Spatio-Temporal Failure Prediction Using LSTGM for Optical Networks

Cheng Xing[1], Chunyu Zhang[1], Yu Wang[2], Zhiyan Duan[3], Wenjie Song[2], Min Zhang[1], and Danshi Wang[1,*]
[1] *State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing, 100876, China;*
[2] *The Intelligent Network Innovation Center of Chinaunicom;*
[3] *China United Network Communications Group Co., Ltd.*
*Email： *danshi_wang @bupt.edu.cn*

Abstract: A Latent Spatio-Temporal Graph Model is proposed for failure prediction in optical networks, which can effectively learn both spatial and temporal distribution of real equipment performance data and achieve F1-score up to 0.9745. © 2023 The Author(s)

1. Introduction

In recent years, machine learning techniques have been widely applied to failure prediction in optical networks [1]. Researchers have explored the failure prediction performance of Gated Recurrent Unit (GRU), Random Forest, and eXtreme Gradient Boosting algorithms, with a focus on capturing time series data [2-3]. However, in real optical network systems, in addition to time series data, there is also topological information that contains node performance parameter information and connectivity information. These pieces of information are crucial for understanding and managing the structure, performance, and operation of optical networks. Therefore, there is an urgent need to propose a method that can simultaneously learn the spatio-temporal information of optical network device data to improve failure prediction performance.

Typical recurrent neural network methods are usually employed to learn the correlations between time series information and struggle to capture spatial information. On the other hand, graph neural network (GNN) methods are typically used to learn spatial information [4], but they struggle to capture time series information. Unlike typical GRU and GNN approaches, the Latent Spatio-Temporal Graph Modeling (LSTGM) is an integrated model that combines spatial graph convolution and time series model [5]. It can learn temporal and spatial information from time series data without the need for prior topological knowledge.

In this paper, a failure prediction scheme based on LSTGM is proposed. The proposed scheme is evaluated on a realistic dataset from operational networks to assess its predictive performance. Typical classification metrics such as F1-score, accuracy (ACC), false negative rate (FNR), and false positive rate (FPR) are employed as evaluation metrics. Additionally, the failure prediction performance of the GRU algorithm that exclusively learns temporal information and the GNN algorithm that exclusively learns topological information are compared with LSTGM. Experimental results demonstrate that the LSTGM approach can effectively capture the spatio-temporal characteristics of failure data and achieve superior predictive performance.

2. LSTGM-Based Spatio-Temporal Failure Prediction

The Schematic of LSTGM-based spatio-temporal failure prediction scheme is shown in Fig.1. The implementation process of the scheme mainly includes three key steps: data collection, spatiotemporal learning and failure prediction. Firstly, the data acquisition module is used to collect operational data from optical network devices. Secondly, the LSTGM module is employed to learn the temporal and spatial features of the input data. Finally, the prediction module is utilized to forecast the operational state of the devices. LSTGM module that aims to reveal potential spatial correlations between multiple optical sensor variables by learning vertex embeddings. Meanwhile, under the framework of the improved GRU, the module captures the unique temporal characteristics of each variable. Specifically, spatial features between multiple data variables are extracted by spatial graph convolution to aggregate the vertex information of the neighborhood while maintaining the weights sharing attribute, as shown in Fig. 1(c). We utilize the adjacency matrix to determine the importance of neighbor vertices in spatial graph convolution.

Learning to aggregate features is an essential part of the LSTGM module. Define the neighbor set of v_i as $N(v_i)=\{v_j \mid A_{i,j}>0\}$, the vertex features take a shared linear transformation which is parameterized by $W \in R^{C_{in}\times C_{out}}$, with C_{in} and C_{out} denoting the number of input channels and output channels, respectively. Given the input signal X_t, the aggregated feature at v_i is computed as follows:

$$Aggregate(X_t, v_i) = \sum_{v_j \in N(v_i) \bigcup \{v_i\}} \frac{1}{K(v_i, v_j)} \tilde{A}_{i,j} W^T X_{t,j} \tag{1}$$

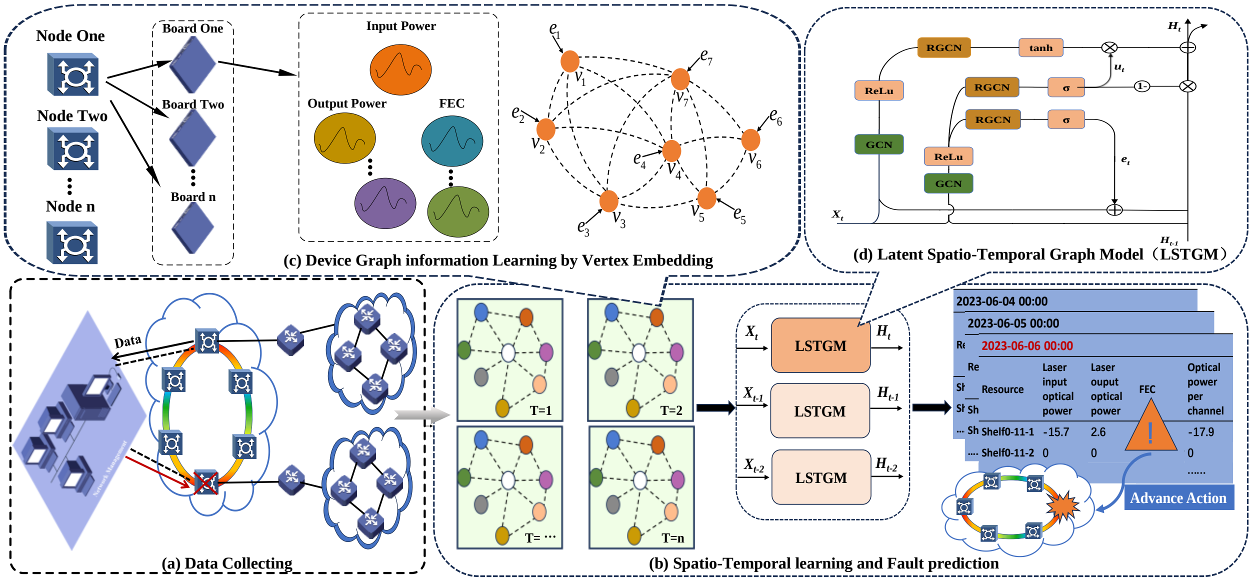

Fig. 1: (a) Data collecting, (b) Spatio-Temporal learning and failure prediction, (c) Device Graph information learning by vertex embedding, (d) Latent Spatio-Temporal Graph Model (LSTGM).

where the normalizing item $K(v_i,v_j)$ is equal to the cardinality of the subset that contains v_i and v_j . It is utilized to balance the contributions of different subsets to the output values. In order to solve the problem of indistinguishable convergent vertex features and over smoothing in backpropagation, the key concept of residual learning is adopted to facilitate a deep GNN structure. A Residual Graph Convolutional Network block (RGCN) is constructed by performing skipping connections on three cascaded Graph Convolutional Network (GCN) layers with *ReLU* activation functions. The adjacency matrix is then attained by calculating the pairwise embedding similarities. Formally, we denote the embedding matrix as $E=(e_1,\dots,e_n)^T$, where e_n is the embedding vector for the vertex v_i. To jointly model the spatio-temporal dependency, we replace the matrix multiplications in Gated Recurrent Units (GRU) with RGCN. Then, the details of the spatio-temporal feature extraction process are formulated as follows:

$$e_t = (\sigma \circ RGCN \circ ReLU \circ GCN)[X_t, H_{t-1}] \tag{2}$$

$$u_t = (\sigma \circ RGCN \circ ReLU \circ GCN)[X_t, H_{t-1}] \tag{3}$$

$$\overline{H_t} = (\tanh \circ RGCN)[X_t, e_t \odot H_{t-1}] \tag{4}$$

$$H_t = (1-u_t) \otimes H_{t-1} \oplus u_t \otimes \overline{H_t} \tag{5}$$

where $H_t \in R^{N\times Ch}$ is the hidden state of LSTGM, with C_h denoting the hidden state size. $\circ$, $\otimes$ and $\oplus$ denote function composition, element-wise multiplication, and element-wise addition, respectively. $\sigma(\cdot)$, $ReLU(\cdot)$ and *tanh* denote the sigmoid, *ReLU* and *tanh* activation function, respectively. The overall structure of LSTGM is presented in Fig. 1(d). The additional GCN layers are used to transform the vertex signals into the hidden state size.

3. Experimental Results and Analysis

In this section, the performance of LSTGM-based failure data prediction in optical networks is evaluated and analyzed. Our dataset is collected over a set of 186 nodes of a real OTN network operated by the Intelligent Network Innovation Center of Chinaunicom, where each node houses multiple boards and each board contains multiple ports. Data is failure-marked based on expert experience, and the ratio of failure data to normal data is 5%. Performance data for each board is recorded every 15 minutes, and the overall data collection period spans between June 1st 2023 and July 1st 2023. After eliminating invalid and missing values, 19 valid features remained.

Firstly, the change of the loss function of the proposed LSTGM method at different times is shown in Fig. 2(a). Moreover, by counting the confusion matrix, we can obtain the predicted accuracy of each day. During the training process of the model, the curves of the training loss function and the testing loss functions are gradually parallel to the X-axis, as well as accuracy, indicating that the model converges well.

Secondly, in order to evaluate the effect of the failure prediction scheme based on LSTGM learning spatio-temporal features, we conduct a comparative analysis of LSTGM with typical methods such as GNN and GRU. We evaluate the performance of the proposed scheme using classic statistical indicators including ACC, F1-score, FNR, and FPR, as shown in Fig. 2(b). It is clear from Fig. 2(b) that the LSTGM-based approach outperforms other methods on metrics such as ACC, F1-score, FNR, and FPR. In contrast, there was a 2.74% improvement in

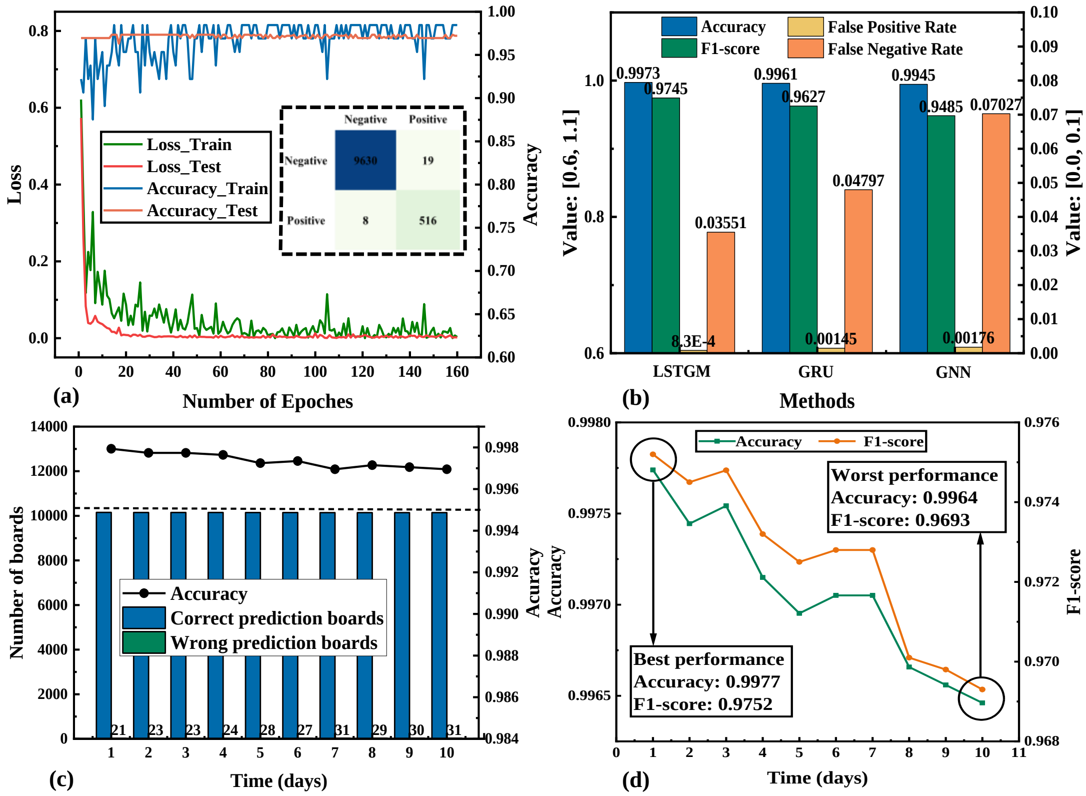

Fig. 2:(a) Loss function for the model; (b) Prediction results on different methods; (c) Number of predicted boards compared with actual boards, prediction accuracy, and confusion matrix based on LSTGM; (d) Prediction index score under different forecast days.

score compared to the GNN method, which solely learned spatial features, and a 1.23% improvement compared to the GRU method, which solely learned temporal features. This suggests a significant enhancement, demonstrating the LSTGM method's capability to effectively learn the temporal and spatial characteristics of optical network device data simultaneously.

Moreover, we compared the number of OTN boards accurately and incorrectly predicted by the LSTGM method in the last 10 days of failure prediction, as shown in Fig. 2(c). Through the analysis of Fig. 2(c), it is found that the operating states of most OTN boards are correctly predicted based on the LSTGM method. The lowest accuracy in the last 10 days is 0.9969. Therefore, the proposed scheme can effectively realize failure prediction.

Furthermore, changes in the number of days predicted for failure also have a significant impact on the results. To explore the effect of prediction days on results of LSTGM method, we compared the results of ACC and F1-score predicted for 1-10 days, respectively, as shown in Fig. 2(d). According to the analysis, it is observed that the scores of the predictive metrics decrease with an increase in the prediction time. When predicting only 1 day, the F1-score and ACC are the highest, while the scores are the lowest when predicting 10 days.

4.Conclusion

We proposed a LSTGM-based failure prediction scheme for optical networks, which can effectively learn both spatial and temporal features of data in optical networks. Moreover, we compared the influence of different forecast days on LSTGM algorithm, and analyzed the number of correct and incorrect prediction boards. The classical predictors (accuracy, F1-score, false negative rate, false positive rate) were used to compare LSTGM with GRU and GNN methods, which proved the ability of LSTGM to learn temporal and spatial characteristics at the same time.

Acknowledgements: This work was supported in part by National Natural Science Foundation of China (No. 62301061, 62171053).

5. References

[1] Z. Wang et al. "Failure prediction using machine learning and time series in optical network". Opt Express, 2017.
[2] C. Zhang et al., "Temporal data-driven failure prognostics using BiGRU for optical networks," J Opt Commun Netw, vol. 12, no. 8, 2020.
[3] C. Zhang et al., "Interpretable Learning Algorithm Based on XGBoost for Failure Prediction in Optical Network," OFC, 2020.
[4] F. Donnangelo et al. "Predicting loss in optical transport segments: A GNN-GRU approach for a nationwide optical network," ICTON, 2023.
[5] J. Yang et al. "Learning Hierarchical Spatial-Temporal Graph Representations for Robust Multivariate Industrial Anomaly Detection," in IEEE Transactions on Industrial Informatics, vol. 19, no. 6, June 2023.