁰⁷fourier analysis ¹⁰⁸Timer ¹⁰⁹Triformer ¹¹⁰patch attention mechanism ¹¹¹triangular variable specific ¹¹²PatchTST ¹¹³Crossformer ¹¹⁴cross dsimension ¹¹⁵DLinear

- تفسیرپذیری نتایج: در بسیاری از حوزه‌های حساس مانند پزشکی یا مالی، توانایی توضیح پیش‌بینی‌ها به اندازهی دقت آن‌ها اهمیت دارد و این موضوع هنوز نیازمند نوآوری‌های جدی است.

جهت‌های تحقیقاتی آینده می‌تواند شامل موارد زیر باشد:

- ترکیب رویکردهای مبتنی بر حوزه‌ی زمان و حوزه‌ی فرکانس: برای مدل‌سازی همزمان الگوهای کوتاه‌مدت ناشی از پرشدگی و الگوهای بلندمدت در دامنه‌های وسیع‌تر.
 - مدل‌های مقاوم‌تر به همبستگی‌های منفی: توسعه‌ی معماری‌هایی که بتوانند وابستگی‌های غیرمنتظره را در داده‌های با فراوانی بالا مدیریت کنند.
 - سیستم‌های انطباق‌پذیر و چندمقیاسی: طراحی مدل‌هایی که بتوانند بسته به نوع نمونه‌گیری و نرخ داده، خود را بهینه‌سازی کنند.
 - پوشش بهتر داده‌های واقعی: ایجاد مجموعه‌داده‌هایی که بازتاب‌دهنده‌ی شرایط واقعی داده‌های با فراوانی بسیار بالا باشند تا امکان ارزیابی عملی‌تر مدل‌ها فراهم شود.
- با توجه به رشد سریع این حوزه، انتظار می‌رود در آینده شاهد ظهور الگوهای فکری و روش‌های جدیدی برای غلبه بر محدودیت‌های فعلی باشیم.

مراجع

- [1] D. Cao and S. Zhang, AD-autoformer: Decomposition transformers with attention distilling for long sequence time-series forecasting, *The Journal of Supercomputing*, **2024** 1–21.
- [2] P. Chen, Y. Zhang, Y. Cheng, Y. Shu, Y. Wang, Q. Wen, B. Yang and C. Guo, Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting, *arXiv preprint arXiv:2402.05956*, 2024.
- [3] W. Chen, W. Wang, B. Peng, Q. Wen, T. Zhou and L. Sun, Learning to rotate: Quaternion transformer for complicated periodical time series forecasting, *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, (2022) 146–156.
- [4] Y. Chen, S. Liu, J. Yang, H. Jing, W. Zhao and G. Yang, A Joint Time-Frequency Domain Transformer for multivariate time series forecasting, *Neural Networks*, **176** (2024) 106334.
- [5] Z. Chen, M. Ma, T. Li, H. Wang and C. Li, Long sequence time-series forecasting with deep learning: A survey, *Information Fusion*, **97** (2023) 101819.
- [6] R.-G. Cirstea, C. Guo, B. Yang, T. Kieu, X. Dong and S. Pan, Triformer: Triangular, Variable-Specific Attentions for Long Sequence Multivariate Time Series Forecasting–Full Version, *arXiv preprint arXiv:2204.13767*, 2022.
- [7] W. Guan, I. Smetannikov and T. Man, Survey on automatic text summarization and transformer models applicability, *Proceedings of the 2020 1st International Conference on Control, Robotics and Intelligent System*, (2020) 176–184.
- [8] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, et al., A survey on vision transformer, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45** no. 1 (2022) 87–110.

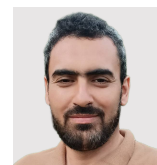
- [9] N. Kitaev, Ł. Kaiser and A. Levskaya, Reformer: The efficient transformer, *arXiv preprint arXiv:2001.04451*, 2020.
- [10] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang and X. Yan, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting, *Advances in Neural Information Processing Systems*, **32** (2019) 5243–5253.
- [11] A. Liang, X. Chai, Y. Sun and M. Guizani, GTformer: Graph-Based Temporal-Order-Aware Transformer for Long-Term Series Forecasting, *IEEE Internet of Things Journal*, (2024) 31467–31478.
- [12] B. Lim, S. Ö. Arık, N. Loeff and T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, *International Journal of Forecasting*, **37** no. 4 (2021) 1748–1764.
- [13] Y. Lin, I. Koprinska and M. Rana, SSDNet: State space decomposition neural network for time series forecasting, *2021 IEEE International Conference on Data Mining (ICDM)*, (2021) 370–378.
- [14] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu and S. Dustdar, Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting, *The Tenth International Conference on Learning Representations (ICLR)*, (2022).
- [15] X. Liu and W. Wang, Deep Time Series Forecasting Models: A Comprehensive Survey, *Mathematics*, **12** no. 10 (2024) 1504.
- [16] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma and M. Long, iTransformer: Inverted Transformers Are Effective for Time Series Forecasting, *The Twelfth International Conference on Learning Representations*, (2024).
- [17] Y. Liu, H. Wu, J. Wang and M. Long, Non-stationary transformers: Exploring the stationarity in time series forecasting, *Advances in Neural Information Processing Systems*, **35** (2022) 9881–9893.
- [18] Y. Liu, H. Zhang, C. Li, X. Huang, J. Wang and M. Long, Timer: Generative Pre-trained Transformers Are Large Time Series Models, *Forty-first International Conference on Machine Learning*, (2024).
- [19] Q. Ma, Z. Liu, Z. Zheng, Z. Huang, S. Zhu, Z. Yu and J. T. Kwok, A survey on time-series pre-trained models, *IEEE Transactions on Knowledge and Data Engineering*, (2024) 7536–7555.
- [20] J. A. Miller, M. Aldosari, F. Saeed, N. H. Barna, S. Rana, I. B. Arpinar and N. Liu, A survey of deep learning and foundation models for time series forecasting. *arXiv 2024, arXiv preprint arXiv:2401.13912*, (2024).
- [21] Y. Nie, N. H. Nguyen, P. Sinthong and J. Kalagnanam, A time series is worth 64 words: Long-term forecasting with transformers, *arXiv preprint arXiv:2211.14730*, (2022).
- [22] K. Olorunnimbe and H. Viktor, Ensemble of temporal Transformers for financial time series, *Journal of Intelligent Information Systems*, (2024) 1–25.
- [23] X. Piao, Z. Chen, T. Murayama, Y. Matsubara and Y. Sakurai, Fredformer: Frequency debiased transformer for time series forecasting, *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, (2024) 2400–2410.
- [24] A. Shabani, A. Abdi, L. Meng and T. Sylvain, Scaleformer: Iterative multi-scale refining transformers for time series forecasting, *arXiv preprint arXiv:2206.04038*, (2022).
- [25] D. V. da Silva, V. Estevam and D. Menotti, Towards a realistic Libras to Portuguese translation, *2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, (2023) 1–6.

- [26] J. Tong, L. Xie, W. Yang, K. Zhang and J. Zhao, Enhancing time series forecasting: A hierarchical transformer with probabilistic decomposition representation, *Information Sciences*, **647** (2023) 119410.
- [27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. U. Kaiser and I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, **30** (2017) 1–15.
- [28] Y. Wang, H. Wu, J. Dong, Y. Liu, M. Long and J. Wang, Deep time series models: A comprehensive survey and benchmark, *arXiv preprint arXiv:2407.13278*, (2024).
- [29] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan and L. Sun, Transformers in time series: A survey, *arXiv preprint arXiv:2202.07125*, (2022).
- [30] G. Woo, C. Liu, D. Sahoo, A. Kumar and S. Hoi, CoST: Contrastive Learning of Disentangled Seasonal-Trend Representations for Time Series Forecasting, *International Conference on Learning Representations*, (2022).
- [31] G. Woo, C. Liu, D. Sahoo, A. Kumar and S. Hoi, Etsformer: Exponential smoothing transformers for time-series forecasting, *arXiv preprint arXiv:2202.01381*, (2022).
- [32] H. Wu, J. Xu, J. Wang and M. Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, *Advances in Neural Information Processing Systems*, **34** (2021) 22419–22430.
- [33] Y. Yu, R. Ma and Z. Ma, Robformer: A robust decomposition transformer for long-term time series forecasting, *Pattern Recognition*, **153** (2024) 110552.
- [34] A. Zeng, M. Chen, L. Zhang and Q. Xu, Are transformers effective for time series forecasting?, *Proceedings of the AAAI Conference on Artificial Intelligence*, **37** no. 9 (2023) 11121–11128.
- [35] J. Zhang, H. Luan, M. Sun, F. Zhai, J. Xu, M. Zhang and Y. Liu, Improving the Transformer Translation Model with Document-Level Context, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, (2018) 533–542.
- [36] Y. Zhang, L. Ma, S. Pal, Y. Zhang and M. Coates, Multi-resolution Time-Series Transformer for Long-term Forecasting, *International Conference on Artificial Intelligence and Statistics*, (2024) 4222–4230.
- [37] Y. Zhang and J. Yan, Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting, *The Eleventh International Conference on Learning Representations*, (2023).
- [38] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong and W. Zhang, Informer: Beyond efficient transformer for long sequence time-series forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** no. 12 (2021) 11106–11115.
- [39] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun and R. Jin, Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting, *International Conference on Machine Learning*, (2022) 27268–27286.
- [40] S. Bandyopadhyay and S. N. Lahiri, Asymptotic properties of discrete Fourier transforms for spatial data, *Sankhyā: The Indian Journal of Statistics, Series A*, **71** no. 2 (2009) 221–259.
- [41] R. S. Tsay, *Analysis of Financial Time Series*, 2nd ed., Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2005, ISBN 978-0-471-69074-0, doi: 10.1002/0471746193.

اسمعیل چیتگر

گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه یزد، یزد، ایران
ec.chitgar@gmail.com

اسمعیل چیتگر دانشجوی مقطع دکتری رشته علوم کامپیوتر دانشگاه یزد می‌باشد. ایشان کارشناسی ارشد علوم کامپیوتر را در سال ۱۳۹۲ از دانشگاه صنعتی شریف و کارشناسی خود را در همین رشته و در سال ۱۳۸۶ از دانشگاه مازندران اخذ کرده است.



مهدیه هاشمی‌نژاد

گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه یزد، یزد، ایران
hasheminezhad@yazd.ac.ir

مهدیه هاشمی‌نژاد استادیار گروه علوم کامپیوتر دانشکده علوم ریاضی دانشگاه یزد است. تحقیقات ایشان در حوزه‌ی ترکیبیات و الگوریتم‌های گراف و نیز داده‌کاوی شبکه‌های اجتماعی است. وی دانش‌آموخته‌ی دکتری پیوسته‌ی رشته ریاضی دانشگاه صنعتی امیرکبیر در سال ۱۳۸۸ می‌باشد.



سپهر ابراهیمی‌مود

گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه یزد، یزد، ایران
s.ebrahimi@yazd.ac.ir

سپهر ابراهیمی‌مود، استادیار و مدیر گروه علوم کامپیوتر دانشکده علوم ریاضی دانشگاه یزد است. حوزه‌های تحقیقاتی و تدریس ایشان شامل مبانی نظریه محاسبات، نظریه محاسبات و محاسبات نرم می‌باشد. وی دانش‌آموخته مقطع دکتری علوم کامپیوتر از دانشگاه شهیدباهنر کرمان در سال ۱۳۹۸ و کارشناسی ارشد همین رشته از دانشگاه صنعتی شریف در سال ۱۳۹۳ است.

