Optical Network Diagnostics Using Graph Neural Networks and Natural Language Processing

Xiaonan Xu(1), Haoshuo Chen(2), Jesse E. Simsarian(2), Roland Ryf(2), Mikael Mazur(2), Lauren Dallachiesa(2), Nicolas K. Fontaine(2), David T. Neilson(2)

(1) Aberdeen, NJ 07747, USA
(2) Nokia Bell Labs, 600 Mountain Ave., Murray Hill, NJ 07974, USA
xuxiaonan2008@gmail.com

Abstract: We propose an AI-powered network diagnostic strategy including alarm clustering and fault localization with >98% accuracy for up-to 16-degree ROADMs and demonstrate the advantages of using NLP in encoding.

1. Introduction

Optical networks create the underlying interconnection for the global internet and enable 5G and future 6G applications. Network interconnection can be visualized by a topological graph showing the arrangement of components, equipment, and fiber connections. A graph neural network (GNN) [1] is one type of artificial neural network that can process the attributes of a graph, such as node or edge features, to update graph representations and extract essential information. GNNs have been widely applied in citation networks, social networks and chemistry for drug discovery [2]. Recently, GNNs have been adopted to tackle problems in optical networks such as fault localization [3] and traffic prediction [4]. Previous work on fault localization [3] can only predict a single failure, which may not be sufficient for practical applications.

In this paper, we propose an artificial intelligence (AI)-powered network diagnostic model that can deal with multiple failure sources through 1) linkage prediction for alarm clustering and 2) fault localization enabled by inductive representation learning on a network topology. Moreover, we introduce a natural language processing (NLP) technique for the generation of alarm feature vectors. In contrast to binary representations using one-hot encoding, we demonstrate that feature vectors created by NLP using the alarm descriptions can not only reduce computational complexity through dimensionality reduction, but also enable zero-shot learning for unobserved alarms during training.

2. Network diagnostics

Figure 1(a) presents a diagram of an optical network consisting of 10 reconfigurable optical add-drop multiplexer (ROADM) nodes within which seven nodes labelled as n_A, ... ,n_G report alarms and three service paths experience network outages. Fig. 1(b) illustrates a 3-degree ROADM composed of line amplifiers (LAs), wavelength selective switch (WSSes), array amplifiers (AAs), multicast switches (MCSes) and optical transponders

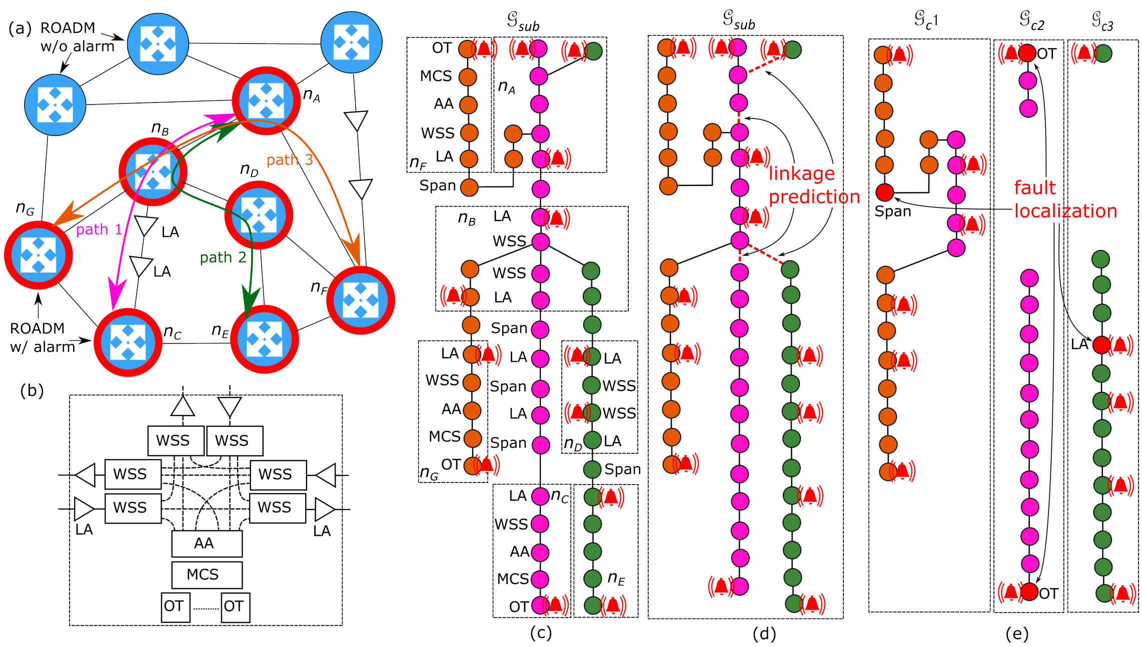

Fig. 1: (a) Diagram of an optical network consisting of 10 reconfigurable optical add-drop multiplexer (ROADM) nodes, (b) illusration of a 3-degree ROADM node, (c) network element card-level sub-graph $\mathcal{G}_{sub}$ of the three service paths with alarms, (d) linkage prediction and, (e) fault localization over three alarm clusters $\mathcal{G}_{c_1}$, $\mathcal{G}_{c_2}$ and $\mathcal{G}_{c_3}$.

(OTs). Service path 1 establishes a connection between n_A and n_C and has an A-Z topology within node n_A of OT→MCS→AA→WSS→LA and the reverse within node n_C as well as LA→WSS→WSS→LA at optical passthrough node (n_B).

To make the network diagnostic solution more flexible, instead of the global graph $\mathcal{G}(\mathcal{V},\mathcal{W},\mathcal{X})$, a sub-graph $\mathcal{G}_{sub}$ with node features such as node alarm embedding $\boldsymbol{x}_v^a$ and path index $\boldsymbol{x}_v^t$, $\forall v \in \mathcal{V}$ is processed, see Algorithm 1. A network element card-level $\mathcal{G}_{sub}$ is constructed as illustrated in Fig. 1(c) after selecting n_A and path 1 as the central node and common path, respectively, and adding interconnected paths 2 and 3 that share network elements with path 1. Note that each service path can contain multiple wavelengths. In order to cope with multiple failure sources, we divide the fault localization problem into two sub-problems: **1)** a linkage problem where a link prediction algorithm estimates whether all the interconnected paths with outages are affected by a common failure source and creates alarm clusters, and **2)** a node classification problem where a classification algorithm locates the node(s) with the main failure for each cluster. Fig. 1(c) and (d) illustrate the process of alarm linkage prediction and fault localization for each alarm cluster. Both linkage and localization models use GraphSAGE [5] which updates node representations by aggregating node features in the local neighborhood. In **Linkage**(), $\boldsymbol{x}_v^a$ is concatenated with $\boldsymbol{x}_v^t$ as a joint node representation, and edge vectors are updated by concatenating the representation of the start and end nodes of the edge. Each edge vector is independently fed through a fully-connected layer, a non-linear function and a second fully-connected layer to determine the linkage between the network elements and create M clustered alarm graphs $\mathcal{G}_{c_j}$, $\forall j \in 1,\ldots,M$. Fault localization **Localizer**() is applied for each $\mathcal{G}_{c_j}$ and only node alarm embedding $\boldsymbol{x}_v^a$ is used as input. Each node representation updated by GraphSAGE independently goes through two sets of fully-connected layers and a non-linear layer for classification. Nodes of $\mathcal{G}_{sub}$ are removed from $\mathcal{G}(\mathcal{V},\mathcal{W},\mathcal{X})$ before a new loop is started, and Algorithm 1 terminates when all of the alarms are processed.

Algorithm 1 Network diagnostics

Input: $\mathcal{G}(\mathcal{V},\mathcal{W},\mathcal{X})$: graph with nodes $\mathcal{V}$, edges $\mathcal{W}$ and features $\mathcal{X}$
D_h: hidden layer dimension, $\mathcal{K}$: aggregation depth
Process:
while # of alarms in $\mathcal{G}(\mathcal{V},\mathcal{W},\mathcal{X})) > 0$ **do**
central node ←select a ROADM with most alarmed OTs
common service path ← pick one alarmed OT associated path
$\mathcal{G}_{sub}$ ← add paths sharing equipment elements
$[\mathcal{G}_{c_1},\ldots,\mathcal{G}_{c_M}]$ ← **Linkage**($\mathcal{G}_{sub}$)
for j = 1 ... M **do**
FaultList ← **Localizer**($\mathcal{G}_{c_j}$)
end for
$\mathcal{G}(\mathcal{V},\mathcal{W},\mathcal{X}) \leftarrow \mathcal{G}(\mathcal{V},\mathcal{W},\mathcal{X})$ - $\mathcal{G}_{sub}$
end while=0
Output: FaultList

3. Implementation and Performance

We implement the proposed network diagnostic model on a network of Nokia 1830 Photonic Service Switch (PSS) nodes, which offers scalable and optimized end-to-end optical transport and switching. Different failure-associated alarm patterns are extracted from the logs of the network with 6 ROADM nodes and then generalized to arbitrary topology and ROADM degrees. We develop algorithms to generate various network topology graphs with alarms triggered by random failures out of 8 fault sources such as various network equipment card failures, broken fiber spans, and OT misconfigurations. To avoid ambiguity in alarm clustering, we apply a rule in the graph construction that each alarm will only be invoked by one failure. A dataset including 17,380,726 nodes and 7,941,340 edges with ROADM degrees from 2 to 16 is prepared and applied to train **Linkage**() with a split ratio of 60%, 20%, and 20% for training, evaluation, and test, respectively. M clustered alarm graphs, $\mathcal{G}_{c_j}$, of $\mathcal{G}_{sub}$ are applied to train the **Localizer**(). In total, 19 types of alarms are used, and four of them, including alarm "PUMPFAIL", are related to optical amplifier failures. To test zero-shot learning, only "PUMPFAIL" is observed during the training and the other three are unseen.

We use an NLP-based bag-of-words (BoW) method and term frequency-inverse document frequency (TF-IDF) to generate alarm embeddings, $\boldsymbol{x}^a$, from the alarm descriptions in the product manual based on the word-count occurrences and the importance of the words. Stop words such as "the", "and", and pronouns are excluded. Principal component analysis (PCA) is applied to further reduce the embedding dimension to 10. Fig. 2(a) shows the calculated cosine similarity between alarm "PUMPFAIL" associated with optical pump failure and the rest of the alarms. Higher similarity with the other optical amplifier related alarms indicates that the NLP encoding scheme can pick up the "meanings" of the alarms, unlike one-hot encoding which represents each alarm as a unique category, i.e., a binary vector with a dimension of 19. In contrast to one-hot encoding, whose embedding increases linearly with the number of alarm types, an NLP scheme can reduce the embedding dimension while maintaining the conceptional meaning of each alarm.

A batch size of 5 and aggregation depth $\mathcal{K}$ of 2 are used for training **Linkage**() in GraphSAGE. Fig. 2(b) shows the linkage accuracy versus the number of iterations for both NLP and one-hot encoding schemes and different D_h,

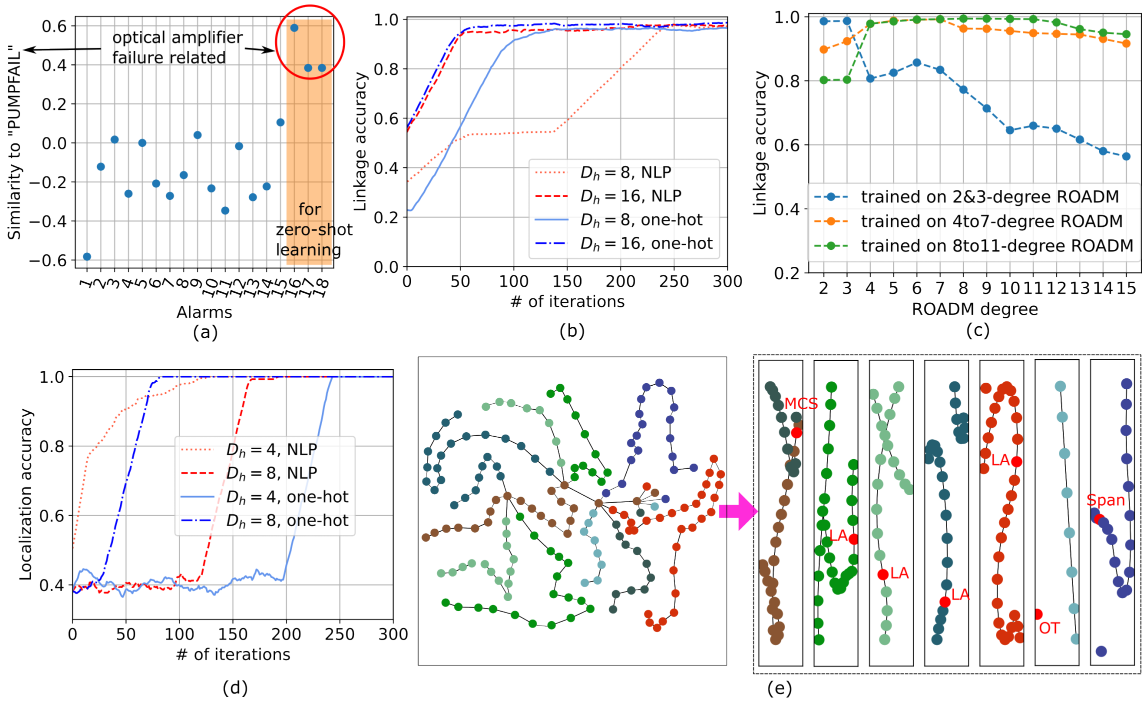

Fig. 2: (a) Calculated cosine similarity between alarm "PUMPFAIL" and the other 18 alarms, (b) alarm linkage accuracy versus the number of iterations in training **Linkage**() and (c) the linkage accuracy for models trained on different ROADM number of degrees, (d) fault localization accuracy versus number of iterations in training **Localizer**() and (e) example of an alarm graph including a 8-degree ROADM with 7 fault sources (left) and seven alarm clusters with an identified main fault (right).

which is the dimension of the hidden layer in GraphSAGE. Linkage accuracy is calculated by the ratio between the number of correctly predicted edges and the total number of edges. A larger D_h provides faster convergence. Both encoding schemes can achieve accuracy >98% after 300 iterations. However, the NLP scheme helps to reduce the model complexity due to its smaller embedding dimensionality. Fig. 2(b) shows the linkage accuracy as a function of ROADM degree for the models trained on different ROADM degrees. For degree >3, accuracy starts to degrade for the cluster model trained on 2&3-degree ROADMs. For the model trained on 4- to 7-degree ROADMs, >90% accuracy can be achieved from 2- up-to 16-degree ROADMs, indicating the flexibility of the cluster model.

For **Localizer**(), a batch size of 10 and aggregation depth $\mathcal{K}$ of 2 are chosen. Fig. 2(d) shows the localization accuracy versus the number of iterations. Localization accuracy is the ratio between the number of clustered alarm graphs with a correct failure estimation and the number of total clustered alarm graphs. 100% accuracy can be achieved for both encoding schemes, but the model using NLP again has lower complexity. A clear advantage of using NLP is observed in the zero-shot learning, where the three optical amplifier failure-related alarms that were not seen in the training are tested. The localizer model using NLP can still localize the faults correctly due to the high similarity between the alarms, however, the model with one-hot encoding completely fails, with zero correct estimations. An example of network diagnostics over $\mathcal{G}_{sub}$ including an 8-degree ROADM with seven independent failure sources is provided in Fig. 2(e) which shows seven alarm clusters after linkage prediction together with correctly identified main faults (red dots) after fault localization.

4. Conclusion

We proposed an AI-powered network diagnostic strategy including alarm clustering and fault localization. >98% clustering and 100% localization accuracy was achieved. We also demonstrated the advantages NLP in embedding generation including the reduction of model complexity and excellent performance in zero-shot learning.

References

1. T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.
2. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and S. Y. Philip, "A comprehensive survey on graph neural networks," IEEE transactions on neural networks and learning systems, 32(1), pp.4-24, 2020.
3. Z. Li, Y. Zhao, Y. Li, S. Rahman, F. Wang, X. Xin, and J. Zhang, "Fault localization based on knowledge graph in software-defined optical networks," Journal of Lightwave Technology, 39(13), pp.4236-4246, 2021.
4. B. Bao, H. Yang, Y. Wan, Q. Yao, A. Yu, J. Zhang, B. C. Chatterjee and E. Oki, "Node-oriented traffic prediction and scheduling based on graph convolutional network in metro optical networks," Proc. ECOC 2021, paper F2G-2.
5. W. Hamilton, Z. Ying and J. Leskovec, "Inductive representation learning on large graphs," Advances in neural information processing systems, 30, 2017.