A SURVEY ON FORECASTING IOT TIME SERIES DATA

Ajit Patil[1], Dr. Amol Potgantwar[2]

[1]Deparatment of Computer Engineering, Savitribai Phule Pune Unniversity, Nashik,Email:ajit.patil1091@gmail.com
[2]Deparatment of Computer Engineering,Savitribai Phule Pune Unniversity Nashik, Email: amol.potgantwar@sitrc.org

ABSTRACT
The rapid growth of technologies such as sensors and the Internet of Things (IoT) has led to the continuous generation of vast amounts of data in the form of time series. Over the years, analyzing and forecasting time series data has become a central area of research due to its wide range of applications. Accurate forecasting plays a vital role in domains such as business, financial markets, weather prediction, electricity demand estimation, and resource management, including the consumption of fuels and power. It is also critical in any domain influenced by seasonal variations or long-term trends. Reliable forecasts provide organizations with essential insights for informed decision-making and strategic planning.
This paper presents a comprehensive survey of the techniques employed for time series forecasting across different types of datasets. It reviews general forecasting models, the algorithms underpinning these models, and optimization strategies developed to enhance predictive performance and accuracy. In addition, the paper highlights the evaluation metrics commonly applied to assess the effectiveness of forecasting approaches. By synthesizing prior research, this study offers a consolidated understanding of the advancements in time series forecasting and serves as a reference point for researchers and practitioners working in this area.

KEYWORDS: Forecasting, Time series, Prediction, Temporal, Data mining

INTRODUCTION
The Internet of Things (IoT) represents a computing paradigm in which interconnected devices, or “things,” exchange data autonomously without requiring constant human intervention [1]. These devices may include sensors, actuators, embedded systems, and smart objects capable of perceiving and responding to their surrounding environment. The availability of low-cost sensing technologies, coupled with efficient computing platforms that can operate in challenging environments, has significantly accelerated the adoption of IoT. Current trends indicate that IoT is expanding at an unprecedented pace, with projections estimating its global economic contribution to reach nearly $11.1 trillion annually by 2025 [2], supported by billions of connected devices worldwide [3].
In parallel with this growth, data mining has emerged as a vital process for extracting meaningful insights from massive and diverse datasets. These datasets may consist of structured or unstructured information in forms such as text, images, audio, and video. In business and market-driven sectors, this accumulation of data is often referred to as Big Data [4]. Through data mining, hidden patterns and relationships can be uncovered, enabling predictive modeling and supporting informed decision-making that fosters organizational growth. While aspects such as data security and integrity are also critical, the core focus of this study lies in prediction techniques and methodologies.
Among the different types of data, temporal data holds a distinct position as it evolves with time and is typically represented through time stamps. Temporal data mining deals with identifying useful patterns within temporal databases and analyzing dynamic data to capture recurring behaviors, dependencies, or trends. The essential functions in temporal data mining revolve around analyzing sequential changes and uncovering predictive patterns that can support applications across business, healthcare, finance, and other time-sensitive domains.
The primary functions of temporal data mining can be summarized as follows:
•Temporal classification: It is the process of finding a model that distinguishes data classes whose class label is unknown in the temporal database.
•Temporal prediction: It is the process of finding unknown future values from historical data.
•Temporal association: It searches for the temporal relationship between attributes.
•Temporal clustering: It is the process of grouping similar data objects from temporal database.
•Temporal regression: It attempts to find a function which models the data with the least error.
•Temporal summarization: It provides a more compact representation of the data set including visualization.
•Temporal outlier detection: Data objects that do not adhere with the generic behavior of other data within the database.
Time series data plays a crucial role in a wide range of real-world applications, such as optimizing information system operations, supporting artificial intelligence for IT management, guiding stock market investments, predicting disease outbreaks, and detecting abnormal equipment behavior. The ability to predict the future trajectory of time-dependent data enables organizations and decision-makers to act proactively rather than reactively. For instance, accurate forecasting of the progression of a particular disease can allow health authorities to adapt prevention strategies in real time, thereby improving control measures and minimizing potential risks.

In practice, time series data is generated across diverse domains, often in the form of continuous data streams. These streams usually differ in their statistical properties, distribution patterns, and levels of variability. As a result, a single predictive approach is rarely sufficient. Instead, multiple prediction models are often required to address these differences effectively. However, relying on numerous specialized models can increase complexity and resource usage. Therefore, the development of flexible and generalized forecasting frameworks has become an important research direction, aiming to balance prediction accuracy with computational efficiency.
Within the broader scope of temporal data mining, time series analysis and prediction occupy a central position [5]. By discovering underlying trends, periodicities, and correlations, time series forecasting not only deepens our understanding of temporal phenomena but also provides actionable insights for strategic decision-making in dynamic environments.

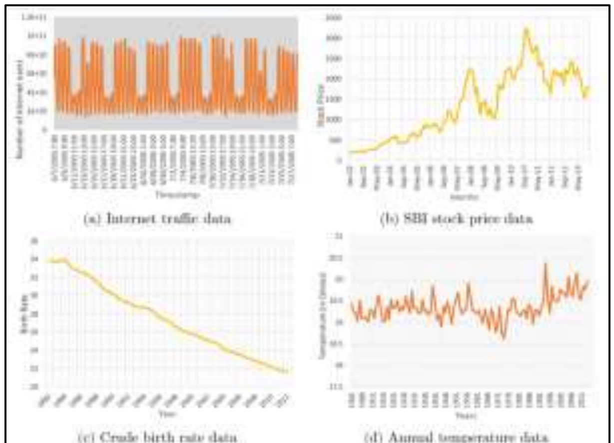

Fig. 1. Different type of time series data from various domains

Time series data refers to a sequence of values collected at uniform time intervals. Depending on the application, the temporal unit may be expressed in years, months, weeks, days, or even seconds. Figure 1 illustrates examples of time series data drawn from different domains. The primary objective of time series analysis is not only to understand the variations present in the observed data but also to forecast future behavior by examining its underlying characteristics and patterns.
A time series can often be described formally using a five-tuple representation: starting time, which indicates when the sequence begins; pattern, which reflects recurring structures or trends in the data; period value, denoting the cycle or seasonality; confidence, which captures the reliability of predictions; and ending time, which marks the completion of the sequence. This structured representation supports both descriptive analysis and predictive modeling.
In recent years, time series prediction has become central to numerous applications that rely on numerical or quantitative data, including finance, healthcare, manufacturing, and information systems. Depending on the analytical objective, predictions may be carried out across different temporal granularities, typically categorized into three types of time spaces: short-term, focusing on immediate or near-future values; medium-term, capturing intermediate trends; and longterm, which projects developments over extended periods. Each of these forecasting horizons has distinct challenges and is suited to different decision-making contexts. Time series prediction has been used in many applications that deal with numerical data in recent years. The prediction can be made using one of three different time spaces:
• Short-Term period
• Mid-Term period
• Long-Term period
In earlier studies, a wide range of machine learning algorithms and models have been employed to forecast temporal data. Some approaches adopt hybrid strategies, combining multiple models or integrating optimization algorithms to enhance predictive accuracy. This paper presents a comprehensive survey of forecasting models applied across different types of time series datasets, including electricity demand, traffic flow, student enrollment, stock market trends, and other seasonal or trend-driven data. The surveyed methods are classified and analyzed to highlight their applications, strengths, and limitations.
To provide clarity and structure, the remainder of the paper is organized as follows: Section II presents an in-depth literature review of IoT-based time series systems; Section III discusses the advantages and limitations of existing forecasting approaches; and Section IV concludes the study and outlines potential directions for future research.

LITERATURE SURVEY

A. Forecasting Techniques and Applications

Forecasting methods for time series data are often classified into three major groups: stochastic models, soft computing–based approaches, and fuzzy-based techniques. Each category brings unique strengths in addressing the challenges of temporal data, such as nonlinearity, noise, and dynamic variations.

In the domain of electricity market forecasting, several studies have combined machine learning and statistical approaches to improve accuracy. For instance, Support Vector Machine (SVM) has been widely integrated with the AutoRegressive Moving Average with External Input (ARMAX) model to create hybrid frameworks [6,7]. Using data from the PJM interconnected electricity market, researchers compared single models such as SVM, Least Squares SVM (LSSVM), and ARMAX against the hybrid SVM-ARMAX approach. The hybrid model consistently outperformed the standalone techniques, demonstrating its robustness in capturing both linear and nonlinear dependencies. This is largely attributed to SVM's ability to manage outliers effectively during training while incorporating linear modules to enhance prediction accuracy.

Further studies extended the application of SVM and LSSVM for mid-term electricity market clearing price (MCP) prediction [8]. Compared to Artificial Neural Networks (ANNs) and Bayesian models, SVM showed notable advantages, including resistance to overfitting, better handling of out-of-sample data, and reduced modeling complexity. Using PJM market data from 2009, experiments demonstrated that SVM and LSSVM could reliably forecast hourly MCP values, with LSSVM offering simplified computation through linear optimization, while SVM employed quadratic programming for model training.

To improve parameter selection in LSSVM models, Rubio et al. [9] introduced a heuristic method using Non-parametric Noise Estimation (NNE) to approximate Gaussian kernel parameters. Cross-validation with optimization schemes such as Levenberg–Marquardt (LM) and Conjugate Gradient (CG) further refined model performance [10]. Multiple-SVM and multiple-LSSVM architectures were also explored for electricity price forecasting [11]. By dividing market prices into categories such as low, medium, high, and peak, researchers demonstrated that multi-layered LSSVM structures were particularly effective at capturing price spikes compared to singlemodel approaches.

Beyond energy markets, SVM-based methods have also been widely applied in financial time series forecasting. Fan et al. introduced a coarse-grained SVM classifier to categorize stocks as "winning" or "losing," followed by regression analysis to estimate returns and rankings. Later, hybrid models combining Support Vector Regression (SVR) with Genetic Algorithms (GA) were proposed to optimize hyperparameters and enhance prediction accuracy [12,13]. These hybrid methods demonstrated superior performance in stock index forecasting compared to conventional SVM or ANN models. Huang further advanced this approach by employing SVR with GA-based feature selection for stock ranking and portfolio optimization.

Other hybrid approaches have integrated optimization algorithms with regression models. For example, an RBFAR (Radial Basis Function–AutoRegressive) model combined gradient-based optimization with GA to predict retail sales data from the U.S. Census Bureau [14]. Comparisons with ANN, ARIMA, TDNN, and SVR models revealed that hybrid methods could better capture both trend and seasonal components in complex time series. Similarly, neural networks have been explored extensively for seasonal and trend-based forecasting [15–17], with models such as C. Hamzacebi's hybrid NN framework showing improvements in handling seasonal variations [16].

In addition to soft computing methods, fuzzy-based models have emerged as a powerful tool for handling uncertainty in temporal data. The Fuzzy Time Series Genetic Algorithm (FTSGA) model [23] combined fuzzy logic with GA optimization techniques, improving accuracy in stock forecasting using the TAIEX dataset. Similarly, Egrioglu et al. [24] proposed a hybrid approach integrating Fuzzy C-Means (FCM) clustering with ANN to generate fuzzy rules and predict enrollment data at the University of Alabama. These methods showed strong performance in handling non-stationary and nonlinear datasets. Fuzzy time series prediction has also been refined through modified algorithms. Shah [25] proposed fuzzy IF–THEN rule-based forecasting for trends such as growth, decline, and constancy, tested across datasets including GDP of India and sales data from the propylene industry. Bas et al. [26] introduced a modified GA for fuzzy time series forecasting to overcome the limitations of random mutations in GA. This approach, validated on datasets such as TAIFEX and traffic accident records in Belgium, demonstrated improved reliability over conventional fuzzy-GA hybrids.

More recent studies have integrated Particle Swarm Optimization (PSO) with FCM clustering to further enhance fuzzy time series forecasting [27]. By optimizing fuzzy relationship matrices through PSO, this approach reduced information loss during fuzzification and achieved strong results on stock market and enrollment datasets. These hybrid fuzzy–soft computing frameworks highlight the potential of combining clustering, evolutionary algorithms, and fuzzy logic to tackle the inherent uncertainty and variability in time series data.

DISCUSSION AND ANALYSIS

Analyzing and modeling time series data presents several challenges, including the vast size of datasets, high dimensionality, non-stationarity, volatility, and the difficulty of making reliable long-term predictions. Designing forecasting systems that are both accurate and computationally efficient remains a central concern. Numerous researchers and statisticians have proposed different techniques to address these challenges, ranging from classical statistical models to advanced machine learning and hybrid frameworks [28,29,30,31,32].

One recurring observation across the literature is that no single model consistently performs best across all domains. Statistical models such as ARIMA work well for stationary and linear data but struggle with nonlinear and highly volatile datasets. On the other hand, machine learning and neural network–based approaches capture complex patterns effectively but are prone to overfitting, require significant computational resources, and often lack interpretability.

Hybrid methods have emerged as a promising solution by combining the strengths of multiple approaches while compensating for their weaknesses. For future work, ensemble-based forecasting frameworks hold considerable potential. In particular, the Ensemble Regression Algorithm that integrates Reduced Error Pruning Tree (REPTree), Support Vector Machine regression (SMOreg), and Multi-Layer Perceptron (MLP) can provide a balanced tradeoff between accuracy, robustness, and efficiency. REPTree offers fast and interpretable decision rules, SMOreg contributes strong generalization ability for nonlinear relationships, and MLP enhances the capacity to model complex temporal dependencies. By combining these models, an ensemble framework could reduce individual weaknesses while improving predictive reliability across diverse time series domains such as finance, energy, healthcare, and traffic forecasting.

Overall, the discussion suggests that the future of time series analysis lies in adaptive, hybrid, and ensemble-based models that can dynamically handle non-stationary and large-scale data. Coupling these models with optimization algorithms, feature engineering techniques, and domain-specific knowledge will further improve their performance, paving the way for low-cost yet highly accurate forecasting systems.

CONCLUSION

Time series data evolves continuously over time, with variations observed across different intervals such as hourly, daily, weekly, monthly, and yearly. By analyzing these variations, it becomes possible to predict future trends using forecasting models that integrate machine learning algorithms and optimization techniques.

This survey presented a comprehensive overview of the different types of time series data and discussed a wide range of forecasting models applied across multiple domains. The analysis highlighted how classical statistical approaches, machine learning methods, and hybrid techniques have been employed to address challenges such as non-stationarity, high dimensionality, and long-term prediction accuracy.

Looking ahead, more adaptive and ensemble-based frameworks are expected to play a pivotal role in improving forecasting reliability. In particular, the proposed Efficient Ensemble Regression Algorithm (EERA), which integrates Reduced Error Pruning Tree (REPTree), Support Vector Machine regression (SMOreg), and Multi-Layer Perceptron (MLP), offers a promising direction. This combination can exploit the strengths of individual models while mitigating their limitations, thereby enabling more robust and accurate forecasting across diverse application areas such as finance, energy, healthcare, and traffic systems.

REFERENCES

1. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of things: A survey on enabling technologies, protocols, and applications,” IEEE communications surveys & tutorials, vol. 17, no. 4, pp. 2347–2376, 2015.
2. J. Manyika, M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon, “The internet of things: Mapping the value beyond the hype,” McKinsey Global Institute, Tech. Rep., 06 2015
3. Hector Espinoza, Gerhard Kling, Frank McGroarty, Mary O’Mahony,´ Xenia Ziouvelou,”Estimating the impact of the Internet of Things on productivity in Europe,”Heliyon,Volume 6, Issue 5, 2020, e03935, ISSN 2405-8440,
4. G. Mahalakshmi, S. Sridevi and S. Rajaram, ”A survey on forecasting of time series data,” 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE’16), Kovilpatti, India, 2016, pp. 1-8, doi: 10.1109/ICCTIDE.2016.7725358.
5. Liu, L. M., Bhattacharyya, S., Sclove, S. L., Chen, R., & Lattyak, W. J. (2001). Data mining on time series: an illustration using fastfood restaurant franchise data. Computational Statistics & Data Analysis, 37(4), 455-476.
6. Xing Yan, Nurul Chowdhury; “Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and autoregressive moving average with external input”,Journal of Electrical Power and Energy Systems, Vol. 63, pp. 64-70,December 2014.
7. Xing Yan, Nurul A Chowdhury; “Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach”, Electrical Power and Energy Systems, Volume 53, pp. 20-26, December 2013.
8. Xing Yan, Chowdhury; “A Comparison Between SVM and LSSVM in Mid-term Electricity Market Clearing Price Forecasting”, 26th Annual IEEE Canadian Conference of Electrical and Computer Engineering, May 2013.
9. G.Rubio, H.Pomares, I.Rojas, LJ. Herrera; “A heuristic method for parameter selection in LS-SVM: Application to time series prediction”, International Journal of Forecasting, Vol. 27, Issue 4, pp. 725-739, 2011.
10. S.An, W.Liu ,S.Venkatesh; “Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression”, Pattern Recognition, Vol.40, Issue 8, pp. 2154-2162, August 2007.
11. Xing Yan, Nurul A Chowdhury; “Mid-term Electricity Market Clearing Price Forecasting Using Multiple Least Squares Support Vector Machines”, The Institution of Engineering and Technology, Generation, Transmission & Distribution, Vol.8, Issue 9, pp.1572-1582, September 2014.
12. H.Frohlich, O.Chapelle, B.Sch¨ olkopf,” Feature selection for support¨ vector machines by means of genetic algorithms”, Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, pp. 142–148, November 2003.
13. CL.Huang, CJ.Wang, “A GA-based feature selection and parameters optimization for support vector machines”, Expert Systems with Applications, Vol. 31, Issue 2, pp.231–240, August 2006.

14. M.Gan, H.Peng, XP.Dong; “A hybrid algorithm to optimize RBF network architecture and parameters for nonlinear time series prediction”, Applied Mathematical Modelling, Vol. 36, Issue 7, pp. 2911-2919, July 2012.
15. GP. Zhang, M.Qi; “Neural network forecasting for seasonal and trend time series”, European Journal of Operational Research, Vol. 160, Issue 2, pp. 501-514, January 2005.
16. C.Hamzacebi,“Improving artificial neural networks’ performance in seasonal time series forecasting”, Journal of Information Science, Vol. 178, Issue 23, pp. 4550-4559, December 2008.
17. TT.Temizel, Matthew C. Casey M; “A comparative study of autoregressive neural network hybrids”, Journal of Neural Networks, Vol. 18, Issue 5-6, pp. 781-789, July 2005.
18. M. Gan, Y.Cheng, K.Liu, GL.Zhang; “Seasonal and trend time series forecasting based on a quasi-linear autoregressive model”, Applied Soft Computing, Vol. 24, pp. 13-18, November 2014.
19. TS. Quah, B.Srinivasan, “Improving returns on stock investment through neural network selection”, Expert Systems with Applications ,Vol. 17, Issue 4,pp.295–301,November 1999.
20. N.Chapados,Y.Bengio,“Costfunctions and model combination for VaRbased asset allocation using neural networks”, IEEE Transactions on Neural Networks Vol. 12, Issue 4, pp.890–906, July 2001.
21. K.Kim, I.Han, “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stockprice index”, Expert Systems with Applications”, Vol.19, Issue 2, pp.125–132, August 2000.
22. H.Peng, T.Ozaki , VH.Ozaki V, Y.Toyoda; “A parameter optimization method for radial bias function type models”, IEEE Transactions on Neural Networks, Vol. 14, Issue 2,pp. 432-438, March 2003.
23. QiSen Cai, Defu Zhang, Bo Wu, Stephen CH Leung; “A novel stock forecasting model based on fuzzy time series and genetic algorithm”, Procedia Computer Science, Vol. 18, pp. 1155-1162, 2013.
24. E.Egrioglu, CH.Aladag, U.Yolcu, “Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks”, Expert Systems with Applications, Vol. 40, Issue 3,pp. 854-857, February 2013.
25. M.Shah; “Fuzzy based trend mapping and forecasting for time series data”, Expert Systems with Applications, Vol. 39, Issue 7, pp. 63516358, June 2012.
26. E.Bas, VR. Uslu, U.Yolcu, E.Egrioglu; “A modified genetic algorithm for forecasting fuzzy time series”, Application Intelligence, Vol. 41, Issue 2, pp. 453-463, September 2014.
27. CH. Aladag, U. Yolcu, E.Egrioglu, Ali Z Dalar; “A new time invariant fuzzy time series forecasting method based on particle swarm optimization”, Applied Soft Computing, Vol. 12, Issue 10, pp. 3291- 3299, October 2012.
28. R. G. Brown, Smoothing, forecasting and prediction of discrete time series. Courier Corporation, 2004.
29. C. C. Holt, “Forecasting seasonals and trends by exponentially weighted moving averages,” International journal of forecasting, vol. 20, no. 1, pp. 510, 2004.
30. G. E. Box and G. M. Jenkins, Time series analysis: forecasting and control, revised ed. Holden-Day, 1976.
31. B. Abraham and J. Ledolter, “Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures,” International Statistical Review/Revue Internationale de Statistique, pp. 5166, 1986.
32. D. F. Findley, K. C. Wills, and B. C. Monsell, “Seasonal adjustment perspectives on a damping seasonal factors: Shrinkage estimators for the x-12-arima program a,” International Journal of Forecasting, vol. 20, no. 4, pp. 551556, 2004.