Deep Learning for Time Series Forecasting: A Survey

Authors: Huang Jiajun

Date: 2024-01-08T00:00:00+00:00

Abstract

The rapid advancement of sensor and network technologies has led to the proliferation of extensive historical time series data, rendering efficient and accurate time series forecasting increasingly critical. In recent years, approaches incorporating deep learning concepts and techniques for time series forecasting tasks have advanced swiftly, yielding substantial achievements. This paper analyzes the current research status of time series forecasting methods both domestically and internationally, expounds upon the relevant theories underlying time series forecasting, and systematically summarizes the traditional methods, machine learning-based methods, and deep learning-based methods utilized for this task. It places particular emphasis on a comparative analysis of the strengths and weaknesses of various deep learning-based approaches, thereby offering perspectives on the future development of deep learning-based time series forecasting methods.

Full Text

Abstract

A Survey of Deep Learning Applications in Time Series Forecasting

Jiajun Huang
Beijing University of Chemical Technology

The rapid advancement of sensor and network technologies has generated vast amounts of historical time series data, making efficient and accurate time series forecasting increasingly critical. In recent years, the application of deep learning concepts and techniques to time series forecasting has developed rapidly and achieved numerous significant results.

This paper analyzes the current state of time series forecasting research both domestically and internationally, discusses the relevant theories involved in time series forecasting, and systematically summarizes the traditional methods, machine learning-based methods, and deep learning-based methods employed for this task. The paper focuses on comparing and analyzing the advantages and disadvantages of various deep learning-based approaches, and consequently provides an outlook on deep learning-based time series forecasting methods.

Keywords: time series forecasting, deep learning, machine learning

1.1 Background and Significance

Time series data finds extensive applications across numerous domains including finance, meteorology, agriculture, industry, and healthcare [**?**]. In recent years, the rapid development of sensor and network technologies has made it possible to generate and accumulate massive amounts of time series data. Within the field of time series data analysis, research typically addresses multiple problems such as classification, anomaly detection, and forecasting, with time series prediction being a primary focus. Unlike conventional regression-based predictive models, time series models rely more heavily on the temporal ordering of data. Time series forecasting encompasses both continuous predictions (numerical forecasting or range estimation) and discrete predictions (event forecasting), with its core objective being to extract patterns from historical temporal data and utilize these patterns to estimate future development trends. Due to its ability to analyze underlying patterns and predict future trends, time series forecasting research holds substantial academic significance and practical value for guiding decision-making processes across various industries.

As the volume of temporal data and the dimensionality of features continue to grow, research methodologies for time series forecasting have undergone continuous improvement, evolving from initial mathematical and statistical approaches to machine learning methods, and subsequently to deep learning techniques. Since natural language processing inherently requires sequential modeling, the remarkable successes of deep learning in NLP tasks have inspired the adaptation of these widely-used techniques to time series research. Consequently, deep learning-based time series forecasting methods have experienced rapid development, necessitating a comprehensive analysis and discussion of their historical evolution and current state. This paper first examines the time series forecasting problem, then systematically reviews its research methodologies, provides detailed analysis and discussion of deep learning-based approaches categorized by network architecture type, and concludes with a forward-looking perspective on the future development of time series forecasting technology.

1.2 Time Series

Time series generally refers to a set of random variables collected by observing the development and change process of a phenomenon at a certain frequency.

Due to inherent potential relationships among the variables in time series data, they often exhibit one or more characteristic properties.

To provide a comprehensive understanding of time series forecasting, this section details several common characteristics. First, **massiveness**: with the upgrading of IoT sensor devices, increased measurement frequency, and expanded measurement dimensions, time series data has grown explosively, with high-dimensional data becoming mainstream. Effective preprocessing at the dataset level is crucial for high-quality time series forecasting. Second, **trend**: current data points are often closely related to previous ones, indicating that time series data typically follows certain patterns influenced by various factors, exhibiting long-term stable increase, decrease, or horizontal trends. Third, **periodicity**: influenced by external factors, data in time series may display alternating fluctuations over extended periods—for instance, tidal heights throughout a week do not follow a simple trend but rather show regular rises and falls. Fourth, **volatility**: over time and under multiple external influences, the variance and mean of time series may undergo systematic changes, affecting prediction accuracy. Fifth, **stationarity**: while time series data may appear randomly variable at individual time points, they exhibit statistical regularity across different time periods, maintaining relatively stable variance and mean. Sixth, **symmetry**: if the distance between an original time series and its reversed version within a certain period falls below a specific threshold, with curves aligning substantially, the series is considered symmetric—for example, in the repetitive operations of port transport vehicles or crane arm movements.

1.3 Time Series Forecasting

Time series forecasting aims to extract core patterns from vast amounts of data and make accurate estimates of future values based on known factors.

For single-step time series forecasting, historical observations and covariates serve as input variables to predict the value at the next time step—a process known as single-step forecasting. However, many forecasting applications require predicting trends over future time intervals, such as electricity consumption in a region, degradation trends of components, or stock price movements. This scenario necessitates forecasting multiple future time steps, typically using historical observations and covariates as inputs to predict values for the next n time steps—a task known as multi-step forecasting.

Multi-step time series forecasting employs five primary strategies: Direct Multi-step Forecast, Recursive Multi-step Forecast, Direct-Recursive Hybrid Forecast, Multiple Output Forecast, and seq2seq Forecast.

2.1 Traditional Time Series Modeling Methods

Traditional time series forecasting methods primarily involve determining parametric models and solving their parameters for future predictions. Representative approaches include ARIMA and Holt-Winters. ARIMA (Auto-Regressive Integrated Moving Averages) is a widely used statistical method [?] that requires stationarity testing, white noise examination, and calculation of ACF (autocorrelation function) and PACF (partial autocorrelation function) for model identification and forecasting. Numerous ARIMA variants exist, such as SARIMA and SARIMAX [?]. The Holt-Winters method extends traditional Holt's method to capture seasonality through prediction equations and three smoothing parameters (α, β, and γ) for level, trend, and seasonal components [?]. While effective for simple forecasting problems, traditional models struggle with high dimensionality and complex patterns, where more sophisticated models become necessary.

2.2 Time Series Decomposition Methods

Time series decomposition has long been a valuable analytical approach, positing that a time series represents the superposition or coupling of multiple components: long-term secular trend (T), seasonal variation (S), cyclical variation (C), and irregular variation (I). Depending on the application, decomposition follows either additive or multiplicative principles. In additive models, components are independent with identical units, while multiplicative models maintain interdependent relationships, typically expressing trend in absolute terms and other components relatively. A prominent decomposition method is Facebook's open-source Prophet model [?], which decomposes time series into trend, seasonal (weekly/monthly), holiday, and noise components. Beyond its direct application, decomposition provides a valuable analytical framework that has influenced the design of many subsequent machine learning and deep learning approaches.

2.3 Machine Learning Methods

Time series forecasting is fundamentally connected to regression analysis in machine learning. Major algorithm categories include Support Vector Regression (SVR), which employs kernel methods to map inputs to high-dimensional space without increased computational complexity, effectively overcoming the curse of dimensionality and demonstrating stable predictive capability for nonlinear time series [?]. Gradient Boosting Regression Trees (GBRT) introduces gradient descent for regression, iteratively computing the negative gradient of the loss function to find optimal models with high robustness to outliers [?]. Hidden Markov Models (HMM) provide a probabilistic framework for multivariate time series forecasting, representing a dual stochastic process with an unobservable Markov chain and observable random functions [?]. As the simplest dynamic Bayesian network, HMM calculates observation sequence probabilities and identifies optimal state sequences.

2.4 Deep Learning Methods

Following deep learning's remarkable achievements in computer vision and natural language processing, these methods have been increasingly applied to time series forecasting. Deep neural networks enable better high-dimensional data representation through various architectures, reducing manual feature engineering requirements and facilitating end-to-end training through defined loss functions. By constructing multiple nonlinear layers, deep neural networks build feature representations of historical time series to learn internal variation patterns, effectively encoding relevant historical information into hidden variables for prediction.

2.4.1 Convolutional Neural Network-Based Methods

Convolutional Neural Networks (CNNs), widely used in computer vision for extracting spatially-invariant local features, can similarly extract local temporal features. Through hierarchical aggregation across layers, CNNs capture hidden information from longer historical sequences. However, traditional CNNs have limited connections between distant inputs and outputs, restricting their forecasting capability. To expand the receptive field, DeepMind proposed WaveNet [?], based on dilated causal convolution for speech generation, which experts later adapted for time series forecasting. Dilated convolution aggregates information from increasingly distant time points by increasing dilation rates layer-by-layer, effectively utilizing long-range historical information [?]. Temporal Convolutional Networks (TCN) treat sequences as one-dimensional objects, capturing long-term relationships through iterative convolutions using causal, dilated, and residual skip connections [?], enabling parallel computation for improved efficiency. Building on this foundation, SCINet employs a hierarchical convolutional structure that iteratively extracts and aggregates features at different temporal resolutions, learning effective representations with enhanced predictability [?]. Its binary tree architecture splits time series into two parts at each SCI-Block, extracting finer temporal information with increasing depth to capture both short- and long-term dependencies.

2.4.2 Recurrent Neural Network-Based Methods

Recurrent Neural Networks (RNNs) have been extensively used for sequence modeling, achieving excellent results in natural language processing tasks. RNNs learn hidden states from all previous time steps as feature representations of past information, combining them with current inputs for next-step predictions. Hidden states are recursively updated with new observations at each time step, making RNN architectures among the earliest deep learning approaches applied to time series forecasting. Early RNN variants faced gradient explosion and vanishing issues when learning long-term dependencies. Long Short-Term Memory (LSTM) networks [?] addressed this through gating mechanisms inspired by logical control gates—output gates control sequence output, input gates determine data reading, and forget gates manage content resetting. Through learned weights, LSTM decides when to remember or ignore inputs, becoming essential for RNN-based forecasting.

DeepAR employs LSTM for time series forecasting [?], using previous actual values and covariates as inputs during training, then substituting predicted values for actual values during forecasting. The Deep State Space model [?] provides a state-space transformation approach that estimates relationships between consecutive hidden states, enabling predictions from current hidden states without requiring previous actual or predicted values, thus resolving training-inference inconsistency. To address data duplication and memory issues during training, interleaved training methods [?] achieve sliding-window-like effects without data replication, balancing short- and long-term forecasting. Additionally, hybrid models combining exponential smoothing with RNNs [?] leverage complementary strengths to improve forecasting performance.

2.4.3 Self-Attention Network-Based Methods

Self-attention networks originated in natural language processing, with Transformer introduced in the 2017 paper "Attention is All You Need" [?], achieving state-of-the-art results for machine translation through parallel computation in the encoder, significantly reducing training and inference time. Current mainstream language models like GPT and BERT are Transformer-based. Due to similarities with NLP tasks, Transformer methods were quickly adapted for time series forecasting. While RNNs must process sequences sequentially from $t-n$ to t, making distant time point connections weak and inefficient, attention mechanisms enable pairwise associations between all input units, with weighted representations passed to subsequent layers for better temporal context interaction.

Literature [?] applied GPT-like Transformer structures to time series forecasting with promising results, though limitations in computation, memory usage, and encoder-decoder architecture hindered direct application to longer sequences. Literature [?] proposed convolutional self-attention, reducing complexity by generating Queries and Keys using causal convolutions. Informer [?] improved computational complexity to $O(\log L)$ by selecting dominant queries and employed a generative decoder for direct long-term forecasting, avoiding cumulative errors in single-step forward prediction. Researchers have also explored frequency-domain attention mechanisms. Autoformer [?] designed a short-term trend decomposition architecture using autocorrelation as an attention module, measuring delay similarity between input signals and aggregating top-k similar subsequences with $O(L \log L)$ complexity. FEDformer [?] applies attention operations in the frequency domain through Fourier and wavelet transforms, achieving linear complexity by randomly selecting fixed-size frequency subsets.

2.4.4 Graph Neural Network-Based Methods

In multivariate time series forecasting, variables exhibit interdependencies that existing methods cannot fully exploit. Graph Neural Networks (GNNs) have demonstrated strong capability in handling relational dependencies. Typically, a graph structure learning layer defines information propagation structures, followed by convolutional operations that capture spatial and temporal dependencies for multivariate forecasting. Literature [?] proposed a general GNN framework for multivariate time series that automatically extracts variable relationships through a graph learning module, interleaving graph convolutional and temporal convolutional modules to capture spatial-temporal correlations using hybrid skip propagation layers. Literature [?] employed fully-connected graphs with N nodes, mapping them to K-node connection graphs based on internal correlations, reducing complexity from $O(N^2)$ to $O(NK)$ while balancing accuracy and computational cost. Graph WaveNet [?] introduced adaptive graph modeling through node embedding to capture hidden spatial dependencies, replacing RNNs with dilated causal 1D convolutions for easier long-term temporal relationship extraction.

2.4.5 Fully Connected Network-Based Methods with Residual Learning

The N-BEATS model [?] achieves excellent forecasting performance using only fully connected layers without RNN, CNN, or Attention mechanisms. Its core concept involves serially connected fully connected blocks (FC Blocks), where each block learns a portion of the input time series information. The input to subsequent blocks excludes information already learned by previous blocks, forcing later blocks to focus on residual information—similar to GBDT's iterative fitting approach. Final predictions aggregate outputs from all blocks, enabling each layer to concentrate on residuals that previous layers could not fit properly, thereby decomposing the time series from trend to details progressively. Specifically, N-BEATS comprises multiple Stacks, each containing several FC Blocks as fundamental modules built from fully connected layers.

To incorporate external features such as date information, holidays, and attributes, N-BEATSx [?] extends the original model by introducing exogenous variables x. GAGA [?] further extends N-BEATS to spatio-temporal forecasting, handling multiple time series with spatial relationships by adding Time Gate and Graph Gate structures to the architecture. Compared to other forecasting models, N-BEATS pioneered an all-fully-connected architecture that decomposes temporal information from coarse to fine granularity through continuous refinement, achieving accurate predictions.

Chapter 3 Summary and Outlook

This paper comprehensively introduced the background of time series forecasting, reviewed early methods including traditional and machine learning approaches, analyzed their connections and limitations, and focused on comparing various deep learning-based model designs. Although deep learning-based time series forecasting has achieved significant progress, future applications demand higher performance. Several key research directions emerge:

First, **integrating domain knowledge**: deep learning models must incorporate domain-specific knowledge to achieve higher accuracy and efficiency, considering particularities and leveraging specialized expertise during model construction. Second, **causal inference**: current forecasting primarily relies on correlation analysis within sequences, but future applications require greater emphasis on causal inference and improved model interpretability to ensure users understand decision processes. Third, **anomaly handling**: real-world data often contains anomalies; improving robustness requires mechanisms to ignore these points and reduce prediction errors. Fourth, **online learning**: current methods primarily use offline batch training, necessitating incremental learning approaches for real-time online analysis and adaptation to new data. Fifth, **hyperparameter optimization using stochastic nature-inspired algorithms**: as deep learning complexity increases, hyperparameter selection critically impacts performance. Nature-inspired optimization algorithms, drawing from swarm intelligence and natural phenomena, randomly generate solutions in the feasible space and iteratively seek global optima, representing a promising research direction. Sixth, **architectures for irregularly-sampled small datasets**: while Transformers excel on large periodic datasets, they underperform on small, irregularly-sampled data. Future research could introduce resampling, interpolation, filtering, or other methods into model architectures to address overfitting. Seventh, **GNN for multivariate forecasting**: given the complex and dynamic correlations in multivariate time series, GNNs offer superior modeling capabilities, with recent temporal polynomial graph neural networks achieving state-of-the-art results in short- and long-term forecasting. Eighth, **differentiable loss functions for shape and temporal dynamics**: existing loss functions often lack invariance to shape distortions and temporal delays; future research should explore differentiable loss functions that simultaneously support precise shape and temporal dynamics for more comprehensive model evaluation.

Through in-depth investigation of these directions, deep learning-based time series forecasting will become more reliable, efficient, and flexible in solving practical problems. As the information era advances, all domains will generate massive temporal data, and discovering patterns within this data to forecast future trends holds enormous economic and scientific value. This paper reviewed the evolution from traditional statistics through machine learning to deep learning, identified underlying development patterns, and provided future research perspectives.

References

[?] Cibicong, Zhang Pinyi. Financial Time Series Prediction Based on ARIMA-LSTM Model[J]. Statistics and Decision, 2022, 38(11): 145-149. DOI: 10.13546/j.cnki.tjyjc.2022.11.029.

[?] Che Changchang, Wang Huawei, Ni Xiaomei, et al. Remaining Useful Life Prediction of Aero-engines Based on 1D-CNN and Bi-LSTM[J]. Journal of Mechanical Engineering, 2021, 57(14): 304-312. DOI: 10.3901/JME.2021.14.304.

[?] Luo Xiaomeng, Li Jianbin, Hu Peng. E-commerce Inventory Optimization Strategy Based on Time Series Forecasting[J]. Systems Engineering, 2014, 32(6): 91-98.

[?] Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50: 159-175. DOI: 10.1016/S0925-2312(01)00702-0.

[?] Li Shaoxiong, Li Benguang. Comparison of Forecasting Based on SARIMA Model and X-12-ARIMA Seasonal Adjustment Method[J]. Statistics and Decision, 2018, 34(18): 39-42. DOI: 10.13546/j.cnki.tjyjc.2018.18.008.

[?] Chatfield C. The holt-winters forecasting procedure[J]. Journal of the Royal Statistical Society, 1978, 27(3): 264-279. DOI: 10.2307/2347162.

[?] Taylor S J, Letham B. Forecasting at Scale [J]. PeerJ Preprints, 2017, 5: e3190v2. DOI: 10.7287/peerj.preprints.3190v2.

[?] Chen Rong, Liang Changyong, Xie Fuwei. A Review of Nonlinear Time Series Prediction Methods Based on SVR[J]. Journal of Hefei University of Technology (Natural Science Edition), 2013, 36(3): 369-374. DOI: 10.3969/j.issn.1003-5060.

[?] Elsayed S, Thyssens D, Rashed A, et al. Do we really need deep learning models for time series forecasting?[Z]. arXiv: 2101.02118, 2021. DOI: 10.48550/arXiv.2101.02118.

[?] Zahari A, Jaafar J. A novel approach of hidden Markov model for time series forecasting[C]//Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication. Bali, Indonesia: ACM, 2015: 91. DOI: 10.1145/2701126.27011

[?] Van den Oord A, Dieleman S, Zen H G, et al. WaveNet: A generative model for raw audio[C]//The 9th ISCA Speech Synthesis Workshop. Sunnyvale, CA, USA: ISCA, 2016: 1

[?] Borovykh A, Bohte S, Oosterlee C W. Conditional time series forecasting with convolutional neural networks[Z]. arXiv: 1703.04691, 2017. DOI: 10.48550/arXiv.1703.04691.

[?] Borovykh A, Bohte S, Oosterlee C W. Dilated convolutional neural networks for time series forecasting[J]. Journal of Computational Finance, 2019, 22(4): 73-101. DOI: 10.21314/JCF.2019.358.

[?] Chen Y T, Kang Y F, Chen Y X, et al. Probabilistic forecasting with temporal convolutional neural network[J]. Neurocomputing, 2020, 399: 491-501. DOI: 10.1016/j.neucom.

[?] Liu M H, Zeng A L, Chen M X, et al. SCINet: Time series modeling and Forecasting with sample convolution and interaction[Z]. arXiv: 2106.09305, 2021. DOI: 10.48550/arXiv.2106.09305.

[?] Hossain M S, Mahmood H. Short-term load forecasting using an LSTM neural network[C]//2020 IEEE Power and Energy Conference at Illinois (PECI). Champaign: IEEE, 2020. DOI: 10.1109/PECI48348.2020.9064654.

[?] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks[J]. International Journal of Forecasting, 2020, 36(3): 1181-1191. DOI: 10.1016/j.ijforecast.2019.07.001.

[?] Rangapuram S S, Seeger M W, Gasthaus J, et al. Deep state space models for time series forecasting[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018: 7796-7805.

[?] Wen R F, Torkkola K, Narayanaswamy B, et al. A multi-horizon quantile recurrent forecaster[Z]. arXiv: 1711.11053, 2017. DOI: 10.48550/arXiv.1711.11053.

[?] Smyl S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting[J]. International Journal of Forecasting, 2020, 36(1): 75-85. DOI: 10.1016/j.ijforecast.2019.03.017.

[?] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017: 6000-6010.

[?] Wu N, Green B, Ben X, et al. Deep transformer models for time series forecasting: the influenza prevalence case[Z]. arXiv: 2001.08317, 2020. DOI: 10.48550/arXiv.2001.08317.

[?] Li S Y, Jin X Y, Xuan Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, BC, Canada: Curran Associates Inc., 2019: 471.

[?] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting[C]//Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Event: AAAI Press, 2021: 11106-11115.

[?] Wu H X, Xu J H, Wang J M, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event, Curran Associates, Inc., 2021, 34: 22419-22430.

[?] Zhou T, Ma Z Q, Wen Q S, et al. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]//International Conference on Machine Learning. Baltimore: PMLR, 2022: 27268-27286.

[?] Wu Z H, Pan S R, Long G D, et al. Connecting the Dots: multivariate time series forecasting with graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event, CA, USA: ACM, 2020: 753-763.

[?] Satorras V G, Rangapuram S S, Januschowski T. Multivariate time series forecasting with latent graph inference[Z]. arXiv: 2203.03423, 2022. DOI: 10.48550/arXiv.2203.03423.

[?] Wu Z H, Pan S R, Long G D, et al. Graph WaveNet for deep spatial-temporal graph modeling[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China: IJCAI.org, 2019: 1907-1913.

[?] Oreshkin B N, Carpov D, Chapados N, et al. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting[C]//8th International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net, 2020.

[?] Olivares K G, Challu C, Marcjasz G, et al. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx[Z]. arXiv: 2104.05522, 2021. DOI: 10.48550/arXiv.2104.05522.

[?] Oreshkin B N, Amini A, Coyle L, et al. FCGAGA: Fully connected gated graph architecture for spatio-temporal traffic forecasting[C]//Thirty-Fifth AAAI Conference on Artificial Intelligence. Virtual Event, AAAI Press, 2021: 9233-9241.

Note: Figure translations are in progress. See original paper for figures.