Analysis of Machine Learning-Based Methods for Network Traffic Anomaly Detection and Prediction

Jingyao Wang[a]
Santa Monica College, California, 90401, U.S.A.

Keywords: Machine Learning, Network Traffic Analysis, Network Traffic Prediction, Deep Learning, Anomaly Detection.

Abstract: In the era of rapid development of network technology, the volume of network data traffic has grown exponentially. Network traffic analysis and prediction can effectively facilitate network management, enable timely detection of network attacks, and enhance security protection and optimization of internet resources. This paper introduces the current application of machine learning in network traffic anomaly detection and prediction, along with key technologies such as data preprocessing, feature engineering, model evaluation, and optimization. It describes the technological advancements in traditional machine learning and deep learning methods for traffic classification, anomaly detection, and traffic prediction. The paper highlights the challenges faced by machine learning in network traffic analysis and prediction, including data complexity, real-time processing, and privacy protection. To address these challenges, machine learning in network traffic analysis and detection will rely on interdisciplinary collaboration and technological innovation to develop more automated, intelligent models that emphasize privacy protection, model interpretability, and real-time processing capabilities.

1 INTRODUCTION

With the widespread adoption and application of networks, the reception, storage, and processing of Internet data have experienced exponential growth. In this context, network data traffic has expanded rapidly and exhibits certain characteristics. Network traffic not only reflects user behavior and activities but also indicates the state of the network. Accurate detection and predictive modeling of network traffic can help collect user behavior data on time, optimize user experience, and enhance network performance. Detecting abnormal traffic signals can safeguard network security and identify potential network attacks. Therefore, finding an efficient and accurate method for network traffic detection and analysis is essential.

Traditional non-machine learning methods for network traffic detection, such as NetFlow technology, lack predictive capabilities and require further data processing and analysis (Bai, 2024). In an era of explosive growth in network traffic and increasing concerns about user privacy, there is an urgent need to develop new, more secure, and efficient methods for network traffic detection and analysis.

Unlike traditional traffic analysis techniques like NetFlow, machine learning methods can automatically learn the characteristics and patterns of network traffic. By building models, machine learning enables efficient traffic classification, anomaly detection, and even data analysis and traffic prediction. This approach reduces the need for extensive data transmission, improves efficiency, and enhances network performance and resource utilization. Moreover, machine learning can process data locally or in a distributed architecture, eliminating the need to upload data to central servers and avoiding centralized data storage, thereby enhancing network security.

This paper systematically reviews the application of machine learning in network traffic anomaly detection and prediction, focusing on the following aspects: basic characteristics of network traffic, an overview of machine learning methods, specific methods for traffic anomaly detection, data

[a] https://orcid.org/0009-0000-0176-8845

Wang, J.
Analysis of Machine Learning-Based Methods for Network Traffic Anomaly Detection and Prediction.
DOI: 10.5220/0013701800004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 550-554
ISBN: 978-989-758-765-8

preprocessing and feature engineering, model evaluation and optimization, and future development directions.

2 OVERVIEW OF MACHINE LEARNING METHODS

2.1 Basic Concepts of Network Traffic Analysis and Prediction

Network traffic exhibits characteristics such as autocorrelation, burstiness, non-stationarity, and periodicity. Autocorrelation refers to the temporal correlation of network traffic over time. Burstiness indicates that network traffic can increase sharply within a short period. Non-stationarity means that the statistical properties of network traffic change over time. Network traffic analysis primarily involves categorizing traffic into different types, such as video streaming, web browsing, and P2P downloads, to analyze user behavior patterns and create user profiles for behavior prediction. Network traffic prediction faces challenges such as data dynamism, complexity, and uncertainty. Traffic patterns change over time, requiring models to adapt dynamically. Additionally, network traffic is influenced by various factors, including user behavior, network topology, and application types. Noise and outliers in network traffic further complicate prediction.

2.2 Traditional Machine Learning Methods

Traditional machine learning methods include supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning, which uses labeled data to learn the mapping between inputs and outputs, is the primary approach for network traffic analysis and prediction. Common supervised learning methods include decision trees, which classify or regress by constructing tree-like models; support vector machines (SVM), which classify by finding optimal hyperplanes; and naive Bayes classifiers, which use Bayesian probability for classification. These methods, combined with algorithm optimization, yield efficient and accurate results in network traffic analysis and detection.

2.3 Deep Learning Methods

Deep learning methods include neural networks (NN) and their derivatives, such as convolutional neural networks (CNN), recurrent neural networks (RNN), and their variants (e.g., LSTM and GRU), as well as the Transformer architecture. NN achieve complex function mapping through the combination of multiple layers of neurons. CNNs are suitable for feature extraction in image and sequence data, while RNNs and their variants (e.g., LSTM and GRU) are better suited for time-series data. The Transformer architecture, based on self-attention mechanisms, captures long-range dependencies.

2.4 Federated Learning and Distributed Learning

Federated learning is a distributed machine learning method that enables model training across multiple devices or institutions while preserving data privacy. It offers significant advantages in processing distributed network traffic data.

3 MACHINE LEARNING-BASED NETWORK TRAFFIC ANALYSIS METHODS

3.1 Traffic Classification

Traffic classification is a fundamental task in network traffic analysis. Machine learning-based methods include feature engineering-based classification, which extracts statistical features (e.g., packet size, transmission rate) and combines them with traditional machine learning algorithms (e.g., SVM, decision trees) for classification. Deep learning methods such as CNN and RNN excel in encrypted traffic classification. For example, CNNs extract local features, while RNNs capture dynamic temporal characteristics (Cui, 2024).

3.2 Anomaly Detection

Anomaly detection identifies abnormal network traffic for further analysis. Statistical metrics (e.g., mean, variance) can be used to detect anomalies, or different models' statistical metrics can be compared to determine accuracy. For instance, Wang Ruixue compared the accuracy of the GAFSA-SVR model with CPSO-SVR and GA-SVR models using mean relative error and root mean square error (Wang, 2013). Algorithms like random forests and autoencoders are also used for anomaly detection. Random forests handle high-dimensional data, while

autoencoders detect anomalies through reconstruction errors.

3.3 Behavior Analysis

Behavior analysis models user network behavior patterns for user profiling and behavior prediction. Deep learning methods like LSTM capture temporal characteristics of user behavior, enabling precise behavior analysis.

4 MACHINE LEARNING-BASED NETWORK TRAFFIC PREDICTION METHODS

Time series prediction is a primary method for network traffic prediction. The ARIMA model is a classic statistical time series model suitable for stationary time series data. LSTM and GRU, based on RNNs, capture long- and short-term dependencies in time series data.

Deep learning models have made significant progress in traffic prediction. For example, the Transformer architecture captures long-range dependencies through self-attention mechanisms, significantly improving prediction accuracy (Ji, 2024).

Distributed and federated learning methods are crucial in network traffic prediction. Federated learning stores models on client devices and exchanges model parameters with servers at specific times, optimizing models without uploading user data. This approach enhances security in network anomaly detection compared to traditional methods, enabling joint traffic detection across multiple base stations while preserving data privacy.

5 APPLICATIONS AND IMPACT OF MACHINE LEARNING IN NETWORK TRAFFIC ANALYSIS AND PREDICTION

This section discusses three application areas: network traffic classification and anomaly detection, network traffic prediction and resource optimization, and network security situational awareness and early warning.

5.1 Network Traffic Classification and Anomaly Detection

Machine learning plays a vital role in network traffic classification. By analyzing traffic features such as packet size, transmission protocol, and source and destination addresses, machine learning algorithms classify traffic and identify different application types or services. This classification helps network administrators better understand network usage and optimize resource allocation. Sun Yu proposed a phishing attribution analysis based on deep learning interpretability methods, focusing on visual features like website logos, buttons, and navigation bars (Sun, 2024). By leveraging multi-modal data and self-attention mechanisms, the model extracts more discriminative features for phishing website detection.

In anomaly detection, machine learning analyzes historical traffic data to learn normal behavior patterns. When traffic deviates from these patterns, machine learning algorithms quickly identify anomalies, which may indicate network attacks or failures. For example, deep learning algorithms can monitor network traffic, device logs, and signal strength in real-time, accurately locating faults or attack sources with over 98% accuracy. Liu Jingrui proposed a deep clustering model, Cluster-AAE, which learns low-dimensional representations of network traffic data and uses the SNNDC algorithm for clustering analysis to establish behavior rule libraries(Liu, 2024) . Test data is then matched against these rules to predict attack behavior.

5.2 Network Traffic Prediction and Resource Optimization

Machine learning models can predict network traffic fluctuations in advance. By analyzing historical data, such as seasonal traffic trends and holiday traffic peaks, machine learning models accurately forecast future network traffic. This prediction helps network operators allocate resources and bandwidth proactively, optimizing resource utilization during traffic peaks. Wang Yuewen developed a wireless cellular network traffic prediction method based on residual networks and RNNs, using attention modules to optimize the model. The method was implemented in a prototype system for wireless cellular network traffic services (Wang, 2021).

Machine learning models also assist in network resource optimization. For example:

Bandwidth Allocation Optimization: Based on traffic predictions, network operators can adjust bandwidth allocation to ensure sufficient resources during peak periods, avoiding congestion and improving user experience (Zhang, 2024).

Routing Optimization: By analyzing network data, optimal data transmission paths can be

determined, reducing latency and congestion and enabling self-healing networks.

Fault Prediction and Maintenance: Machine learning predicts network component failures by analyzing historical and real-time data, enabling proactive maintenance and reducing network downtime. Ji Jingchan utilized CNNs and RNNs for feature extraction and pattern recognition, identifying abnormal network activities(Ji, 2024). By integrating various machine learning techniques, including neural networks, SVMs, random forests, decision trees, deep learning, and ensemble learning, significant improvements were achieved in traffic prediction, real-time traffic classification, network resource optimization, anomaly detection, and security threat analysis, enhancing the overall performance and efficiency of 5G networks.

5.3 Network Security Situational Awareness and Early Warning

Machine learning is widely used in network security monitoring and early warning systems. Sardar Shan Ali Naqvi proposed a DDoS attack detection model based on multi-level autoencoder feature learning (Naqvi, 2024). Using unsupervised learning, the model combines multiple shallow and deep autoencoders with multi-kernel learning (MKL) to detect DDoS attacks in smart grids, enabling timely security situational awareness and early warning.

Machine learning algorithms are shifting from passive defense to proactive prevention. By analyzing past attack patterns and current system activities, these algorithms can issue warnings before potential threats materialize.

5.4 The Role of Machine Learning in Network Traffic Analysis and Prediction

Machine learning has driven a paradigm shift from rule-based to data-driven approaches in network traffic analysis and detection. Traditional methods relied on predefined rules and feature matching, which struggled to cope with increasingly complex and diverse network attacks. Machine learning, especially deep learning, enables systems to automatically learn and extract features from vast amounts of network traffic data, achieving precise anomaly detection. Moreover, machine learning models are adaptive and scalable, continuously updating and optimizing as network environments and attack methods evolve. This ensures effective network security protection even against new and unknown threats.

Machine learning enhances the accuracy and efficiency of network traffic analysis through automated feature extraction, efficient handling of complex patterns, and real-time data analysis. In network automation, machine learning algorithms are applied to log analysis, fault prediction, and resource optimization, enabling automated and intelligent operations.

6 CHALLENGES AND FUTURE DIRECTIONS

6.1 Challenges in Machine Learning for Network Traffic Analysis and Prediction

Despite significant progress, machine learning in network traffic anomaly detection and prediction faces several challenges. The dynamic and complex nature of network traffic data limits model accuracy and generalization. Real-time analysis of big data demands more efficient algorithms and computational power. Imbalanced data samples, adversarial data, and the need for multi-source data fusion further complicate model generalization. Balancing model performance and interpretability is also a critical issue, requiring tailored approaches based on specific needs and scenarios.

In summary, machine learning in network traffic analysis and prediction faces challenges related to data complexity, real-time processing, and privacy protection (Liu etal, 2024). Addressing these issues requires technological innovation and interdisciplinary collaboration.

6.2 Future Directions for Machine Learning in Network Traffic Analysis and Prediction

Future developments in machine learning for network traffic analysis and detection will focus on the following areas:

The application of deep learning and reinforcement learning is developing rapidly, driving the construction of automated and intelligent decision-making systems. These systems use distributed processing frameworks to enhance their processing capabilities, so that they can efficiently process and analyze large amounts of data. At the same time, with the rise of multi-source data fusion

and cross-domain learning, the system can more comprehensively understand and utilize information from different sources.

In this process, the interpretability, transparency, and visual interface of the model become particularly important, which help users understand the decision-making process of the model and enhance the trust of the system. In addition, the application of real-time processing and stream computing frameworks enables data to be analyzed and processed at the moment of generation, meeting the demand for rapid response.

In order to achieve a higher level of technology, cooperation and innovation in different fields become essential, especially the collaboration between computer science, statistics, and network security. This interdisciplinary cooperation will promote the birth and application of new technologies and lay the foundation for future development.

In the future, machine learning in network traffic analysis and detection will become more automated and intelligent, emphasizing privacy protection, model interpretability, and real-time processing capabilities. Interdisciplinary collaboration and technological innovation will be key drivers of progress in this field.

7 CONCLUSIONS

This paper reviewed the application of machine learning in network traffic anomaly detection and prediction, covering key technologies such as data preprocessing, feature engineering, model evaluation, and optimization. It described the advancements in traditional machine learning and deep learning methods for traffic classification, anomaly detection, and traffic prediction. The paper highlighted the challenges of data complexity, real-time processing, and privacy protection in network traffic analysis and prediction. To address these challenges, machine learning will rely on interdisciplinary collaboration and technological innovation to develop more automated, intelligent models that prioritize privacy protection, model interpretability, and real-time processing capabilities.

REFERENCES

Bai, F., Yao, M., Li, C., 2024. A real-time network traffic analysis system based on big data. In Tianjin Electronic Industry Association, Proceedings of the 2024 Annual Conference of the Tianjin Electronic Industry Association. China Telecom Tianjin Branch; Tianjin Information and Communication Industry Association, 9.

Cui, X., 2024. Research on network anomaly detection method based on efficient federated learning. PhD thesis, Qilu University of Technology. DOI:10.27278/d.cnki.gsdqc.2024.000714.

Ji, J., 2024. Application of machine learning algorithms in 5G network diversion enhancement. Yangtze River Information and Communication, 37(09), 193-195. DOI:10.20153/j.issn.2096-9759.2024.09.057.

Liu, J., 2024. Research on density-based deep clustering algorithm and its application in intrusion detection. PhD thesis, Northwest Normal University. DOI:10.27410/d.cnki.gxbfu.2024.000091.

Liu, W., Wen, B., Ma, M., et al., 2024. A network traffic anomaly detection model based on multiple deep learning fusion. In China Computer Federation, Proceedings of the 39th National Computer Security Academic Exchange Conference. Key Laboratory of Data Science and Smart Education, Ministry of Education; School of Information Science and Technology, Hainan Normal University, 5. DOI:10.26914/c.cnkihy.2024.043726.

Naqvi, A. S. S., 2024. Machine learning-based DDoS attack detection in smart grid. PhD thesis, North China Electric Power University (Beijing). DOI:10.27140/d.cnki.ghbbu.2024.000194.

Sun, Y., 2024. Phishing website detection method based on multimodal information fusion. PhD thesis, Qilu University of Technology. DOI:10.27278/d.cnki.gsdqc.2024.000715.

Wang, R., 2013. Research on the perception and prediction of network traffic based on learning machines. PhD thesis, Jiangnan University.

Wang, Y., 2021. Research on traffic prediction of wireless cellular network based on deep learning. PhD thesis, China University of Mining and Technology. DOI:10.27623/d.cnki.gzkyu.2021.001168.

Zhang, L., Li, X., & Chen, Y., 2024. A hybrid approach for network intrusion detection using deep learning and ensemble methods. Journal of Network and Systems Management, 32(2), 456-478.