Transformer-based Alarm Context-Vectorization Representation for Reliable Alarm Root Cause Identification in Optical Networks

Jinwei Jia, Danshi Wang*, Chunyu Zhang, Hui Yang, Luyao Guan, Xue Chen, and Min Zhang

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China, email: *danshi_wang@bupt.edu.cn

Abstract *A Transformer-based alarm context-vectorization representation technique is proposed for alarm root cause identification and correlation analysis. Three common root alarms are identified with an accuracy of 94.47%, and other correlated alarms are obtained with quantified correlation degrees.*

Introduction

In optical networks, the alarm analysis is essential for network operation, administration, and maintenance. When a fault occurs, the alarm root cause identification is extremely important for fault location and troubleshooting. However, the modern network management system (NMS) often received massive false or nuisance alarms distributing in different network layers with different severity levels, which brings great challenges to alarm root cause identification among these large amounts of redundant and intricate alarms.

To overcome above challenges, several schemes have been proposed for alarm correlation analysis and alarm root cause identification. They can be mainly classified into three types: 1) the case-based reasoning method that was based on the historical case base[1]; 2) the rule-based method that relied on the correlative rule set and summarized engineer expertise[2]; 3) the graph-based method that was derived from an alarm knowledge graph[3]. However, there are two major problems in these methods in terms of the alarm representation. Alarm representation refers to a mapping from discrete alarms to numerical data which can be easily processed by the model. The first is that the alarm representations of cases, rules, and graphs are strongly dependent on the engineering knowledge and expertise, which have been summarized as the alarm information manual (AIM) in the form of text data. The second is all these alarm representations from cases, rules, or graphs can only be abstracted into symbolic representations, such as the identity (ID) or abbreviated name of the alarm. This kind of symbolic representation cannot express the detailed description of the alarm context in AIM, which leads to the original information loss and identification models degradation. Thus, the prospective alarm analysis is still desiring the more advanced techniques to develop the capacity of intelligent operation without engineer intervention, high-fidelity representation without information loss, and accurate identification model without high complexity.

Recently, with the rapid development of deep learning-driven natural language processing (NLP), the current technology can understand words and sentences like humans, and has the potential to replace engineers to construct expertise. Accordingly, with the help of NLP, it is possible to directly use the alarm context as model's input, and let the model 'read' sentences of each alarm context and 'understand' semantics for getting a vector to identify root cause alarm. This kind of alarm representation based on context-vectorization will retain the original information of multiple dimensions to the greatest extent, and it can also facilitate the use of downstream models.

In this paper, the NLP technology is introduced to vectorize the alarm context for high-fidelity alarm representation. The powerful BERT (Bidirectional Encoder Representations from Transformers) is adopted as pre-training model to vectorize the alarm context, and the Transformer Encoder is used to realize alarm root cause identification. Experimental results show that by the combination of alarm vector and Transformer Encoder, three main root faults are identified with a satisfactory accuracy, and other correlated alarms are also digged out with quantified correlation degrees.

Operating Principle

When a failure occurs in optical network, the MS will receive *N* alarms reported from multiple network elements (NEs), and record them in the

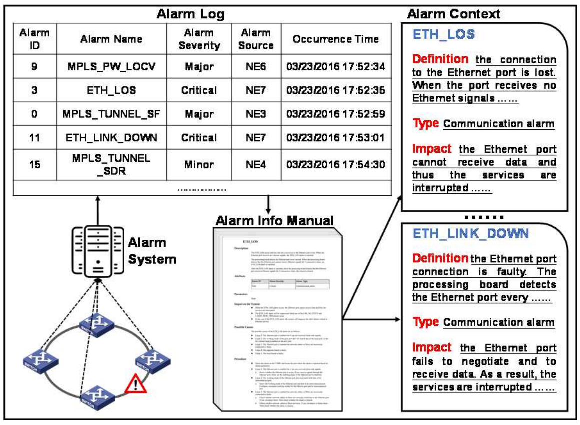

Fig. 1: Examples of alarm data and alarm context.

978-1-6654-3868-1/21/$31.00 ©2021 IEEE

alarm log, as shown in Fig. 1. In the alarm log, an alarm can only be represented by the alarm ID, alarm name, alarm severity, alarm source, and occurrence time. However, as mentioned above, these symbolic representations cannot accurately characterize the alarm information. Hence, in our approach, the official alarm description in AIM is used to represent various alarms, which is called as *alarm context* in text form. In Fig. 1, the alarm context includes the *alarm definition*, *alarm type*, and *impact on the system*. The schematic diagram of alarm root cause identification scheme based on the context-vectorization representation is illustrated in Fig. 2. First, we form an alarm transaction according to the occurrence time. Then, the model will directly read the alarm context for getting a vector to represent the alarm instead of symbolic representation from expertise. Finally, an identification model suitable for vectorization alarm representation is implemented. The whole structure consists of the following three modules.

a) Alarm Encoder

Alarm encoder is used to mine the alarm context and obtain vector representations, where the core is BERT model. BERT is a Transformer-based machine learning technique for NLP pre-training developed by Google[4]. It can dynamically generate word vectors according to the semantic of the word, solve the problem of polysemous words, and contain the positional and semantic information of the word to better express the information. Actually, BERT has been officially trained well by Google, and the parameters of the model are also open. Thus, the BERT pre-trained model can be used to extract features from the alarm context.

In our task, we input the words of the alarm context into the BERT model (see Fig. 2). The BERT model is powered by the multi-head self-attention mechanism of Transformer Encoder to extract vector expressions. Then the word vectors output by the model will be performed by MEAN_POOLING operation, and then the generated average matrix is the alarm vector[5]. In order to distinguish the same alarm in the transaction, after obtaining the vectorization representations of all alarms in the transaction, we also add the alarm vector and the embedded representation of the alarm NEs' ID to get the final output of the alarm encoder, which is represented by a_i.

b) Transaction Encoder

Transaction encoder is to learn the relationship among alarms to generate a transaction representation, and give the higher weight to the alarms that are more similar to the root-cause alarm. We use the Transformer Encoder[6] model in this module, which is a fully parallel model

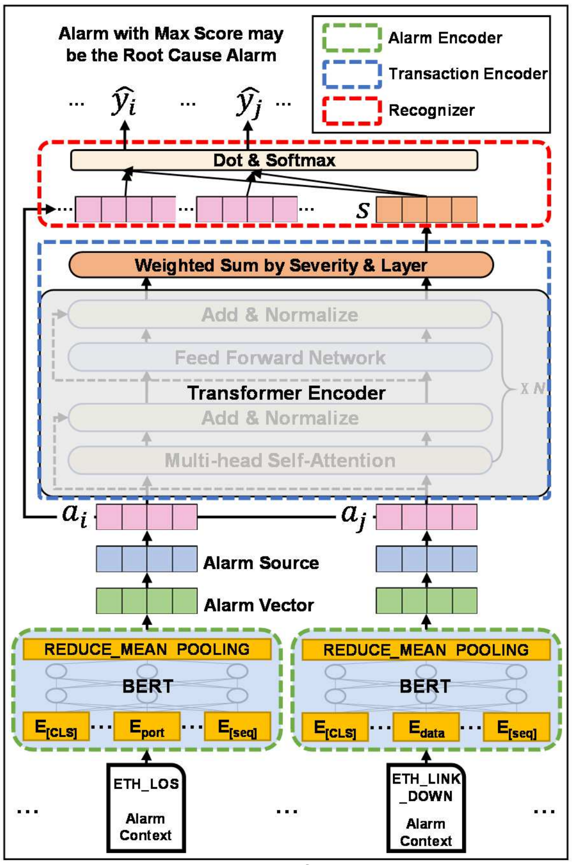

Fig. 2: Root cause alarm identification model based on the context-vectorization alarm representation.

structure, dispensing with recurrence and convolutions entirely. It solely relies on attention mechanism to draw the global dependence between input and output, so that the output could pay more attention to root alarm.

After extracting from Transformer Encoder, a new transaction representation $[a_1', a_2', \ldots, a_N']$ is obtained. And we use the weighted average to aggregate the alarm transaction. The weight here is the joint priority w_i of the alarm severity and the alarm layer. The higher the alarm severity and the lower the alarm layer, the higher the priority, and vice versa. Finally, the alarm transaction representation s is obtained:

$$w_i' = \frac{w_i}{\sum_N w_i}$$
$$s = \sum_N w_i' a_i' \tag{1}$$

c) Recognizer

Recognizer is used to identify which alarm in the alarm transaction is a root alarm. The key to this module is using a simple dot product to estimate alarm-transaction correlation score. The formula is shown in Eq. (2).

$$\hat{y}_i' = a_i^T s$$
$$\hat{y}_i = \frac{\exp(\hat{y}_i')}{\sum_N \exp(\hat{y}_i')} \tag{2}$$

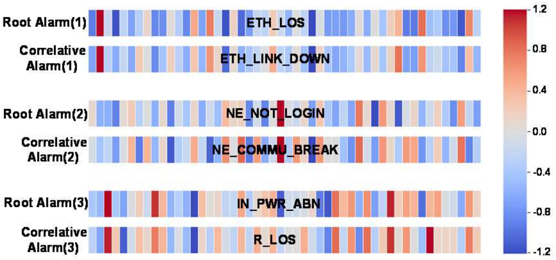

Fig. 3: Visualization (heatmap) of the 50 dimensions of the alarm vector of 3 groups.

In the end, an alarm with the highest correlation score is more likely to be a root alarm.

Experimental Results and Analysis

In order to verify the effectiveness of our proposed method, the alarm transaction data with the known root alarm labels were collected from the practical Packet Transport Network (PTN) provided by the operator. Total of 4950 and 1230 alarm transactions are used for training and testing, respectively. The maximum length of alarm transaction N is 15. The joint priority w_i of the alarm is composed of the alarm severity (Critical, Major, Minor, Warning) and the alarm layer (Physical, Data-Link, Tunnel, PW, Ethernet). The BERT pre-training model selects the BERT-Base with 768 dimensions released by Google[7].

First, six selected alarm representations based on context-vectorization are visualized in the form of heatmap, where different colours correspond to different vector values, as shown in Fig. 3. Here, the three main root alarms and three corresponding correlated alarms learned from expertise and manual are displayed. Each alarm vector is extracted by BERT, and the 50 dimensions of the 768-dimensional vector are displayed. It can be seen that every two vectors in the three groups (such as *ETH_LOS* and *ETH_LINK_DOWN*) look similar in the feature expression of the vectors. However, the vector expressions between different groups are distinctly different in many dimensions. In fact, by asking engineers and checking the alarm manual, we find out that the two alarms in the group do have a strong correlation, and it is also reflected in the context description of the alarm definition or impact. Meantime, we also learned from the manual that alarms such as *ETH_LOS* and *NE_NOT_LOGIN* is irrelevant in many ways. Correlative alarms do have similar feature expressions and irrelevant ones are not. The eigenvector representation generated by our method produces a similar awareness of the alarm to humans, demonstrating that we use context to represent alarms is effective, and the method of understanding context content and generating vectors through advanced language models can better represent the alarm.

During the training process the model can be

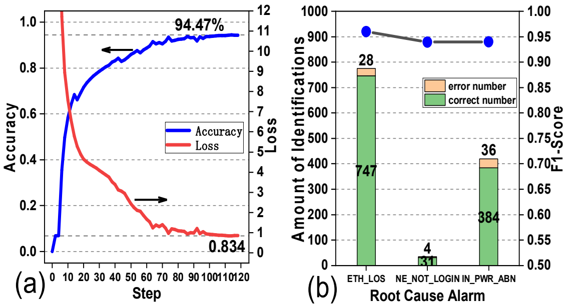

Fig. 4: (a) The performance of Identification model; (b) Amount of identifications and F1-score for each root alarm.

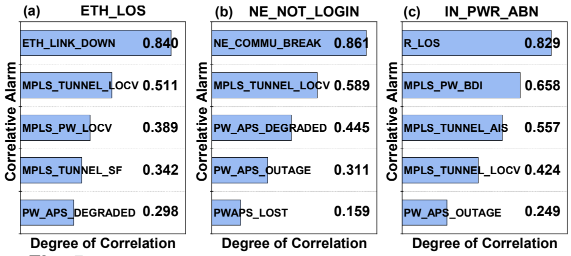

Fig. 5: The top five alarms correlated to (a) ETH_LOS; (b) NE_NOT_LOGIN; (c) IN_PWR_ABN.

converged after 120 steps. Then the trained model is used to identify 1230 alarm transactions for test. The identification accuracy is 94.47% and loss is 0.834, as shown in Fig. 4(a). Through statistical analysis, it is found that root alarms are mainly concentrated into three types: *ETH_LOS*, *NE_NOT_LOGIN*, *IN_PWR_ABN*, and the statistical results are summarized in Fig. 4(b). In the identification of each root cause alarm, our model can maintain an F1-score of ~0.95. Moreover, through calculating the correlation score from Eq. (2), we also figure out 5 derivative alarms with the quantified correlation degree for each root alarm, as shown in Fig. 5. It can be seen that the most relevant derivative alarms for the three root alarms are *ETH_LINK_DOWN*, *NE_NOT_LOGIN*, and *R_LOS*, which is consistent with the visualized vector results in Fig. 3, proving the effectiveness of the identification model using vectorized alarm representation in the task.

Conclusions

In this work, an NLP technology was introduced to vectorize the alarm context to represent the alarm. The obtained alarm representation is not only more accurate than the traditional symbolic representation, but also can replace the engineer to extract the expertise and read the alarm manual. Three common root alarms were identified with an accuracy of 94.47%, and other correlated alarms were also found out with quantified correlation degrees.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No. 61871415, 61975020), in part by Fund of State Key Laboratory of IPOC (BUPT) (No. IPOC2020ZT05), P. R. China, by the Key Laboratory Fund (No. 6142104190207).

References

[1] N. Amani *et al.*, "A case-based reasoning method for alarm filtering and correlation in telecommunication networks." In *Canadian Conference on Electrical and Computer Engineering, 2005.*, pp. 2182-2186. IEEE, 2005.

[2] D. S. Kim *et al.*, "An alarm correlation algorithm for network management based on root cause analysis." In *13th International Conference on Advanced Communication Technology (ICACT2011)*, pp. 1233-1238. IEEE, 2011.

[3] Z. Li *et al.*, "Demonstration of fault localization in optical networks based on knowledge graph and graph neural network." In *Proc. OFC 2020*, pp. Th1F-5. Optical Society of America, 2020.

[4] J. Devlin, *et al.*, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv*:1810.04805, 2018.

[5] Linyi Y. et al., "Html: Hierarchical transformer-based multi-task learning for volatility prediction." In Proceedings of The Web Conference 2020, pp. 441-451. 2020.

[6] A. Vaswani, et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, p. 6000–6010, 2017

[7] google-research, "bert," [Online]. Available: https://github.com/google-research/bert. [Accessed 2021]