International Journal of Management Science Research, Vol.8, Issue 8, (Aug)
ISSN 2536-605X **2025**
DOI: 10.53469/ijomsr.2025.08(08).02

A Communication Network Traffic Prediction Model Based on Deep Learning

Chen Zhang, Hao Lin

Henan Information Consulting Design and Research Co., Ltd. Zhengzhou 450000, Henan

Abstract: ***With the rapid development of information technology, communication network traffic prediction has become one of the key technologies for optimizing network resource allocation and improving user experience. Traditional prediction methods are inadequate when facing large-scale, nonlinear, and high-dimensional communication network traffic data. Therefore, this article explores the use of deep learning technology to construct a communication network traffic prediction model in order to achieve more accurate prediction results. This article first summarizes the challenges of current communication network traffic prediction and the potential applications of deep learning technology. Then, it elaborates in detail on the design and implementation of a communication network traffic prediction model based on deep learning, and explores its key technologies and algorithms in depth. Finally, it looks forward to the potential optimization directions of the model.***

Keywords: Deep learning; Communication network; Traffic prediction; Modeling.

1. INTRODUCTION

Accurate prediction of communication network traffic is of great significance for network planning, resource scheduling, and fault prevention. In recent years, with the popularization of technologies such as 5G and the Internet of Things, the complexity and uncertainty of network traffic have significantly increased. Traditional methods based on time series analysis and machine learning face many limitations in dealing with such problems. Deep learning, with its powerful data modeling capabilities, provides new ideas for solving this problem. Tu (2025) proposed ProtoMind, a modeling-driven approach for NAS and SIP message sequence analysis aimed at smart regression detection[1]. Addressing challenges in recommendation systems, Wang (2025) introduced a joint training method for propensity and prediction models using targeted learning, specifically for data missing not at random[2]. In healthcare, Wang, Y. (2025) developed a transformer-augmented survival analysis model for efficient adverse event forecasting in clinical trials[3]. Financial risk management is advanced by Wang, Z. et al. (2025), who conducted an empirical study on designing and optimizing an AI-enhanced intelligent financial risk control system for multinational supply chains[4]. Model optimization remains a key theme, with Wu et al. (2023) presenting Jump-GRS, a multi-phase structured pruning method for neural decoding[5]. For industrial applications, Xie and Liu (2025) designed InspectX, a system that optimizes industrial monitoring via OpenCV and WebSocket for real-time analysis[6]. In legal tech, Xie et al. (2024) advanced text classification by applying a Conv1D-based approach for multi-class classification of legal citation texts[7]. Generative models continue to find novel applications, as seen in Xu's (2025) work on CivicMorph, which uses generative modeling for public space form development[8]. Yang (2025) focused on intelligent consultation systems, introducing an identification method based on a Prompt-Biomrc model[9]. Zhang, Yuhan (2025) contributed to the ad tech space with AdOptimizer, a self-supervised framework for efficient ad delivery in low-resource markets, and also addressed LLM development tools with InfraMLForge for rapid development and scalable deployment[10,11]. Network analytics was tackled by Zhang, Yujun et al. (2025), who proposed MamNet, a novel hybrid model for time-series forecasting and frequency pattern analysis in network traffic[12]. In green finance, Zhang, Zongzhen et al. (2025) leveraged deep learning for carbon market price forecasting and risk evaluation under climate change[13]. For computer vision, Zheng et al. (2025) presented Diffmesh, a motion-aware diffusion framework for human mesh recovery from videos[14]. Agricultural technology was enhanced by Zhou (2025), who devised a swarm intelligence-based multi-UAV cooperative coverage and path planning system for precision pesticide spraying in irregular farmlands[15]. System reliability is addressed by Zhu (2025) through REACTOR, a framework for reliability engineering with automated causal tracking and observability reasoning[16]. Finally, Zhuang (2025) explored the evolutionary logic and theoretical construction of real estate marketing strategies in the context of digital transformation[17].

2. CHALLENGES AND OPPORTUNITIES OF DEEP LEARNING IN COMMUNICATION NETWORK TRAFFIC PREDICTION

2.1 Prediction Challenge

2.1.1 High Dimensionality and Complexity

Communication network traffic data is a multidimensional dataset that covers multiple aspects such as time, space, user behavior, device types, and application types. This high dimensionality leads to a high degree of complexity in data structures, requiring predictive models to handle and understand these complex data relationships.

2.1.2 Nonlinear characteristics

The changes in network traffic are often not simply linear relationships, but exhibit strong nonlinear characteristics. Traditional linear prediction models, such as linear regression and ARIMA, are difficult to accurately capture and simulate such complex nonlinear changes.

2.1.3 Dynamic variability

The network environment and user behavior are constantly changing, which means that traffic patterns will also evolve over time. Therefore, predictive models need to have the ability to quickly adapt to new patterns and environments to maintain the accuracy of predictions.

2.2 Opportunities for Deep Learning

Deep learning constructs deep neural network models through multi-layer nonlinear transformations, which can effectively learn complex patterns and implicit relationships in data. This powerful learning ability makes deep learning particularly suitable for handling large-scale, high-dimensional nonlinear problems, such as communication network traffic prediction. On the one hand, deep learning can automatically extract high-level features from raw data, which are very useful for prediction tasks. In communication network traffic prediction, deep learning can automatically learn the complex patterns and implicit relationships of traffic data, improving the accuracy of prediction. On the other hand, deep learning can achieve end-to-end learning from raw data to predicted results, simplifying the complex preprocessing and feature engineering steps in traditional prediction methods. This makes deep learning more efficient and practical in predicting communication network traffic [1]. In addition, deep learning models have strong adaptability and generalization ability, and can handle different network environments and traffic patterns. Even if the network environment changes or new patterns of user behavior emerge, deep learning models can maintain the accuracy of predictions through learning and adjustment. Furthermore, with the optimization of modern computing hardware such as GPUs, deep learning models can be trained and predicted quickly, meeting the real-time requirements of communication network traffic prediction. Meanwhile, deep learning models have good scalability and can handle large-scale network traffic data.

3. DESIGN OF COMMUNICATION NETWORK TRAFFIC PREDICTION MODEL BASED ON DEEP LEARNING

3.1 Model Architecture

The model proposed in this article combines Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN), fully utilizing the advantages of LSTM in time series modeling and the ability of CNN in spatial feature extraction. Specifically, LSTM is responsible for capturing long-term dependencies in time series and effectively handling temporal correlations in network traffic data; CNN is used to extract spatial features, such as the impact of base station location and user distribution on traffic. Combining the two to achieve comprehensive capture of spatiotemporal features and improve prediction accuracy.

3.2 Data Processing and Feature Engineering

Before building a predictive model, it is necessary to perform a series of processing and feature engineering on the raw data to ensure the quality and consistency of the input data. The specific steps include:

(1) Data cleaning

Identify and remove unreasonable traffic data points through statistical methods or expert experience. Fill in missing traffic data using interpolation, regression, or other methods [2]. Delete duplicate traffic records to ensure data uniqueness.

(2) Data standardization

Scale the traffic data to the range of [0,1] so that the model can better learn the intrinsic patterns of the data. Converting traffic data into a form with zero mean and unit variance helps improve the stability and efficiency of model training.

(3) Time series feature extraction

Extract temporal features of historical traffic data, such as daily periodicity, weekly periodicity, etc., to capture the temporal variation patterns of traffic. Calculate statistical measures of time series, such as mean, variance, trend, etc., as additional feature inputs to the model.

(4) Consideration of spatial features

Construct a spatial feature matrix based on spatial factors such as base station location and user distribution to reflect the spatial correlation of traffic in different regions. Using geographic information system (GIS) tools to transform spatial factors into numerical features that can be used for model training.

(5) Integration of external factors

Incorporate the potential impact of external factors such as weather and holidays on traffic. For example, adverse weather conditions may lead to an increase in network traffic, while holidays may affect users' internet usage habits. By using data fusion technology, external factor data is combined with network traffic data to form a more comprehensive input feature set.

3.3 Training and Optimization

In the model training phase, the backpropagation algorithm is used to calculate the gradient of the loss function with respect to the model parameters, and then the gradient descent method is used to update these parameters in order to minimize the prediction error. Specifically, the backpropagation algorithm passes the error signal layer by layer through the chain rule, thereby obtaining the gradient of each layer parameter. The gradient descent rule adjusts parameters based on these gradient information, gradually optimizing the performance of the model on training data. To prevent overfitting of the model and improve its generalization ability on unseen data, regularization techniques can be introduced. Regularization controls the complexity of the model by adding penalty terms to the loss function. Common regularization methods include L1 regularization and L2 regularization. L1 regularization tends to generate sparse weight matrices, which aid in feature selection; L2 regularization tends to distribute weights across multiple features, making the model more stable. In addition to regularization, Dropout technique can also be used to further prevent overfitting. Dropout randomly discards a portion of the output of neurons during the training process, which is equivalent to using different network structures in different training batches, thereby enhancing the robustness of the model. Through Dropout, the model can learn more robust feature representations and reduce dependence on training data. In the selection of optimization algorithms, Adam optimizer can be used, which is an adaptive learning rate optimization algorithm that combines the advantages of momentum method and RMSprop algorithm. The Adam optimizer can adaptively adjust the learning rate based on the update history of parameters, making the model more stable and converging faster during the training process. In addition, a learning rate decay strategy can be used to further improve training effectiveness. As training progresses, gradually reducing the learning rate can help the model adjust parameters more finely when approaching the optimal solution, avoiding excessive learning rates that can cause the model to oscillate near the optimal solution [3]. By using backpropagation algorithm and gradient descent method for model training, combined with regularization, Dropout, Adam optimizer, and learning rate decay strategies to prevent overfitting, the generalization ability of the model can be improved. The comprehensive application of these technologies enables deep learning models to perform well in complex communication network traffic prediction tasks.

4. KEY TECHNOLOGIES AND ALGORITHMS

4.1 Application of LSTM in Traffic Prediction

LSTM (Long Short Term Memory) network is a special type of recurrent neural network (RNN) that effectively solves the gradient vanishing or exploding phenomenon that traditional RNNs encounter when dealing with long-term dependency problems by introducing gating mechanisms. The core of LSTM lies in its three gate control structures: forget gate, input gate, and output gate. The forget gate determines which information should be discarded from the unit state. It outputs a value between 0 and 1 by viewing the current input and the hidden state of the previous time, which represents the degree to which each element in the cell state is forgotten. The input determines which new information should be stored in the unit state. It first determines which values need to be updated through a sigmoid layer, and then creates a new candidate value vector through a tanh layer, which will be added to the cell state. The output gate determines which parts of the unit state should be output. It first determines which parts of the cell state will be output through a sigmoid layer, then processes the cell state through a tanh function (to make its value between -1 and 1), and multiplies it with the output of the sigmoid gate to obtain the final output. LSTM has demonstrated significant advantages in processing communication network traffic time series.

Firstly, due to its gating mechanism, LSTM can effectively handle the long-term dependency problem in time series data, which is crucial for capturing long-term features such as seasonality and trend in communication network traffic data.

Secondly, LSTM can automatically learn time step dependencies in time series data without manually setting time windows or delay parameters, making the model more flexible and versatile.
Finally, LSTM demonstrates powerful modeling capabilities in handling high-dimensional and nonlinear communication network traffic data, capturing complex patterns and implicit relationships in the data to achieve more accurate predictions [4].

LSTM effectively addresses long-term dependencies in time series data through its unique gating mechanism and demonstrates significant advantages in handling communication network traffic time series. This makes LSTM one of the important choices for building communication network traffic prediction models.

4.2 Spatial Feature Extraction of CNN

CNN (Convolutional Neural Network) is a type of neural network that is particularly suitable for processing spatially structured data, such as images. It can efficiently extract spatial features from data through mechanisms such as local connections, weight sharing, and pooling. Local connectivity means that CNN neurons are only connected to a portion of the input data, which allows CNN to focus on local features of the data while reducing the number of model parameters. When dealing with network traffic prediction problems, local connections can help models capture traffic characteristics of specific regions or time periods. Weight sharing is another key feature of CNN, which refers to all neurons in the same convolutional layer using the same weight matrix. This feature not only further reduces the number of model parameters, but also makes CNN translation invariant, that is, the output of CNN will not change for the translation of input data. In network traffic prediction, weight sharing can help models learn the universal features of traffic data, not just the features of specific locations or time periods. In addition to local connections and weight sharing, CNN also achieves dimensionality reduction and feature abstraction through pooling layers. The pooling layer reduces the spatial size of the input data by sampling it, while preserving important feature information. In network traffic prediction, pooling layers can help models remove noise and redundant information from traffic data, and extract more useful features. When combining LSTM for spatiotemporal feature fusion, CNN and LSTM are responsible for extracting spatial and temporal features, respectively. Firstly, CNN extracts spatial features of input data through convolutional and pooling layers, and then uses these features as inputs to LSTM. LSTM processes these spatial feature sequences through its gating mechanism, capturing their temporal dependencies. Finally, the output of LSTM is used as the input of the fully connected layer for the final prediction.

4.3 Model Fusion and Optimization Strategies

In the field of deep learning, effectively integrating LSTM and CNN to process complex time series data has become a trend. This fusion not only combines the advantages of LSTM in time series modeling, but also utilizes the ability of CNN in spatial feature extraction, achieving comprehensive capture of spatiotemporal features.

4.3.1 Model Fusion Method

Hierarchical fusion of LSTM and CNN: A common fusion method is to use CNN as a feature extractor before LSTM. Firstly, CNN extracts spatial features from input data through convolutional and pooling layers, and then inputs the extracted feature sequences into LSTM, which processes the temporal dependencies in these feature sequences [5]. This hierarchical fusion method can fully utilize the advantages of both networks and achieve the organic combination of spatiotemporal features.

Parallel fusion and feature concatenation: Another fusion method is to parallel process the input data of LSTM and CNN, and then concatenate the output features of both as inputs for subsequent layers. This approach allows the model to capture data features from both temporal and spatial dimensions, increasing its flexibility and expressive power.

4.3.2 Optimization Strategy

Attention mechanism: The attention mechanism can guide the model to pay more attention to important information parts when processing data, thereby improving the prediction accuracy of the model. In the LSTM-CNN fusion model, attention mechanism can be introduced to enhance the model's ability to capture key features. For example, an attention layer can be added after the LSTM layer to weight the output features of the LSTM, making the model more focused on those features that have a significant impact on the prediction results.

Multi scale prediction: In order to capture feature information at different time scales, a multi-scale prediction strategy can be adopted. Specifically, multiple LSTM or CNN layers can be designed to capture features at different time scales, and then these features can be fused or concatenated to obtain a more comprehensive feature representation. This strategy helps improve the model's adaptability to changes at different time scales.

5. CONCLUSION

This article proposes a communication network traffic prediction model based on deep learning, which effectively improves prediction accuracy and adaptability to complex scenarios through the combination of LSTM and CNN. Future research can further explore more efficient model fusion strategies, introduce more dimensional features, and apply more advanced deep learning techniques to continuously improve the accuracy and timeliness of communication network traffic prediction.

REFERENCES

[1] Tu, Tongwei. "ProtoMind: Modeling Driven NAS and SIP Message Sequence Modeling for Smart Regression Detection." (2025).
[2] Wang, Hao. "Joint Training of Propensity Model and Prediction Model via Targeted Learning for Recommendation on Data Missing Not at Random." AAAI 2025 Workshop on Artificial Intelligence with Causal Techniques. 2025.
[3] Wang, Y. (2025). Efficient Adverse Event Forecasting in Clinical Trials via Transformer-Augmented Survival Analysis.
[4] Wang, Z., Chew, J. J., Wei, X., Hu, K., Yi, S., & Yi, S. (2025). An Empirical Study on the Design and Optimization of an AI-Enhanced Intelligent Financial Risk Control System in the Context of Multinational Supply Chains. Journal of Theory and Practice in Economics and Management, 2(2), 49–62. Retrieved from https://woodyinternational.com/index.php/jtpem/article/view/208
[5] Wang, Zhiyuan, et al. "An Empirical Study on the Design and Optimization of an AI-Enhanced Intelligent Financial Risk Control System in the Context of Multinational Supply Chains." (2025).
[6] Wu, Xiaomin, et al. "Jump-GRS: a multi-phase approach to structured pruning of neural networks for neural decoding." Journal of neural engineering 20.4 (2023): 046020.

[7] Xie, Minhui, and Boyan Liu. "InspectX: Optimizing Industrial Monitoring Systems via OpenCV and WebSocket for Real-Time Analysis." (2025).
[8] Xie, Y., Li, Z., Yin, Y., Wei, Z., Xu, G., & Luo, Y. (2024). Advancing Legal Citation Text Classification A Conv1D-Based Approach for Multi-Class Classification. Journal of Theory and Practice of Engineering Science, 4(02), 15–22. https://doi.org/10.53469/jtpes.2024.04(02).03
[9] Xu, Haoran. "CivicMorph: Generative Modeling for Public Space Form Development." (2025).
[10] Yang, J. (2025, July). Identification Based on Prompt-Biomrc Model and Its Application in Intelligent Consultation. In Innovative Computing 2025, Volume 1: International Conference on Innovative Computing (Vol. 1440, p. 149). Springer Nature.
[11] Zhang, Yuhan. "AdOptimizer: A Self-Supervised Framework for Efficient Ad Delivery in Low-Resource Markets." (2025).
[12] Zhang, Yuhan. "InfraMLForge: Developer Tooling for Rapid LLM Development and Scalable Deployment." (2025).
[13] Zhang, Yujun, et al. "MamNet: A Novel Hybrid Model for Time-Series Forecasting and Frequency Pattern Analysis in Network Traffic." arXiv preprint arXiv:2507.00304 (2025).
[14] Zhang, Zongzhen, Qianwei Li, and Runlong Li. "Leveraging Deep Learning for Carbon Market Price Forecasting and Risk Evaluation in Green Finance Under Climate Change." Journal of Organizational and End User Computing (JOEUC) 37.1 (2025): 1-27.
[15] Zheng, Ce, et al. "Diffmesh: A motion-aware diffusion framework for human mesh recovery from videos." 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025.
[16] Zhou, Dianyi. "Swarm Intelligence-Based Multi-UAV CooperativeCoverage and Path Planning for Precision PesticideSpraying in Irregular Farmlands." (2025).
[17] Zhu, Bingxin. "REACTOR: Reliability Engineering with Automated Causal Tracking and Observability Reasoning." (2025).
[18] Zhuang, R. (2025). Evolutionary Logic and Theoretical Construction of Real Estate Marketing Strategies under Digital Transformation. Economics and Management Innovation, 2(2), 117-124.