IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 8, 2019, accepted May 28, 2019, date of publication June 6, 2019, date of current version June 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921329

Accurate Fault Location Using Deep Belief Network for Optical Fronthaul Networks in 5G and Beyond

AO YU[1], HUI YANG[1], QIUYAN YAO[1], YAJIE LI[1], HUIFENG GUO[2], TAO PENG[2], HAIBIN LI[2], AND JIE ZHANG[1]

[1]State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China

[2]ZTE, Shenzhen 518057, China

Corresponding author: Hui Yang (yang.hui.y@126.com)

This work was supported in part by the NSFC project (61871056), the Young Elite Scientists Sponsorship Program by CAST (YESS) under Grant 2018QNRC001, the Fundamental Research Funds for the Central Universities under Grant 2018XKJC06, the Open Fund of SKL of IPOC (BUPT) under Grant IPOC2018A001, and ZTE Research Fund.

ABSTRACT In face of staggering traffic growth driven by fifth generation (5G) and beyond, optical fronthaul networks which host such connections require efficient and reliable operational environments. Fault location has become one of the primary factors for post-fault responses. In this paper, we propose a Deep Belief Network (DBN) based fault location (DBN-FL) model to locate single-link fault of optical fronthaul network in 5G and beyond. The DBN-FL model contains two phases including the hybrid pre-training phase and the Levenberg Marquardt (LM) algorithm-based fine-tuning phase. In the hybrid pre-training phase, we combine the supervised and unsupervised learning to reduce the demand for training samples. In the fine-tuning phase, the LM algorithm is adopted to fine-tune the DBN-FL model. The experimental results indicate that the proposed DBN-FL model can realize high-accuracy fault location as a classifier (accuracy over 96%), and outperforms traditional deep learning (DL) approaches both in terms of location accuracy and training efficiency.

INDEX TERMS 5G and beyond, fronthaul, fault location, deep belief network.

I. INTRODUCTION

The rapid growth of ubiquitous services and ultra-high bitrate internet applications in fifth generation (5G) and beyond, such as internet of vehicles (IoV) and augmented reality (AR), has enforced the evolution of optical fronthaul networks to develop towards higher service capacity [1]. Today's fronthaul network solutions including wavelength-division-multiplexed passive optical network (WDM-PON) and active WDM/optical transport network (OTN), can offer extremely high bit rates and bandwidth through ultra-high dense data transmission [2], [3]. However, when a fault occurs in such intensive service access scenarios, hundreds of services may be interrupted and volumes of traffic data (Petabyte level) are lost. Network operators are currently focusing not only on how to restore the interrupted links as soon as possible but also on how to locate the fault accurately. With the increase of users' demand for efficient and stable networks in 5G and beyond, it is much more important for the operators to locate faults accurately [4].

The fault location detection in optical fronthaul networks is based on the alarm information collected from the optical performance monitoring equipment. A considerable number of alarms are generated after the fault occurs, most of which are false alarms. For instance, as depicted in Fig.1, a link fault occurs in the optical fronthaul network, causing multiple false alarms to unexpectedly arrive at the monitoring system. In the case of optical fronthaul networks in 5G and beyond, more monitoring nodes will be deployed to ensure the smooth operation of the network. As a result, the monitors will generate even more false alarms once a fault occurs in the system, and locating faults accurately from a large number of false alarms becomes a more challenging task.

In the last few decades, a growing body of intelligent approaches have been proposed for optical network fault location. Machine learning (ML) techniques, such as neural

The associate editor coordinating the review of this manuscript and approving it for publication was Martin Reisslein.

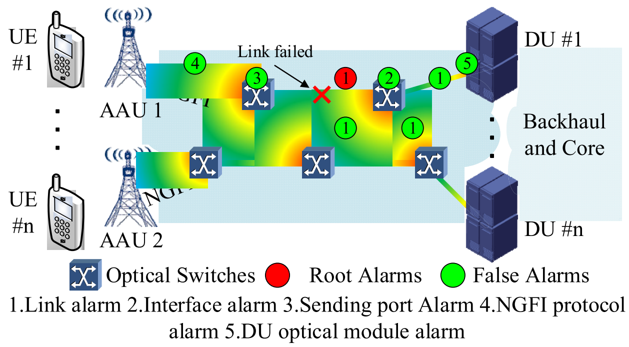


FIGURE 1. The procedure of alarm generation after the link fault occurs. UE, user equipment; NFGI, Next generation fronthaul interface; AAU, active antenna unit; DU, distribute unit.

networks (NN) and support vector machine (SVM), have been successfully applied to optical network fault location [5], [6]. The outstanding advantages of these data-driven algorithms are the high computation speed and satisfactory fault classification ability. Since the number of false alarm continues growing in 5G and beyond, there are limitations in data processing efficiency and locating accuracy of the existing ML algorithms.

In recent years, deep belief network (DBN) model, as an important branch of deep learning (DL), has gained in popularity as a successful implementation of an efficient learning technique that stacks simpler models known as restricted Boltzmann machines (RBMs) [7]. The training process can be divided into two phases in terms of the pre-training phase and the fine-tuning phase. In the pre-training phase, DBN model is pre-trained from bottom-to-top in an unsupervised way to obtain the initial parameters between adjacent layers. Through this process, the DBN largely avoids the vanishing gradient problem that may occur in the case of no pre-initialization of traditional deep neural network. The pre-training phase also improves model performance by avoiding overfitting and enhancing the model generalization. This phase is critical for accurate fault location given the alarm data with massive false alarm of optical fronthaul networks in 5G and beyond. In the fine-tuning phase, the parameters are fine-tuned from top-to-bottom to optimize parameters of the whole DBN in a supervised back propagation (BP) algorithm, producing a multilayer model able to perform the desired task (e.g., alarm classification). With such network architecture and training method, feature selection and extraction of alarm information can be automatically carried out without extra special data refinement. Therefore, the aforementioned two-step training process of DBN showed the high potential of locating single-link fault in optical fronthaul networks.

However, there are still some works to be done before applying DBN to solve the problem of fault location. The optical fronthaul networks are typically managed via conservative design approaches, which makes the probability of failures negligible. In other words, the number of training samples is insufficient. Thus, the ability of feature extraction cannot be fully utilized in the unsupervised pre-training phase, which requires a considerable volumes of training data. Moreover, the DBN model has an inefficient convergence rate during the fine-tuning phase, because the BP algorithm cannot quickly optimize to the complex nonlinear objective function when classifying the alarm information [8]. To the best of our knowledge, DBN has not been well studied in fault location of optical networks.

In this paper, we study the single-link fault location problem, where the faults are independent of each other in optical fronthaul networks. We propose a DBN based fault location (DBN-FL) model to explore and extract latent features from the alarm information and identify false alarms from the generated vast alarms in the network management system. We first train the DBN-FL combined the supervised with unsupervised learning to extract features from alarm information data. Then deep features are generated to analyze the differences in the latent features between the real alarm and false alarm. In addition, we adopt the Levenberg Marquardt (LM) algorithm to accelerate the training of fine-tuning phase. Our results show that the proposed DBN-FL model acquires over 96% accuracy of fault location. Additionally, compared with the advanced deep-learning based fault location methods, the training time is greatly reduced, which means that the DBN-FL model can be used in large-scale and complex networks. The contributions of this paper are summarized as follows.

- First, we consider the single-link fault location in optical fronthaul networks as a classification issue of real alarm and false alarm. Accurate identification of false alarm is an effective approach to improve the accuracy of fault location.
- We then propose a DBN-FL model to characterize difference in real alarm and false alarm for single-link fault location. In the hybrid pre-training phase, we transfer the learned parameters of supervised learning-based RBM to the unsupervised learning-based model as the initial parameters. In the fine-tuning phase, the LM algorithm is adopted to replace the BP algorithm to fine-tune the DBN-FL model. These two approaches can effectively reduce the demand for training samples and improve the training efficiency.
- Furthermore, we perform fundamental analysis on the factors that may affect the performance of DBN-FL model, training efficiency of the DBN-FL as well as the location accuracy of DBN-FL.

The remainder of this paper is organized as follows. Section II reviews the relevant works. In Section III, we discuss the motivation of accurate fault location, the basics of alarm information, types of alarm information, and the data preparation for further processing. Section IV describes the construction, training and implementation of DBN-FL model. Experiments are conducted in Section V to verify the effectiveness and limitations of the proposed DBN-FL model. Finally, Section VI concludes this paper and points out the future works.

II. RELATED WORKS

In this section, we first review the traditional solutions proposed so far for the fault location in optical networks. This is followed by a discussion of the artificial intelligence based fault location solutions.

A. TRADITIONAL SOLUTIONS

Traditional fault location methods are classified into active location and passive location. The principle of active location is to set up the monitor at the destination node in the network. Wu *et al.* [9] presented the monitoring-cycles to detect fault in optical network. This method can set up a monitor to report the status information of links on the monitoring ring upward, which is the most direct fault location method by detecting the state of the equipment. However, this approach wastes a lot of time of operators to distinguish the false alarm and real alarm. Authors in [10] also proposed monitoring trail technology to improve localization efficiency. The passive location technology is to locate faults according to the state of the working path in the network. This method can improve the efficiency of fault location, but the accuracy cannot meet the requirements of future fronthaul optical networks. Khair *et al.* [11] considered Limited-perimeter Vector Matching (LVM) protocol to locate multi-link faults. In order to deal with multi-link faults, they separated the faults in each region, and then located the faults in parallel. This technology can locate multi-link faults by locating faults simultaneously. However, the accuracy of fault location cannot be guaranteed, because this algorithm did not fully consider the content of alarm information including the occurrence and the type of alarms. The authors in [12] and [13] proposed a fault location algorithm based on fuzzy mathematics. By calculating the fuzzy membership degree of each link and setting a certain threshold, the set of suspected fault links in the network can be located. These methods can improve the accuracy of fault location by abstracting and quantifying the fault location process. However, the fault cannot be located in time due to the high computational complexity. Carmen *et al.* [14] presented a failure location algorithm to locate both single and multiple failures in transparent optical networks. However, the computational complexity of this approach become unacceptable in a complex network environment such as the fronthaul optical network, because the equipment is more closely linked and more factors need to be considered.

These works focus on the fault location by monitoring fault information or analyzing alarm information. With the development of optical networks, it is more difficult to analyze the fault location efficiently by using traditional algorithms. Artificial intelligence technologies are essential to the fronthaul networks in 5G and beyond.

B. ARTIFICIAL INTELLIGENCE BASED SOLUTIONS

In recent years, the importance of optical network reliability has been widely recognized, and a lot of artificial intelligence based solutions have been proposed. Gosselin *et al.* [15] presented a probabilistic graphical model based on machine learning algorithm for fault location in optical access networks. This method used a Bayesian network to encode alarm information and developed an inference engine called probabilistic tool to efficiently detect fault in passive optical networks. Moreover, Ruiz *et al.* [16] also modeled the fault location problem of optical networks through a Bayesian network. A fault location method was proposed to improve the quality of services in the virtual network topology. However, the Bayesian network has lower accuracy in identifying new faults that occur in the network compared with other artificial intelligence approaches. Zhang *et al.* [17] proposed a cognitive mechanism for software defined optical networks (SDON). This work aimed to detect faults in centralized networks by exchanging the messages periodically between controllers and switches. This method has a high accuracy of fault location, but leads to a longer locating time due to the extra analysis of the information interaction process. Rafique *et al.* [18] presented a cognitive assurance architecture for optical network fault management, which can make recommendations after fault occurs by analyzing monitored data through pre-trained machine learning model. The locating time of this approach is satisfactory, but the algorithm cannot guarantee the accuracy when the training data set is small.

III. BASICS OF FAULT LOCATION

This section briefly introduces the fault location requirements of optical fronthaul networks in 5G, followed by the basics of alarm information, types of alarm information and the data preparation for fault location.

A. MOTIVATION

In the future 5G era, there will be dozens, hundreds or more optical terminals inside the network [19]–[23]. With the dynamic change of the location and relationship between people and devices, massive high-speed data exchange takes place between these terminals all the time. If the corresponding network resource regulation is done manually, the fault and repair time of network equipment is calculated based on the human response time. Considering the huge amount of state information brought by network equipment, it has become impossible for human beings to find out the critical fault alarm, locate the fault cause, propose solutions, and finally manually complete the resource scheduling to solve the fault. Facing the huge and complicated alarm space, the operator can only be liberated from the original and inefficient manual configuration management process through artificial intelligence based fault location technology.

The effectiveness of fault location is another important requirement. With the expansion of the network scale and the diversity of network resources such as computing, storage, routing etc., it is inevitable that the connection between devices becomes closer. In this case, once the fault occurs, the affected scope of network would be wider. In addition, the efficiency of fault location model is required in optical networks. As the main carrier of data, optical networks faults will affect more users and causes more serious economic losses. Therefore, how to quickly locate the fault from the alarm information and narrow the scope of inspection of operators has become the key to improving the efficiency of operation and maintenance.

B. BASICS OF ALARM INFORMATION

In this subsection, we introduce the typical composition of alarm information. The fault in optical fronthaul networks contains two main types, which are link fault and device fault. Since the number of device fault and multi-link fault only accounts for a very small proportion compared with link fault in actual operating process, only the single-link fault is considered in this paper. The goal of fault location algorithm is to locate the faulty link by classifying alarm information which is collected by the control plane through network management system.

TABLE 1. Typical alarm information.

Name	Description	Examples
Alarm_id	Serial number of alarm information	7651
Dn	Identifying name of a network device	Flexi EDGE BTS
Address	Port name	TM2.5G-K1-1-1-16-N2PQ1-33
Occur_time	Occur time of the alarm	11 Sep 2018 15:03:33.241
Clear_time	Clear time of the alarm	11 Sep 2018 17:28:14.391
Org_type	Type of the alarm	3080
Org_severit	Level of the alarm	Emergency alarm
Alarm_text	Content of the alarm	100 1 100 A898FSMF 1 16 path=N/AadditionalFaultId:3080
Title	Title of the alarm	BASE STATION OPERATION
Activestatus	Active state of the network device	BTS reference clock missing

Table 1 gives an example of the typical alarm information. The raw alarm information consists of network anomalies information pre-defined on the equipment, such as the alarm name, the port name, the name network elements, equipment manufacturers, the name of the circuit, the start time and end time, the types of alarms, and the alarm level, etc.

C. CONSIDERED TYPES OF ALARM INFORMATION

The possible types of alarm information in optical networks considered in this paper can be divided into three categories according to the location of the monitors.

- Interface Alarms. The interface alarms are usually caused by too high or too low optical power received by devices. One example is that the received optical power is dropped when the optical fiber is cut or bent. In this article, the interface alarms mainly include the distributed unit (DU) next generation fronthaul interface (NGFI) alarm, the DU IR interface alarm, the S1/X2 interface alarm, and the radio frequency (RF) unit NGFI alarm.
- Link Alarms. This alarm is the result of services interruption. For example, the service interruption carried by the RF unit will result in the RF unit maintenance link alarm. The link alarms include the user surface link alarm, stream control transmission protocol (SCTP) link alarm, RF unit maintenance link alarm, and antenna maintenance link alarm.
- Module Alarms. The abnormal sending or receiving optical power of optical module might cause module alarm. For instance, the transmission power of the optical module is lower after an upstream link is interrupted, which leads to the DU optical module alarm. In general, the module alarms include the DU IR optical module alarm, the DU optical module alarm, and the dispersion compensation unit (DCU) optical module alarm.

These three alarms cover most of the alarms that may occur in fronthaul optical networks and are important samples that will be used in the feature selection and extraction of the DBN model.

D. DATA PREPARATION

The data-intensive environment of optical fronthaul networks in 5G and beyond would generate a larger volume of alarms when a fault or network problem occurs. These alarms contain lots of detailed but very fragmented, or even useless information about the faults. To improve the efficiency of DL, it is essential to transform the raw alarm information to be well-structured and high quality samples that are suitable for additional processing. The transformation process contains three parts including data reduction and data cleaning.

1) DATA REDUCTION

The raw alarm information obtained from the network management system contains lots of irrelevant and repetitive data. The real essential information includes the alarm name, alarm position, device port information, device manufacturer, alarm reason, and the alarm start time. By analyzing the data in Table 1, we can obtain the alarm device manufacturer, specific physical device, and port location from the indicator called *Address* which is used to represent the port-related information of the device. The correlation and sequence mode of alarm information can be obtained in *Occur_time* which is used to represent the occurrence time of an alarm. The severity of alarm information can be obtained in *Org_severit* which represents the alarm level. We can preliminarily determine the alarm correlation and possible cause of the faults by analyzing *Alarm_text* which is the main content of the alarm. Thus, the raw alarm information can be simplified to a data set containing *Address*, *Occur_time*, *Org_severit*, and *Alarm_text*.

2) DATA CLEANING

Data loss or inconsistence always exists since the raw alarm information is collected from the network management system automatically, which may lead to false interpretations afterwards. Therefore, we modify the samples with missing content and remove the noise data manually to generate complete and consistent samples for fault location.

3) DATA UNIFIED

It is necessary to abstract various data into tensor representation before inputting data into the DL model for further processing. We need to unify different types of data and present them in a standard way. For instance, the English word for *September* can be represented by a number 9. In addition, some long characters can be omitted if they appear in each alarm and are only explanatory, such as $path = N/AadditionalFaultId$.

IV. DBN-FL MODEL CONSTRUCTION AND IMPLEMENTATION

In this section, we give the details of the proposed DBN-FL model. The basic idea of hybrid pre-training phase is to transfer the learned parameters of one supervised learning-based RBM to another unsupervised learning-based RBM as the initial parameters. The two training processes in the hybrid pre-training phase are independent. After the pre-training phase ends, the LM algorithm is adopted to replace the BP algorithm to fine-tune the DBN-FL model.

A. MODEL CONSTRUCTION

The DBN-FL model is constructed by multiple restricted Boltzmann machines (RBMs), one Softmax layer (output layer), and one temporary output layer. Each RBM consists of a visible layer (input layer) and a hidden layer, and the output of the previous RBM serves as the input of the next RBM. Since the RBM cannot be used as a classification model alone, we add a Softmax layer at the top of DBN to classify the fault links. The features extracted by hidden layer are taken as the input of the Softmax layer and the result of alarm classification is obtained through activation function processing [24]. Moreover, one temporary output layer (only for supervised learning) is stacked on top of the first RBM of supervised learning, and its network parameters serve as the initial parameters for the first RBM' (RBM'_1) of unsupervised learning. The existence of the temporary output layer enables to conduct supervised learning algorithm.

The training samples of DBN-FL model are divided into labeled sample and unlabeled sample. Given L and U as sets of labeled sample and unlabeled sample respectively, x and y represent the input vector and the output vector respectively. The input vector contains various types of alarm information, including alarm name, device port information, and the alarm start time while the output vector only contains the alarm types including the real alarm and the false alarm. Then, the labeled sample for supervised learning can be denoted as $L(x, y)$ and the unlabeled sample for unsupervised learning can be denoted as $U(x)$. The number of $L(x, y)$ is far less than $U(x)$ because it is hard to obtain labeled samples in practical optical fronthaul networks. In this paper, all the training samples are divided into Mini-Batch data sets.

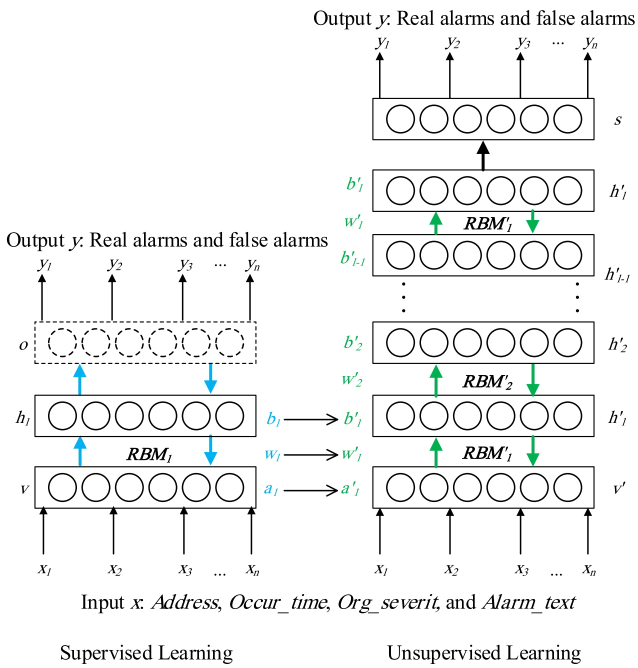


FIGURE 2. Structure of the pre-training phase.

B. HYBRID PRE-TRAINING

Fig. 2 depicts the hybrid pre-training phase of DBN-FL model. We use w_l to represent the weights matrix. Let a_l and b_l represent the bias vector of visible layer and hidden layer respectively. Moreover, we use w_1 and b_1 to represent the weights matrix and bias vector of RBM_1 obtained by supervised learning respectively, and o represents the temporary output layer. Firstly, we use the stochastic gradient descent (SGD) algorithm to train the RBM_1 with labeled sample $L(x, y)$, and obtain the weights matrix w_1 and bias vector b_1 of RBM_1 respectively. Secondly, we train the entire DBN-FL model by contrastive divergence (CD) algorithm with unlabeled sample $U(x)$, and use w_1 and b_1 to initialize the RBM'_1. In addition, s represents the Softmax layer. After training, the weights matrix and bias vector of each RBM' are obtained. Finally, the pre-training process of DBN-FL model is completed.

The training method of DBN-FL model are described in detail below.

We first train the RBM_1 with the SGD algorithm. The Mini-batch learning method is used to divide the training set into several subsets, and train one subset in each iteration.

After all subsets are trained in several times according to Eq. (1), we can obtain the parameters $\theta_1\{w_1, a_1, b_1\}$ of RBM_1.

$$\theta_1 \leftarrow \theta_1 - \alpha \frac{\partial J(\theta_1)}{\partial \theta_1} \tag{1}$$

In which, α is the learning rate of SGD algorithm and $J(\theta)$ is the least squares error function of training set. It is easy to verify that the computation complexity per iteration is $O(mn)$, where m means the number of hidden layer nodes and n indicates the number of visible layer nodes.

Then, a new energy function of DBN-FL model is given by Eq. (2) where $\theta = \{w' = (w'_{ij})_{n\times m}, a' = (a'_i)_n, b' = (b'_j)_m\}$ in which n and m are the number of visible unit v'_i and hidden unit h'_j.

$$\begin{aligned} E(v,h,\theta) = &-w_{11}v_1h_1 - a_1v_1 - b_1h_1 \\ &- \sum_{i=2}\sum_{j=2} w'_{ij}v'_ih'_j - \sum_{i=2} a'_iv'_i - \sum_{j=2} b'_jh'_i \end{aligned} \tag{2}$$

The joint probability distribution of visible and hidden elements can be expressed as Eq. (3). In Eq. (3a), $\theta' = \{w' = (w'_{11})_{n\times m}, a' = (a'_1)_n, b' = (b'_1)_m\}$, which means that the parameters of RBM'$_1$ have been trained by unsupervised learning. Eq. (3b) is defined as the probability distribution of the other RBM's.

$$P(v'_1, h'_1|\theta_1) = \frac{e^{-E(v'_1,h'_1,\theta_1)}}{\sum_{v_1,h_1} e^{-E(v'_1,h'_1,\theta_1)}} \tag{3a}$$

$$P(v', h'|\theta') = \frac{e^{-E(v',h',\theta')}}{\sum_{v,h} e^{-E(v',h',\theta')}} \tag{3b}$$

The binary state probability of the hidden unit $h'_j = 1$ can be calculated as Eq. (4) when a random visible vector v' is given. In Eq. (4), σ is the sigmoid activation function, which is defined as $\sigma(x) = \sigma(x) = \left(1 + e^{-x}\right)^{-1}$. Eq. (4a) is the binary state probability of the hidden unit $h'_j = 1$ in RBM'$_1$, while Eq. (4b) is the binary state probability of the hidden unit $h'_j = 1$ in other RBM's except RBM'$_1$.

$$P(h'_1 = 1|v') = \sigma(w_{11}v'_1 + b_1) \tag{4a}$$

$$P(h'_j = 1|v') = \sigma(\sum_i w'_{ij}v'_i + b'_j) \tag{4b}$$

Similarly, the binary state probability of the visible unit $v'_i = 1$ can be expressed as Eq. (5), when a random hidden vector h is given. Eq. (5a) is the binary state probability of the hidden unit $v'_i = 1$ in RBM'$_1$, while Eq. (5b) is the binary state probability of the visible unit $v'_i = 1$ in other RBMs except RBM'$_1$.

$$P(v'_1 = 1|h') = \sigma(w_{11}h'_1 + a_1) \tag{5a}$$

$$P(v'_i = 1|h') = \sigma(\sum_j w'_{ij}h'_j + a'_i) \tag{5b}$$

By solving the negative gradient of the logarithmic likelihood function, the optimal θ can be obtained by using the CD algorithm which can be expressed as Eq. (6). In Eq. (5), K represents the number of iterations and η is the learning rate of CD algorithm. Eq. (6a) is the parameter update process of RBM'$_1$, while Eq. (6b) applies for other RBM's.

$$\begin{aligned} w'_{11} &\leftarrow w_{11} + \eta[P(h'_{1,0} = 1|v')v'_{1,0} - P(h'_{1,1} = 1|v')v'_{1,1}] \\ a'_1 &\leftarrow a_1 + \eta(v'_{1,0} - v'_{1,1}) \\ b'_1 &\leftarrow b_1 + \eta[P(h'_{1,0} = 1|v') - P(h'_{1,1} = 1|v')] \end{aligned} \tag{6a}$$

$$\begin{aligned} w'_{ij} &\leftarrow w'_{ij} + \eta[P(h'_{j,K-1} = 1|v')v'_{i,K-1} - P(h'_{j,K} = 1|v')v'_{i,K} \\ a'_i &\leftarrow a'_i + \eta(v'_{i,K-1} - v'_{i,K}) \\ b'_j &\leftarrow b'_j + \eta[P(h'_{j,K-1} = 1|v') - P(h'_{j,K} = 1|v')] \end{aligned} \tag{6b}$$

We can estimate that the time complexity of CD algorithm is $O(Nmn)$, where N denotes the number of sample particles. The pseudocode of hybrid pre-training phase is shown in Fig. 3.

Algorithm: Hybrid Pre-training Phase

```
Input:  x; //the input vector of the training sample
        y; //the output vector of the training sample
        V; //the weight matrix of temporary output layer
        εC; //learning rate of un-supervised learning
Output: w; //the weight matrix of RBM
        a; //the bias vector of visible layer
        b; //the bias vector of hidden layer
        e; //prediction error
Initialize η = {w, a, b, V};
  For i = 1 do
    vi = xi; //input  x  to RBM
    f(xi) = (w · xi + b) · V; //output of temporary output layer
    e = f(xi) - y;
  end for
  For j = 2 → m do
    For k = 2 → r do //r, the number of visible units of RBM'
      For g = 2 → t do //t, the number of hidden units of RBM'
        hj ← sigmoid(Σ (wj+w1) · vj + ((bj+b1)); //hidden units
          vk' ← sigmoid(Σ wk · hj + ak); //visible units
          hg' ← sigmoid(Σ wg · vk' + bg); //hidden units
      end for
    end for
  end for
  if (l > 1) Then // network layer l
    η ← η - εC · ∂e/∂η; //update parameterS
    breaks;
  end if
```

FIGURE 3. The Pseudocode of hybrid pre-training phase.

After the training of supervised learning, the parameters $\theta_1\{w_1, a_1, b_1\}$ of RBM_1 can be obtained according to Eq. (1). Because the labeled training samples used in the supervised learning have basic characteristics of the whole training data set, the learned parameters are useful to the following unsupervised learning. Then we use unsupervised learning algorithm to train the RBM'$_1$ with the initial parameters $\theta_1\{w_1, a_1, b_1\}$, and the parameters after training are expressed as $\theta'_1\{w'_1, a'_1, b'_1\}$ according to Eq. (6a). Next, the activation probability of the hidden unit of RBM'$_1$ is taken as the input vector of RBM'$_2$. According to Eq. (6b), the parameters $\theta'_2\{w'_2, a'_2, b'_2\}$ are obtained. The activation probability of the hidden unit of RBM'$_2$ is taken as the input vector of RBM'$_3$

according to Eq. (6b). The rest RBM' can be trained in the same manner until the top layer ends. Finally, we finish the pre-training phase and obtain the parameters $\theta'\{w', a', b'\}$ of DBN-FL model.

Note that, when the training samples for supervised learning are too large, the alarm classification accuracy is satisfactory, but the training time is inefficient. The best proportion of labeled samples in supervised learning is 20%, which is determined from the fault location simulations by trial-and-error method. We try to maintain the advantage of unsupervised learning in feature extraction on the basis of initial parameters provided by supervised learning. Moreover, benefited from the learned parameters, the unsupervised learning can greatly reduce the calculation cost of initialization for parameters and be quickly trained.

C. LM-BASED FINE-TUNING

After the hybrid pre-training phase, we use the Levenberg Marquardt (LM) algorithm to fine-tune the parameters $\theta'\{w', a', b'\}$ of DBN-FL model. For the given training sample set $L\{x, y\}$, the relationship between input and output can be represented by Eq. (7) where f is a non-linear function and K is the number of iterations.

$$\hat{y}_K = f_K(x, \theta) \tag{7}$$

Unlike the regression problem, such as traffic prediction and price analysis, the fault location is a classification problem. In other words, the accuracy of fault location is a discrete binary function and the optimal solution cannot be obtained by derivation. To solve this problem, we use the Least Squares Polynomial Curve Fitting algorithm to fit the original discrete error function into a new continuous error function. Thus, we design a new accuracy function depicted in Eq. (8) where g is a continuous differentiable polynomial function designed as an approximate curve of discrete accuracy data set of fault location.

$$e_K = g_K(y_K - f_K(x, \theta)) \tag{8}$$

Thus, the error loss function of fault location is defined as the error sum of squares as shown in Eq. (9), where y_K is the real location of the fault in Kth iteration.

$$E(\psi) = \sum_{K=1} e_K^2 = \sum_{K=1} g_K^2(y_K - f_K(x, \theta)) \tag{9}$$

In this paper, the LM algorithm is used to maximize the location accuracy rate based on the parameters $\theta\{w, a, b\}$, consequently improving the performance of DBN-FL model. Then, the updated parameters $\psi_K\{w, a, b\}$ can be obtained from Eq. (10).

$$\Delta\psi_k = -[J^T(\psi_k)J(\psi_k) + \mu I]^{-1} J^T(\psi_k)E(\psi_k) \tag{10}$$

In Eq. (10), the damping factor μ is a positive constant, I is the Identity matrix, and $J(\psi_k)$ is the Jacobi matrix given by Eq. (11).

$$J(\psi_k) = \begin{bmatrix} \frac{\partial e_1(\psi_k)}{\partial \psi_1} & \frac{\partial e_1(\psi_k)}{\partial \psi_2} & \cdots & \frac{\partial e_1(\psi_k)}{\partial \psi_p} \\ \frac{\partial e_2(d_k)}{\partial \psi_1} & \frac{\partial e_1(\psi_k)}{\partial \psi_2} & \cdots & \frac{\partial e_2(\psi_k)}{\partial \psi_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N(\psi_k)}{\partial \psi_1} & \frac{\partial e_N(\psi_k)}{\partial \psi_2} & \cdots & \frac{\partial e_N(\psi_k)}{\partial \psi_p} \end{bmatrix} \tag{11}$$

The goal of each iteration of the LM algorithm is to find a suitable damping factor μ. When $\mu = 0$, the LM algorithm is transformed into the Gauss-Newton method with the least square solution. When $\mu > 0$, the LM algorithm is converted to the Steepest Gradient Descent method. Therefore, the LM algorithm has the global convergence characteristics for fault location with initial drop and fast iteration based on the advantage of the above two methods. The time complexity of LM algorithm is $O(e^{-2})$ in this paper, because the LM algorithm is quadratic convergence.

It should be mentioned that when tuning the parameters of LM algorithm, the effective setting of damping factor can decrease the calculation cost. In this work, we first use a relatively small value of damping factor μ to train the LM algorithm. When the training error starts to increase, we increase the value of μ and use Steepest Gradient Descent method to quickly search the minimum of the loss function. Then, we decrease the value of μ and use Gauss-Newton method to search the minimum of the loss function until the LM algorithm converges to the global optimal solution. Here, the LM algorithm-based fine-tuning phase is proposed to fine-tune the whole DBN-FL model, which overcomes the defects of the BP algorithm including computation cost and gradient diffusion for deep structure as well as the low accuracy.

D. IMPLEMENTATION OF DBN-FL MODEL

DBN-FL model can be used for optical fronthaul networks to locate single-link faults in 5G and beyond. DBN-FL model contains two phases as the hybrid pre-training phase and LM based fine-tuning phase.

In the hybrid pre-training phase, we first use supervised learning to train the bottom RBM by adding a temporary output layer, and then use unsupervised learning to train the whole DBN-FL model based on the initial parameters obtained from supervised learning. In the fine-tuning phase, the LM algorithm is adopted to fine-tune the parameters obtained from the pre-training stage. As shown in Fig. 4, the main steps of the DBN-FL model are as follows.

Step 1: Collect the training samples set and pre-process these data for later processing.

Step 2: Use the hybrid pre-training method to obtain the parameters of DBN-FL model. The initial parameters of RBM_1 are obtained through SGD algorithm. Then, we give the parameters of RBM_1 as the initial parameters of RBM'_1 and use CD algorithm to train the whole DBN-FL model.

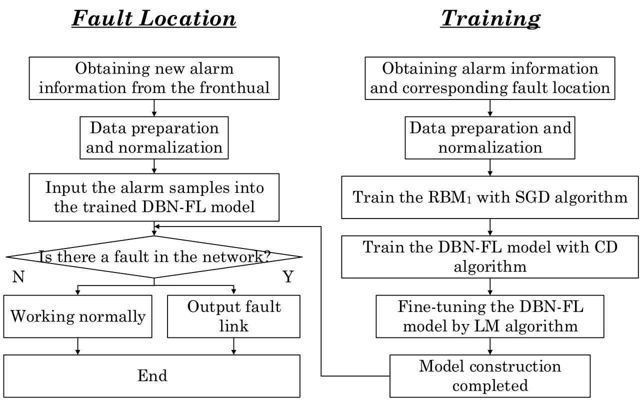


FIGURE 4. Full process of DBN-FL model in fault location.

Step 3: Use LM algorithm to optimize the parameters of DBN-FL model. Supervised learning based training is adopted to fine-tune the parameters of DBN-FL model and obtain the optimal parameters.

Step 4: After the training process, input the new alarm information to the DBN-FL model. Then, we can obtain the location of the fault.

V. EXPERIMENTAL SETUP AND RESULTS

This section gives extensive experiments to evaluate the single-link fault location performance of the proposed DBN-FL model, which assumes that all the fault datasets and alarm information are available during the training process. In the following, the experimental preparation will be explained first, and then the fault location results and analysis will be shown. At last, we briefly introduce the limitations of our work.

A. TRAINING DATA SET

To verify the effectiveness of the proposed algorithms, we collect the fault data and the corresponding alarm information from the management system of network operator. The received fault types include device fault and link fault. Since the number of device fault is far less than the link fault (less than 5%), we screen out all the device fault and focus on the link fault in our experiment. During the 120 days collection period, 1148 faults occur in 123,249 links and a total of 341,139 alarm information are generated.

In addition, we generate another dataset contains synthetic alarms through the generative adversarial network (GAN) to increase the number of training samples [25]. To ensure that the distribution of fault location and false alarm are similar to the actual data, we define two custom distribution to help us evaluate the effectiveness of the training sample. First is the alarm location distribution. Different from ordinary uniform probability distribution, the distribution of fault position is relatively concentrated. In other words, links with heavy load or exposed in the air are prone to failure. Second is the false alarm number distribution. This distribution is piecewise uniform, because the location of the fault would affect the number of false alarm, and the number of false alarms caused in similar location obeys Gaussian normal distribution. We test the validity of the synthetic dataset by checking the "alarm location distribution" and "false alarm number distribution" of the real sample and the generated sample. The overall accuracy of the training data mixed with the actual data and the synthesized data is 96.8%. The ratio of synthetic data to real data is about 1:3. Thus, we have 1531 total faults and 454,852 alarm samples after combining the actual data with the synthesized data. Each fault occurs independently, which means that all faults are single-link faults. We divide all the faults into 300 labeled samples and 1231 unlabeled samples which are both split to a ratio of 70/15/15 for training/validation/testing.

B. EXPERIMENTAL PREPARATION

We use a multi-core server with 12 2.5GHz Intel Core i5-7200U CPU cores, 2 NVIDIA TITAN XP GPU cores, and 32GB RAM to accelerate the training of DBN-FL model. This server runs Ubuntu 16.04, and we compile our code in Tensorflow framework using python 2.7/3.5.

The input of DBN-FL model is the dataset of each alarm information, including *Address*, *Occur_time*, *Org_severit*, and *Alarm_text*. The output of is the classification results of real alarms and false alarms.

To determine the structure of DBN-FL model, we test the number of hidden layers and the number of units in each hidden layer. To find the network architecture suitable for DBN-FL model, we choose the number of input layer unites from 1 to 10, the number of hidden layer from 1 to 10, and the number of hidden units from {32, 64, 128, 256, 512}. The most suitable architecture for DBN-FL model consists of 3 input layer unites, 4 hidden layers, and the hidden units is [256, 128, 128, 128] (bottom-top).

In the pre-training phase, the maximum iteration number of SGD algorithm and CD algorithm are 80 and 120 respectively. The initial learning rate of SGD algorithm and CD algorithm are 0.01 and 0.001, and the Mini-batch size is 4. In the fine-tuning phase, the maximum iteration number is 200, the learning rate of LM algorithm is 0.01, the Mini-batch size is 3, and the target value of error function is 0.0001.

Meanwhile, besides the proposed DBN-FL model, five other fault location models are also implemented for comparison. The first basic-DBN model is the traditional fault location model without any modification. This model adopts the pure-unsupervised learning algorithm in the pre-training phase and uses back propagation (BP) algorithm to fine-tune the parameters of DBN [7]. The second convolutional neural network (CNN) model is one of the advanced DL algorithm by taking the advantages of locally shared weights. This model shares the architecture consisting of 4 hidden layers and [256, 128, 128, 128] (bottom-top) units in each hidden layer that is similar to DBN-FL model [26]. The third CNN-Hybrid model uses the same network as CNN. But we exploit the hybrid training method to train the CNN, which is similar to the hybrid pre-training phase of DBN-FL model [27]. This aims to verify the effectiveness of the hybrid training method. The fourth SVM model is one of the most widely used fault location model for optical networks. This model uses Radial Basis Function as kernel function and is trained by one-against-all decomposition [28]. The fifth naive Bayesian (NB) network is another commonly used tool to build the fault location model, which uses a probabilistic graphical model to represent the binary variables (e.g., real alarm or false alarm) and their probabilistic independencies [29].

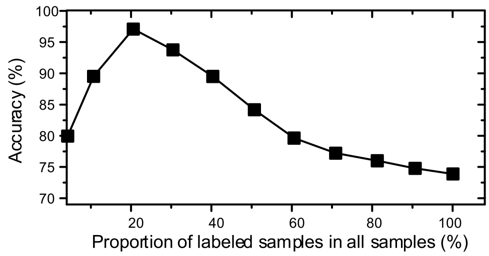


FIGURE 5. The accuracy of DBN-FL model based on different proportion of hybrid pre-training method.

C. FAULT LOCATION RESULTS AND ANALYSIS

First, we study the optimal proportion of labeled samples in whole samples. In the hybrid pre-training phase, we use supervised learning with different proportion of labeled samples to obtain the initial parameters of the RBM_1, then use the rest samples as the unlabeled samples to train the whole DBN-FL model. In the fine-tuning phase, LM algorithm is used to optimize parameters in all the cases. When the proportion of labeled samples is zero, it means that only unsupervised learning is adopted to train the DBN-FL model. As shown in Fig. 5, the accuracy of DBN-FL model is improved when we use the hybrid pre-training method. That is because the supervised learning step enables the DBN to converge to an ideal state quickly. However, when the proportion of labeled samples exceeds 20%, the accuracy of the DBN-FL model decreases. That is because unsupervised learning requires a large number of samples for fault classification. Thus, the combination of supervised learning and unsupervised learning can effectively improve the accuracy of DBN-FL model in the case of lacking training samples.

Then, we compare the performance of the proposed LM based fine-tuning algorithm and the traditional BP algorithm. Both two algorithms are deployed given the same network parameters obtained in the hybrid pre-training phase. BP algorithm uses gradient descent method for the update process of all parameters, which reduces the loss function through back propagation. Technical details of BP algorithm are shown in Ref. [30]. As shown in Fig. 6, the overall accuracy is improved by about 10% through the LM based fine-tuning phase. In the early iteration, the LM algorithm has the global convergence characteristics of maximum Steepest Gradient Descent method with initial drop and fast iteration. In the late iteration, the LM algorithm has the advantage of

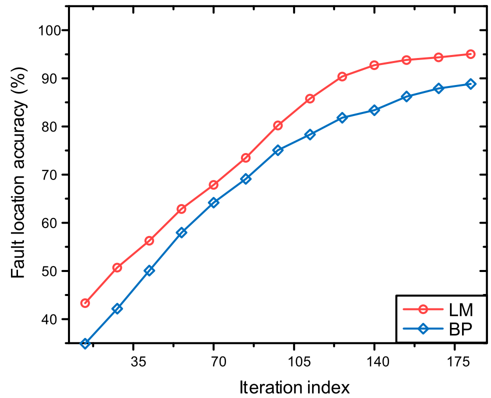


FIGURE 6. The accuracy of DBN-FL model based on two different fine-tuning algorithm.

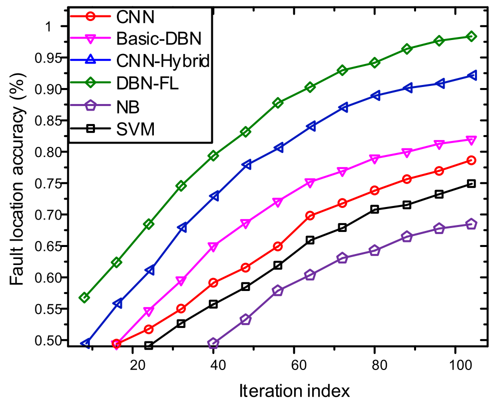


FIGURE 7. The fault location accuracy of DBN-FL model, basic-DBN model, CNN model, CNN-Hybrid model, SVM model, and NB model.

local convergence of the Gauss-Newton method and avoids the shortcoming of the Steepest Gradient Descent method.

We set up a series of experiments to illustrate the overall performance of DBN-FL model, including the fault location accuracy, the training time, and the error rate during training. We use the basic-DBN model, the CNN model, the CNN-Hybrid model, the SVM model, and the NB model as comparison algorithms.

Fig. 7 indicates that the accuracy of DBN-FL model is over 96% and much higher than the other five models. That is because the optimized DBN-FL model can extract fault features more accurately through the hybrid pre-training phase when the training samples are insufficient. The hybrid pre-training method of DBN-FL model can bring the network parameter initialization near a local optimal solution or even the global optimal solution. The CNN-Hybrid model, which also adopts a hybrid training method, also has over 90% fault location accuracy. The basic-DBN performs better than the CNN model, SVM model, and NB model, because the unique stacking structure of RBMs can fully describe the propagation characteristics of alarms through layer by layer training. As the iteration gradually increases, the locating accuracy of all six models tends to be stable. This indicates that the features contained in the alarm information have an upper limit, and the positioning accuracy will not be improved when the models have extracted the main fault feature.

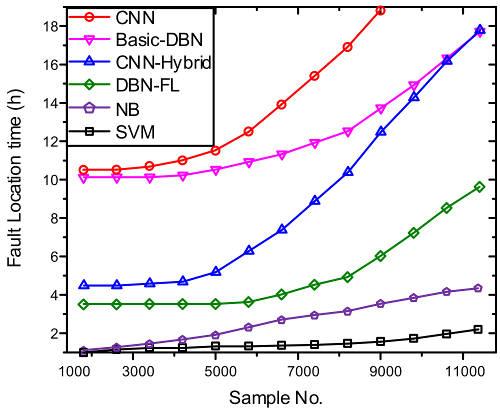


FIGURE 8. **The offline training time of DBN-FL model, basic-DBN model, CNN model, CNN-Hybrid model, SVM model, and NB model.**

The fault location time refers to the offline training time before the arrival of the new alarm information. Less training time will allow researchers to locate faults in a larger and more complex environments. Fig. 8 shows that the proposed DBN-FL model has the minimum training time of the DL approaches for different number of samples. This is because the hybrid pre-training phase and LM based fine-tuning phase of DBN-FL model can greatly reduce the convergence time and the model can be well trained in a short time. Moreover, the training time of DBN-FL model is similar with the CNN-Hybrid model when the sample number is small, because the hybrid training reduces the convergence time of the model. But the training time of DBN-FL model does not increase as sharply as CNN-Hybrid model with the growth of samples. That is because the LM based fine-tuning phase enables the DBN-FL model to quickly converge to the global optimum. In addition, the training time of two machine learning methods including the NB model and SVM model are both smaller than the DL based models due to the relatively few extracted features. The NB and SVM models are suitable for low precision fault location scenarios required fast model training.

The difference between the output of the classification models and the actual output of the sample is called the error rate. Our goal is to minimize the error of the DBN-FL model. We use the test error as an approximation of the learning error to evaluate the output. The error rate of six models in the training process is shown in Fig. 9. Note that the error rate of all the six models is very high at the beginning. The reason is that the training time is not enough for the NN to extract features. After 60 iterations of training, the differences between algorithms become more obvious. The error rate of DBN-FL model is lower than that of the other five methods. That is because the DBN-FL model optimizes the initial parameters during the pre-training phase and fine-tuning phase. Moreover, the DBN-FL model and CNN-Hybrid model are both

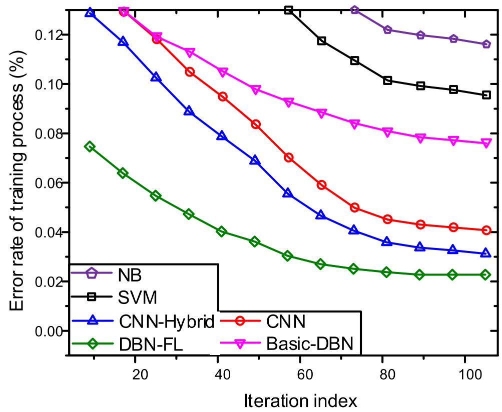


FIGURE 9. **The error rate during the model training process by using DBN-FL model, basic-DBN model, CNN model, CNN-Hybrid model, SVM model, and NB model.**

more accurate than the CNN model, the basic-DBN model, the SVM model, and the NB model, because the hybrid training method can reduce the training error by optimizing the parameters of the classification models.

D. LIMITATIONS

The main limitation of this work is that the approach is not suitable for scenarios that are more complex than single-link failures, such as multi-link fault location and fault prediction. In fault location studies with data from existing networks, it is difficult to obtain a large number of training samples. Only about 200 faults per year may occur on the municipal fronthaul transmission network, which is far from meeting the training demands of nonlinearity problems. Although the proposed DBN-FL model can reduce the demand for samples to some extent, more samples are necessary when problems become more complex.

During the process of determining the best structure of the DBN-FL model, we notice that the performance of the 2-layer RBM is very close to the optimal average performance of the 4-layer DBN-FL. This phenomenon might indicate that there might be an overfitting problem, and it might not allow our DBN-FL model to release its full potential. We could attenuate the effect of this problem by adding additional dropout steps, but it would dramatically increase the training time consumed [31].

VI. CONCLUSION

This paper presents an accurate and fast DBN-based fault location model for single-link fault of optical fronthaul networks in 5G and beyond. The DBN-FL model is constructed through the hybrid pre-training phase and LM based fine-tuning phase. In the hybrid pre-training phase, we combine the supervised learning with unsupervised learning to train the DBN-FL model. In the fine-tuning phase, the LM algorithm is adopted to fine-tune the parameters obtained from the pre-training stage. Our experiment results show that the DBN-FL model outperforms many DL based approaches in terms of fault location accuracy and offline training time.

In our future work, DBN-FL model may be used in applications with more complicated features, such as location of multi-link faults. In addition, the proposed solution would be evaluated with larger datasets.

REFERENCES

[1] H. Yang, A. Yu, X. Zhao, Q. Yao, J. Zhang, and Y. Lee, "Multi-dimensional resources allocation based on reconfigurable radio-wavelength selective switch in cloud radio over fiber networks," *Opt. Express*, vol. 26, no. 26, pp. 34719–34733, 2018.

[2] K. Honda, H. Nakamura, K. Hara, K. Sone, G. Nakagawa, Y. Hirose, T. Hoshida, J. Terada, and A. Otaka, "Wavelength adjustment of upstream signal using AMCC with power monitoring for WDM-PON in 5G mobile era," in *Proc. OFC*, 2018, pp. 1–3, Paper Tu3L. 4.

[3] V. Eramo, M. Listanti, F. G. Lavacca, and P. Iovanna, "Dimensioning of OTN/WDM rings for the transport of Ethernet/CPRI flows in 5G scenario," in *Proc. ICTON*, 2018, pp. 1–6.

[4] X. Zhao, H. Yang, H. Guo, T. Peng, and J. Zhang, "Accurate fault location based on deep neural evolution network in optical networks for 5G and beyond," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2019, pp. 1–3, Paper M3J. 5.

[5] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1383–1408, 2nd Quart., 2019.

[6] T. Panayiotou, S. P. Chatzis, and G. Ellinas, "Leveraging statistical machine learning to address failure localization in optical networks," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 10, no. 3, pp. 162–173, Mar. 2018.

[7] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[9] B. Wu, P.-H. Ho, K. L. Yeung, J. Tapolcai, and H. T. Mouftah, "Optical layer monitoring schemes for fast link failure localization in all-optical networks," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 1, pp. 114–125, 1st Quart., 2011.

[10] Y. Zhao, S. Xu, X. Wang, and S. Wang, "A new heuristic for monitoring trail allocation in all-optical WDM networks," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–5.

[11] M. Khair, J. Zheng, and H. T. Mouftah, "Distributed multi-failure localization protocol for all-optical networks," in *Proc. Int. Conf. Opt. Netw. Design Modeling (ONDM)*, Feb. 2009, pp. 1–6.

[12] J. Sadeh and H. Afradi, "A new and accurate fault location algorithm for combined transmission lines using adaptive network-based fuzzy inference system," *Electr. Power Syst. Res.*, vol. 79, no. 11, pp. 1538–1545, 2009.

[13] C. K. Jung, K. H. Kim, J. B. Lee, and B. Klöckl, "Wavelet and neuro-fuzzy based fault location for combined transmission systems," *Int. J. Elect. Power Energy Syst.*, vol. 29, no. 6, pp. 445–454, 2007.

[14] C. Mas, I. Tomkos, and O. K. Tonguz, "Failure location algorithm for transparent optical networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 8, pp. 1508–1519, Aug. 2005.

[15] S. Gosselin, J.-L. Courant, S. R. Tembo, and S. Vaton, "Application of probabilistic modeling and machine learning to the diagnosis of FTTH GPON networks," in *Proc. ONDM*, May 2017, pp. 1–3.

[16] M. Ruiz, F. Fresi, A. P. Vela, G. Meloni, N. Sambo, F. Cugini, L. Poti, L. Velasco, and P. Castoldi, "Service-triggered failure identification/localization through monitoring of multiple parameters," in *Proc. 42nd Eur. Conf. Opt. Commun. (ECOC)*, Sep. 2016, pp. 1–3.

[17] X. Zhang, L. Guo, W. Hou, Q. Zhang, and S. Wang, "Failure recovery solutions using cognitive mechanisms based on software-defined optical network platform," *Opt. Eng.*, vol. 56, no. 1, 2017, Art. no. 016107.

[18] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth, and J.-P. Elbers, "Cognitive assurance architecture for optical network fault management," *J. Lightw. Technol.*, vol. 36, no. 7, pp. 1443–1450, Apr. 1, 2018.

[19] H. Yang, Q. Yao, A. Yu, Y. Lee, and J. Zhang, "Resource assignment based on dynamic fuzzy clustering in elastic optical networks with multi-core fibers," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3457–3469, May 2019.

[20] H. Yang, J. Zhang, Y. Ji, and Y. Lee, "C-RoFN: Multi-stratum resources optimization for cloud-based radio over optical fiber networks," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 118–125, Aug. 2016.

[21] H. Yang, J. Zhang, Y. Zhao, Y. Ji, J. Han, Y. Lin, and Y. Lee, "CSO: Cross stratum optimization for optical as a service," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 130–139, Aug. 2015.

[22] H. Yang, Y. He, J. Zhang, Y. Ji, W. Bai, and Y. Lee, "Performance evaluation of multi-stratum resources optimization with network functions virtualization for cloud-based radio over optical fiber networks," *Opt. Express*, vol. 24, no. 8, pp. 8666–8678, Apr. 2016.

[23] H. Yang, J. Zhang, Y. Ji, Y. He, and Y. Lee, "Experimental demonstration of multi-dimensional resources integration for service provisioning in cloud radio over fiber network," *Sci. Rep.*, vol. 6, Jul. 2016, Art. no. 30678.

[24] G. E. Hinton and R. R. Salakhutdinov, "Replicated softmax: An undirected topic model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1607–1614.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (ICONIP)*, 2014, pp. 2672–2680.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (ICONIP)*, 2012, pp. 1097–1105.

[27] S. Li, G. Liu, X. Tang, J. Lu, and J. Hu, "An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis," *Sensors*, vol. 17, no. 8, p. 1729, Jul. 2017.

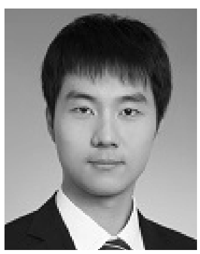
[28] P. K. Dash, S. R. Samantaray, and G. Panda, "Fault classification and section identification of an advanced series-compensated transmission line using support vector machine," *IEEE Trans. Power Del.*, vol. 22, no. 1, pp. 67–73, Jan. 2007.

[29] L. Li, Y. Zhang, W. Chen, S. K. Bose, M. Zukerman, and G. Shen, "Naïve Bayes classifier-assisted least loaded routing for circuit-switched networks," *IEEE Access*, vol. 7, pp. 11854–11867, 2019.

[30] S.-I. Horikawa, T. Furuhashi, and Y. Uchikawa, "On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 801–806, Sep. 1992.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

AO YU received the B.S. degree from the China University of Petroleum (UPC), Shandong, China, in 2014. He is currently pursuing the Ph.D. degree in the Beijing University of Posts and Telecommunications (BUPT), Beijing, China. His main research interests include deep learning, optical networks, fault location, and traffic prediction.



HUI YANG received the B.S. degree in electronical information science and technology and the Ph.D. degree in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT), in 2010 and 2014, respectively, where he is the Vice Dean and Associate Professor. He has authored or coauthored 100 papers in prestigious journals and conferences, and is the first author of more than 50 of them. He has served as the Guest Editor for IEEE JSAC, the Associate Technical Editor of IEEE ComMag, and the General Chair of ISAI'16. He received the Best Paper Award in NCCA'15 and the Young Scientist Award in IEEE ICOCN'17. He is an active reviewer or TPC member for several journals and conferences. His research interests include SDN, fixed-mobile access networks, cross-stratum optimization, data center network, flexi-gridoptical networks, blockchain, and so on.

QIUYAN YAO received her M.S. degree from the Hebei University of Engineering, Handan, China, in 2014. She is currently pursuing the Ph.D. degree with the Beijing University of Posts and Telecommunications (BUPT), Beijing, China. Her research mainly focuses on SDN, network architecture, access networks, flexi-grid optical networks, and so on.

YAJIE LI received the B.S. degree in electronical information science and technology and the Ph.D. degree in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT), in 2014 and 2018, respectively, where he is an Associate Professor. His research interests include deep learning, optical access networks, and so on.

HUIFENG GUO received the B.S. degree in electronical information science and technology and the M.S. degree in radio physics from the Beijing University of Posts and Telecommunications (BUPT), in 2005 and 2007, respectively. He is a Senior Engineer with Zhongxing Telecommunication Equipment Corporation (ZTE). Her research interests include Artificial Intelligence, optical access networks, fault location, and so on.

TAO PENG received the B.S. degree in electronical information science and technology and the M.S. degree in radio physics from the Beijing University of Posts and Telecommunications (BUPT), in 2010 and 2013, respectively. He is a Senior Engineer with Zhongxing Telecommunication Equipment Corporation (ZTE). His research interests include Artificial Intelligence, optical access networks, fault location, and so on.

HAIBIN LI received the B.S. degree in electronical information science and technology and the M.S. degree in radio physics from the Beijing University of Posts and Telecommunications (BUPT), in 2010 and 2013, respectively. He is a Senior Engineer with Zhongxing Telecommunication Equipment Corporation (ZTE). His research interests include Artificial Intelligence, optical access networks, fault location, and so on.

JIE ZHANG is a Professor and Dean with the Institute of Information Photonics and Optical Communications at BUPT. He is sponsored by more than 10 projects of the Chinese government. He has published eight books and more than 100 articles. Seven patents have also been granted. He served as a TPC member for ACP 2009, PS 2009, ONDM 2010, and so on. His research focuses on optical transport networks, packet transport networks, and so on.