Front. Optoelectron. 2021, 14(4): 513–521
https://doi.org/10.1007/s12200-020-1079-y

RESEARCH ARTICLE

A QoT prediction technique based on machine learning and NLSE for QoS and new lightpaths in optical communication networks

Yongfeng FU (✉)[1], Jing CHEN[1], Weiming WU[1], Yu HUANG[2], Jie HONG[1], Long CHEN[1], Zhongbin LI[3]

1 Hainan Power Grid Co., Ltd., Haikou 570100, China
2 Power Dispatching Control Center of China, Southern Power Grid, Shenzhen 5180008, China
3 China Southern Power Grid Digital Power Grid Research Institute Co., Ltd., Guangzhou 511458, China

Abstract In this paper, we proposed a quality of transmission (QoT) prediction technique for the quality of service (QoS) link setup based on machine learning classifiers, with synthetic data generated using the transmission equations instead of the Gaussian noise (GN) model. The proposed technique uses some link and signal characteristics as input features. The bit error rate (BER) of the signals was compared with the forward error correction threshold BER, and the comparison results were employed as labels. The transmission equations approach is a better alternative to the GN model (or other similar margin-based models) in the absence of real data (i.e., at the deployment stage of a network) or the case that real data are scarce (i.e., for enriching the dataset/reducing probing lightpaths); furthermore, the three classifiers trained using the data of the transmission equations are more reliable and practical than those trained using the data of the GN model. Meanwhile, we noted that the priority of the three classifiers should be support vector machine (SVM) > K nearest neighbor (KNN) > logistic regression (LR) as shown in the results obtained by the transmission equations, instead of SVM > LR > KNN as in the results of the GN model.

Keywords optical networks, quality of transmission (QoT), quality of service (QoS), link establishment, physical performances, bit error rate (BER), machine learning

1 Introduction

Traditional optical network performance studies focus on calculating network layer parameters, such as the network throughput and blocking probability, which are only based on the available capacity and traffic load of the network [1]. However, in real network scenarios, particularly in the case of transparent optical networks, the hindrance to the lightpath is no longer only determined by the network layer factors; the physical layer factors, such as the quality of transmission (QoT), also play a role [2–6]. Therefore, QoT must be predicted before deploying new lightpaths in transparent optical networks. Generally, QoT is associated with several physical layer parameters, such as the optical signal-to-noise ratio (OSNR), bit error rate (BER), and Q factor. These physical layer parameters reflect the performance of the optical signal. They can be quantitatively measured to verify if the QoT meets a predetermined requirement, and they are affected by several design indexes, such as the modulation format, bit rate, channel launch power, and physical links in the network. The optimization of these abundant parameters is not straightforward; consequently, system engineers often find it difficult to handle all possible lightpath deployment combinations manually.

Thus far, QoT prediction techniques have been divided into two major categories [7]. The first category comprises traditional techniques without machine learning. This category also includes two approaches: the accurate analysis model (e.g., transmission equations) for estimating the physical layer impairments and providing accurate results [8–10] and the approximation formula (e.g., Gaussian noise (GN) model) that is computationally fast but not highly accurate [11]. The approximation formula often introduces high margins, leading to underutilized

Received July 29, 2020; accepted October 1, 2020

E-mails: lizhou@csg.cn, 710256653@qq.com

network resources. The second category is based on machine learning classifiers, which have many advantages over traditional techniques. Machine learning classifiers are promising predictors that meet high precision and real-time requirements. Furthermore, they can automatically predict the QoT of unestablished lightpaths [12–18]. However, due to the inaccurate estimation of the nonlinear noise in the GN model, the classifiers obtained from the training data are unreliable. Considering this disadvantage of the GN model, we employed the wavelength division multiplexing (WDM) transmission equations to obtain the synthetic data for further training and testing to ensure more reliable results.

This paper presents a prediction technique based on machine learning classifiers, with the synthetic data generated using WDM transmission equations. This technique uses some link and signal characteristics as input features and selects the comparison results of signal BER and the forward error correction (FEC) threshold BER as the output labels. According to the FEC recommended by ITU-T G.975.1, the FEC threshold BER equals 4×10^{-3} and is denoted by T. We used three different dataset cases for the training and testing. The training dataset of case I comprises random instances extracted from the synthetic data generated by the transmission equations. The training dataset of case II is composed of instances randomly extracted from the synthetic data generated by the GN model. The training dataset of case III comprises 50% of the randomly extracted instances of the training datasets of cases I and II. The test datasets of the three cases are identical, consisting of the remaining instances in the synthetic data generated by the transmission equations. Subsequently, the performances of the three frequently used classifiers, including K nearest neighbor (KNN), logistic regression (LR), and support vector machine (SVM), were evaluated and compared using the three dataset cases.

2 Operating principle

2.1 Classifier description

Figure 1 describes that the proposed classifiers considered the following factors: number of spans, length of span, modulation format, bit rate, and channel launch power, denoted by a, b, c, d, and e, respectively. The predicted label of the classifiers denoted by f is a Boolean logic variable that equals 1 only when the signal BER is less than T; otherwise, it equals 0. We believe that there is a sophisticated function between the five input features and the output label. Based on this, we decided to use some common machine learning classifiers to fit this relationship as accurately as possible. The considered classifiers include KNN, LR, and SVM. When the difference between the predicted output and the label given in the training set reaches the stopping condition, the training phase is considered over. Therefore, we expected the trained machine learning classifiers to have excellent generalization performance, which means that they can be generalized to predict instances that are not employed during the training phase.

2.1.1 K nearest neighbors

KNN [19] is an essential and straightforward classification and regression method. A training dataset is expressed as follows:

$$D = \{(x_1,y_1),(x_2,y_2),...,(x_N,y_N)\}, \quad (1)$$

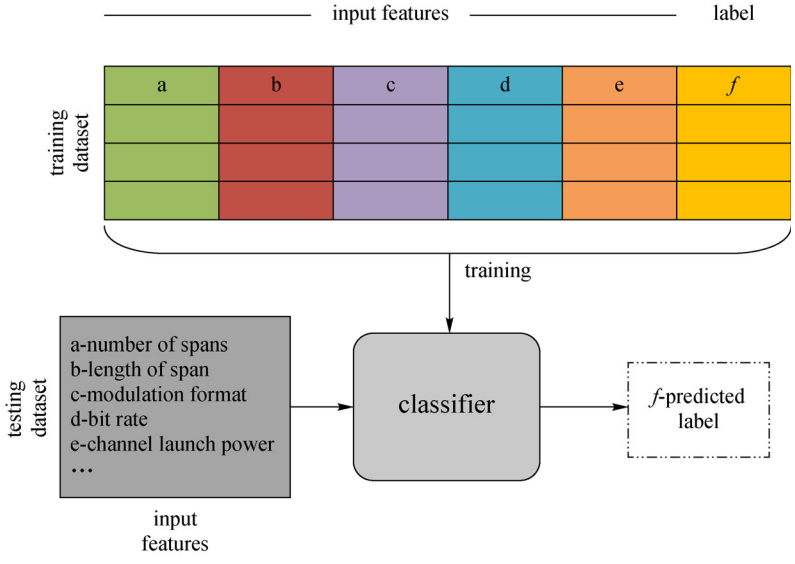

Fig. 1 Classifier structure

where $x_i \in X \subseteq \mathbf{R}^n$ is the feature vector for instances, $y_i \in Y = \{c_1,c_2,...,c_k\}$ is the class label, and $i = 1,2,...,N$ is the number of the samples. According to the given distance measure (commonly used distance metrics are Euclidean distances) [19], the k points closest to x are found in training dataset D, and the neighborhood of x covering the k points is denoted by $N_k(x)$. In $N_k(x)$, the class label, y of x, is determined by the classification decision rules (e.g., majority voting), as follows:

$$y = \arg\max_{c_j} \sum_{x_i \in N_k(x)} I(y_i = c_j),$$

$$i = 1,2,...,N,\ j = 1,2,...,K. \tag{2}$$

In Eq. (2), I is the indicator function. $I = 1$ when $y_i = c_j$; otherwise, $I = 0$.

2.1.2 Logistic regression

The following conditional probability distribution is called the LR model [19].

$$P(Y = 0|x) = \frac{1}{1 + \mathrm{e}^{w \cdot x + b}},$$

$$P(Y = 1|x) = 1 - P(Y = 0|x) = \frac{\mathrm{e}^{w \cdot x + b}}{1 + \mathrm{e}^{w \cdot x + b}}. \tag{3}$$

Here, $x \in \mathbf{R}^n$ is the input vector, and $Y \in \{0,1\}$ is the output. $w \in \mathbf{R}^n$ is called the weight vector, and $b \in \mathbf{R}$, the bias. Occasionally, for convenience, the weight vector and the input vector are expanded as follows:

$$w = \left(w^{(1)},w^{(2)},...,w^{(n)},b\right)^{\mathrm{T}},$$

$$x = \left(x^{(1)},x^{(2)},...,x^{(n)},1\right)^{\mathrm{T}}. \tag{4}$$

At this point, the LR model becomes

$$P(Y = 0|x) = \frac{1}{1 + \mathrm{e}^{w \cdot x}},$$

$$P(Y = 1|x) = \frac{\mathrm{e}^{w \cdot x}}{1 + \mathrm{e}^{w \cdot x}}. \tag{5}$$

For a given training dataset, $D = \{(x_1,y_1),(x_2,y_2),..., (x_N,y_N)\}$, where $x_i \in \mathbf{R}^n$ and $y_i \in \{0,1\}$, we can use the maximum likelihood estimation method [19] to estimate the model parameter (w) and subsequently obtain the LR model.

2.1.3 Support vector machine

SVM [19] is a machine learning algorithm based on statistical learning theory. It uses the principle of structural risk minimization instead of empirical risk minimization to solve the problem of small sample learning. It also employs the idea of kernel function to transform the problem of nonlinear space into linear space, which reduces the complexity of the algorithm. Due to its complete theoretical basis and good learning performance, SVM is uniquely suitable for solving finite samples, nonlinearity, and high-dimensional pattern recognition problems. It has received significant attention in the field of machine learning and has been successfully applied in many fields.

For linearly separable problems, given a dataset, D, where $x_i \in \mathbf{R}^n$ and $y_i \in \{0,1\}$, α_i is the Lagrange multiplier. The steps below are followed:

Constructing and solving constrained optimization problems:

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i,$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i = 0,$$

$$\alpha_i \geqslant 0, \quad i = 1,2,...,N. \tag{6}$$

Finding the optimal solution, $\alpha^* = (\alpha_1^*,\alpha_2^*,...,\alpha_N^*)^{\mathrm{T}}$. Thereafter, calculating

$$w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i. \tag{7}$$

Selecting a positive component of α^*, i.e., α_j^*, after which b^*, which is the displacement term, is obtained, as follows:

$$b^* = y_j - \sum_{i=1}^{N}\alpha_i^* y_i (x_i \cdot x_j). \tag{8}$$

The separating hyperplane and classification decision functions are obtained as follows:

$$w^* \cdot x + b^* = 0,\ f(x) = \mathrm{sign}(w^* \cdot x + b^*), \tag{9}$$

where $\mathrm{sign}(\cdot)$ denotes the symbolic function.

For nonlinearly separable problems, we selected the appropriate kernel function $K(x,z)$ and parameter C, to construct and solve the corresponding optimization problem, and obtain the optimal solution, $\alpha^* = (\alpha_1^*,\alpha_2^*,...,\alpha_N^*)^{\mathrm{T}}$. Thereafter, we obtain the corresponding b^*. Thus, the classification decision function can be constructed at this time:

$$f(x) = \mathrm{sign}\left[\sum_{i=1}^{N}\alpha_i^* y_i K(x \cdot x_i) + b^*\right]. \tag{10}$$

2.2 Data generation

2.2.1 Data generation using the Gaussian noise model

First, we used candidate lightpaths to calculate the nonlinear OSNR from the ratio of the channel launch power (P_{TX}) to the sum of linear (P_{ASE}) and nonlinear noises (P_{NLI}), as expressed in Eq. (11).

$$\mathrm{OSNR}=\frac{P_{\mathrm{TX}}}{P_{\mathrm{ASE}}+P_{\mathrm{NLI}}}. \tag{11}$$

Here, P_{ASE} denotes the power introduced by the amplified spontaneous emission (ASE), as expressed in Eq. (12), where N_{S} represents the number of span, F is the noise figure, G is the amplifier gain, h is Planck's constant, v_0 is the channel central frequency, and B_{N} is the noise bandwidth.

$$P_{\mathrm{ASE}}=N_{\mathrm{S}}\times F\times(G-1)\times h\times v_0\times B_{\mathrm{N}}. \tag{12}$$

P_{NLI} is the power introduced by nonlinear interference, obtained using the GN model and the Matlab code given in Ref. [20].

The ratio of the energy per bit to the noise power spectral density (E_{b}/N_0) is obtained using Eq. (13), as follows:

$$E_{\mathrm{b}}/N_0=\mathrm{OSNR}-10\times\log_{10}(\log_2 M)-10\times\log_{10}(D_{\mathrm{rate}}/2\times\log_2 M/B_{\mathrm{ref}}), \tag{13}$$

where M is the number of states per symbol, D_{rate} of the signal, and B_{ref} is the channel bandwidth. Thus, we can calculate the BER as a function of E_{b}/N_0 using Eq. (14); we obtain $a=0.5$ and $d=\log_2 M$, and $a=3/8$ and $d=10$ for phase shift keying (PSK) and 16 quadrature amplitude modulation (16QAM), respectively, where a and d are constant coefficients related to the modulation formats.

$$\mathrm{BER}=a\times\mathrm{erfc}\left(\sqrt{\frac{E_{\mathrm{b}}/N_0}{d}}\right). \tag{14}$$

After the above steps, we can use the BER comparison with T to obtain the label of each candidate lightpath. The label equals 1 when the BER $<T$; otherwise, it equals 0.

2.2.2 Data generation using transmission equations

Numerical simulation is essential for the research and design of WDM systems. The nonlinear Schrödinger equation (NLSE) is written as a set of coupled differential equations [8–10]. Here, cross-phase modulation (XPM) is considered in terms of coupling, and those that correspond to the four-wave mixing (FWM)are omitted. Therefore, the transmission equation for the complex amplitude A_i of the ith channel can be expressed as

$$\frac{\partial A_i}{\partial z}+\frac{\alpha}{2}A_i+\beta_{1i}\frac{\partial A_i}{\partial t}-\frac{\mathrm{j}}{2}\beta_{2i}\frac{\partial^2 A_i}{\partial t^2}-\frac{1}{6}\beta_{3i}\frac{\partial^3 A_i}{\partial t^3}=-\mathrm{j}\gamma_i\left(|A_i|^2+2\sum_{\substack{k\\k\neq i}}|A_k|^2\right)A_i. \tag{15}$$

Here, α and β_{1i} denote the fiber loss and constant group delay, respectively. β_{2i}, β_{3i}, and γ_i represent the fiber dispersion and nonlinearity, respectively.

The solution of the nonlinear operator causes self-phase modulation (SPM) and XPM, resulting in phase modulation; $A_i(t,z+\Delta z/2)$ is the complex amplitude of channel i after the first linear half-step of the symmetric split-step Fourier method (SSFM) and before the nonlinearity. $A_i'(t,z+\Delta z/2)$ is the input complex amplitude for the next linear half-step after considering the nonlinearity. Here, Δz denotes one split-step of length.

$$\begin{aligned}A_i'\left(t,z+\frac{\Delta z}{2}\right)&=A_i\left(t,z+\frac{\Delta z}{2}\right)\\&\times\exp\left(-\mathrm{j}\gamma_i\Delta z\left(\left|A_i\left(t,z+\frac{\Delta z}{2}\right)\right|^2+2\sum_{\substack{k\\k\neq i}}\left|A_k\left(t,z+\frac{\Delta z}{2}\right)\right|^2\right)\right)\\&=A_i\left(t,z+\frac{\Delta z}{2}\right)\\&\times\exp\left(-\mathrm{j}\left(\phi_{i,\mathrm{SPM}}(t,z)+\sum_{\substack{k\\k\neq i}}\phi_{ik,\mathrm{XPM}}(t,z)\right)\right).\end{aligned} \tag{16}$$

According to Ref. [8], the XPM-induced phase modulation can be expressed in the frequency domain, as follows:

$$\begin{aligned}&\phi_{ik,\mathrm{XPM}}(f,z)\\&=2\gamma_i\frac{1-\exp\left(-(\alpha-\mathrm{j}2\pi f d_{ik})\Delta z\right)}{\alpha-\mathrm{j}2\pi f d_{ik}}\\&\times\exp\left((\alpha-\mathrm{j}2\pi f d_{ik})\frac{\Delta z}{2}\right)F\left\{\left|A_k\left(t,z+\frac{\Delta z}{2}\right)\right|^2\right\},\end{aligned} \tag{17}$$

where d_{ik} is the walk off between channels i and k, and F denotes the Fourier transform. Equation (17) expresses the average of all relative positions within one split-step. Thus, this method enables a relatively large step size, which reduces the computation time. Meanwhile, we neglected the conversion of the induced phase modulation to the intensity modulation, and we used Eq. (12) to simulate the ASE noise introduced by Erbium-doped fiber amplifiers (EDFAs). Based on the above preparations, the complex amplitude of the receiving signal can be simulated by solving the NLSE using the SSFM under various lightpaths, and the corresponding BER can be obtained after electrical compensation and final decision. The label equals 1 when BER $< T$; otherwise, it equals 0.

3 Simulation setup and result analysis

Considering the difficulty associated with obtaining real field monitoring data, we used synthetic data to train and test the proposed machine learning-based classifiers. For building a knowledge base (KB) consisting of the synthetic data, it is necessary to calculate the BERs corresponding to the varying different input features. As mentioned above, compared with the approximate formula method using the GN model, the method of obtaining BERs by solving the WDM transmission equations with the SSFM can provide relatively accurate results. Moreover, although the computational complexity is high, the generation of synthetic data can be regarded as the preliminary preparation stage of the classifier training, and the time spent is not considered in the performance index of the classifiers in future predictions. Therefore, we used a relatively time-consuming solution to obtain the BERs by solving the WDM transmission equations using SSFM. Meanwhile, to verify the performance improvement using the synthetic data generated by the transmission equations, we generated synthetic data using the GN model for comparison. Therefore, KB comprises two parts: part I is generated by the transmission equations and part II by the GN model.

For KB part I, generated by the transmission equations, Fig. 2(a) shows the simulation system structure used in the

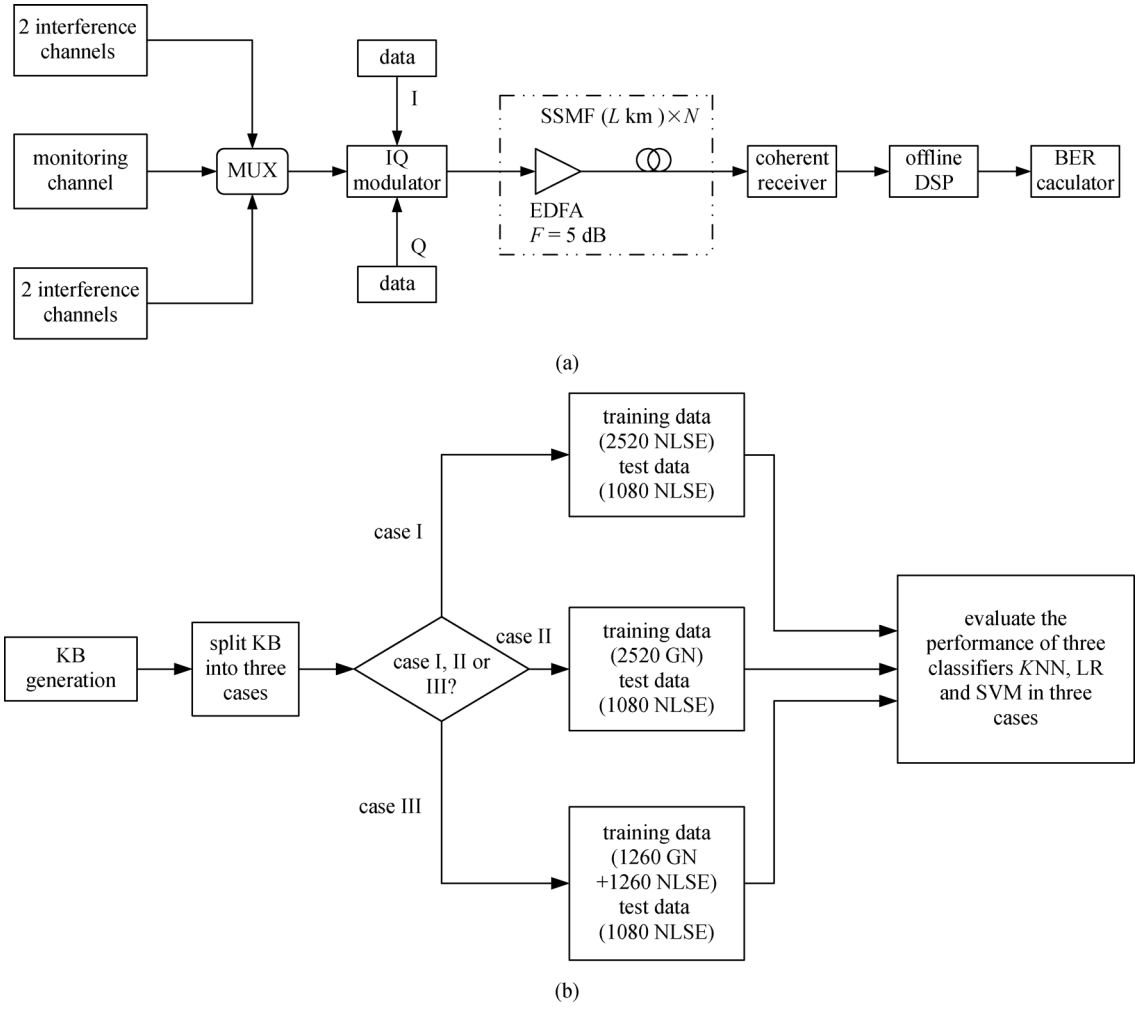

Fig. 2 (a) Simulation setup. (b) Workflow of the classification process. MUX: multiplexer, IQ modulator: in-phase quadrature modulator, SSMF: standard single-mode fiber, EDFA: Erbium-doped fiber amplifier, DSP: digital signal processing

process of generating the synthetic data. It comprises five transmitters, an optical link with N optical fiber spans and amplifiers, a digital coherent receiver integrated with an electric domain compensation function, and a BER calculator. The amplifier accurately compensates for the fiber loss but also adds noise. Without the loss of generality, the center channel is assumed to be a probe channel, and its performance and the XPM effect are studied, whereas the other four channels are interference channels. The fiber parameters include the group velocity dispersion coefficient, fiber loss, and nonlinear parameters. The receiving end performs the demodulation and the electric domain compensation to determine the BERs of the probe channel. Each instance comprises the total lightpath length ranging from 50 to 5000 km; span lengths of 50, 80, and 100 km; the number of spans ranging from 1 to 50; channel launch power varying from –10 to 4 dBm with a 2 dBm interval; and three modulation formats, quadrature phase shift keying (QPSK), 16QAM, and 64QAM. We set the EDFA noise to 5 dB. The channel spacing was set to 50 GHz. The noise bandwidth was 32 GHz, and the symbol rate used was 32 GBaud. According to the principles introduced in Section 2.2.2, the resulting partial KB contains BER observations of 3600 different lightpaths and corresponding labels. Among them, there are 1118 instances of $\mathrm{BER} < T$ and 2482 instances of $\mathrm{BER} > T$.

For KB part II, generated by the GN model, each instance comprises the total lightpath length ranging from 50 to 5000 km; span lengths of 50, 80, and 100 km; the number of spans ranging from 1 to 50; and channel launch power ranging from –10 to 4 dBm. The power interval is 2 dBm, and there are three modulation formats, QPSK, 16QAM, and 64QAM. We set the noise of the EDFAs to 5 dB. The channel spacing is set to 50 GHz. The noise bandwidth is 32 GHz, and the symbol rate is 32 GBaud. First, the GN model is used to roughly estimate the nonlinear noise, and the OSNR is calculated by combining the nonlinear noise with the ASE noise. Thereafter, the corresponding BER is obtained by the mapping relationship between OSNR and BER. According to the principles introduced in Section 2.2.1, the dataset based on the GN model has a total of 3600 instances, where the numbers of $\mathrm{BER} < T$ and $\mathrm{BER} > T$ instances are 1269 and 2331, respectively.

Figure 3 shows that the KB was divided into three cases. According to the division method described below, these three cases all have 2520 instances for training and 1080 instances for testing. For case I, the training dataset was entirely generated by 2520 instances randomly extracted from KB part I, and the test dataset comprises the remaining 1080 instances. For case II, the training dataset was obtained by 2520 instances randomly extracted from KB part II, and the test dataset remains unchanged from case I. For case III, we randomly extracted 50% of the training dataset of case II and 50% of the training dataset of case I and merged them to form a new training dataset, whereas the test dataset remains the same as case I and case II. Note that the test dataset is the same dataset generated by the transmission equations in the three cases. This is because these data are sufficiently similar to the real field data to accurately measure the performance of the trained classifiers.

We used the various machine learning classifiers mentioned earlier to classify the lightpaths; thereafter, we compared their performances and selected the best machine learning classifier, as shown in Fig. 2(b). We employed the trained optimal parametric classifiers to predict the BERs of the lightpaths in the test dataset. If the BER of a lightpath is below T, it is a Boolean logic variable equal to 1; otherwise, it is equal to 0, which is the label predicted by the classifier. Here, a label equal to 1 means that it is a “good QoT”, and a label equal to 0 implies that it is a “poor QoT”.

The confusion matrixes in Fig. 4 show the classification results achieved by the classifiers considered in this work

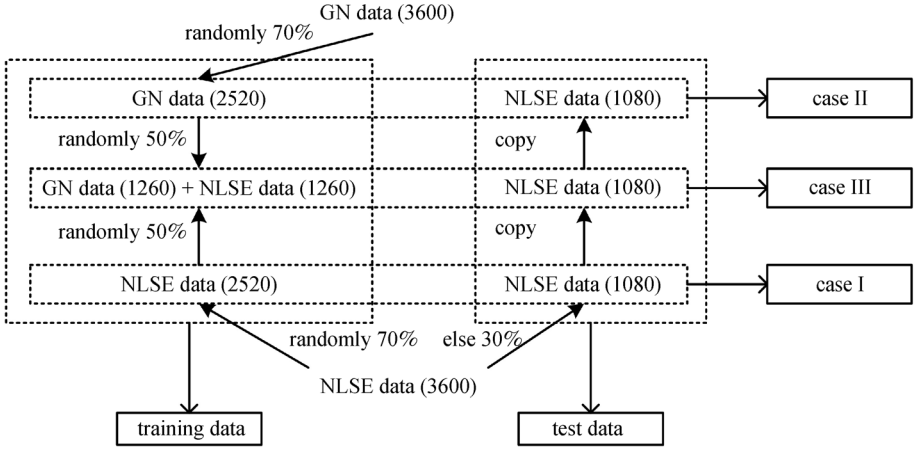

Fig. 3 Three KB cases

Table 1 Performance of the classifiers

classifier type	case I			case II			case III		
	*K*NN	LR	SVM	*K*NN	LR	SVM	*K*NN	LR	SVM
accuracy/%	97.87	89.63	99.17	83.24	84.44	88.24	97.13	85.37	99.35
error rate/%	2.13	10.37	0.83	16.76	15.56	11.76	2.87	14.63	0.65
false positives rate/%	1.02	4.63	0.37	9.26	8.33	5.00	1.02	6.39	0.37

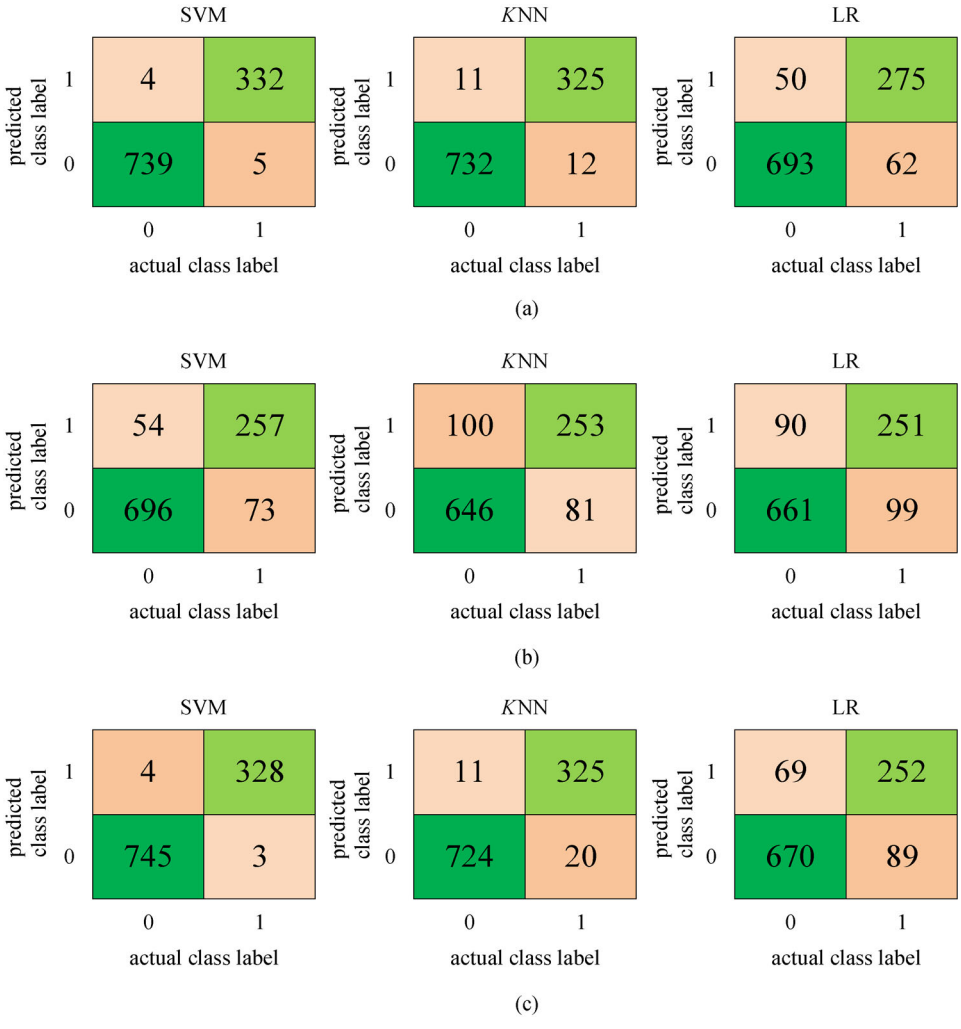

Fig. 4 Confusion matrix for (a) case I, (b) case II, and (c) case III

for the three different cases. The columns of the matrixes describe the actual classes of the test instances, whereas the rows are the predicted classes by each classifier. Two metrics were generally used to evaluate the classifiers: the classification accuracy and the false positive rate. The false positive (anticipated instances of "good QoT", when the actual class is "poor QoT") rate can be used to further refine the prediction performance of each classifier. Figures 4(a)–4(c) show the corresponding confusion matrix for the three cases, respectively. Table 1 depicts that after comparing the results of the three cases, we found that when the dataset is composed of case I, the three classifiers, *K*NN, LR, and SVM, achieve the highest classification accuracies, i.e., 97.87%, 89.63%, and 99.17%, respectively. However, when the dataset comprises case II, the classification accuracies of the three classifiers are the lowest, i.e., 83.24%, 84.44%, and 88.24%, respectively. When the dataset is composed of case III, the classification accuracies of the three classifiers are between those of the previous two, i.e., 97.13%, 85.37%, and 99.35%, respectively. By observing the relationship between the false positive rate and the classification accuracy in the three cases, it is easy to find that the false positive rate is negatively correlated to

the classification accuracy. This indicates that the higher the accuracy of the classifier, the lower the false positive rate of the classifier. It is evident that the classifier with the highest classification accuracy also has the lowest false positive rate, accounting for its best overall performance.

The above comparison results indicate that the classifiers corresponding to case I exhibit the best performance when tested using the test data that are sufficiently similar to real field data. We believed that this is mainly because the training dataset of case I are composed entirely of accurate synthetic data generated by the transmission equations, and the trained classifiers are the most reliable and practical. In case II, the performance is poor when tested using the same data. We believed this is because the training dataset is composed entirely of inaccurate synthetic data generated by the GN model, and the trained classifiers are unreliable. The accuracies of cases I and II also indicate the extent of the GN model deviation from the more realistic transmission equations. This is further verified by case III. In case III, because the training dataset is a mixture of the data generated by both the GN model and the transmission equations, the performance of the trained classifiers was between those of the previous two. Comparing the results obtained in cases I and II, we found that SVM is the best classifier in both cases, although *K*NN exhibits a different trend. In case I, the classification accuracy of *K*NN is slightly lower than that of SVM but considerably higher than that of LR. In case II, *K*NN is not as accurate as LR. The transmission equations approach constitute a better alternative to the GN model (or other similar margin-based models) in the absence of real data (i.e., at the deployment stage of a network) or the case that real data are scarce (i.e., for enriching the dataset/reducing probing lightpaths). We believed the priority of the three classifiers should be SVM > *K*NN > LR, as shown in case I, rather than SVM > LR > *K*NN, shown in case II.

4 Conclusions

In this paper, we presented three supervised learning classifiers for predicting the QoT of unestablished lightpaths. By comparing the performance of the three different classifiers: *K*NN, LR, and SVM, in three different dataset cases, we found that when the training dataset comprises the data generated by the transmission equations, the classification accuracies of *K*NN, LR, and SVM are 97.87%, 89.63%, and 99.17%, respectively. However, when the training dataset comprises the data generated by the GN model, the classification accuracies are 83.24%, 84.44%, and 88.24%, respectively. Considering that the transmission equations approach constitute a better alternative to the GN model (or other similar margin-based models) in the absence of real data (i.e., at the deployment stage of a network) or the case that real data are scarce (i.e., for enriching the dataset/reducing probing lightpaths), the three classifiers trained using the data generated by the transmission equations are more reliable and practical than those trained by the data generated by the GN model. Moreover, we noticed that the priority of the three classifiers should be SVM > *K*NN > LR, instead of SVM > LR > *K*NN. In future works, the optimization methods for reducing the KB size would be investigated to propose other novel classifiers with better performance.

Acknowledgements This work was supported by the Open Foundation of China Southern Power Grid Co., Ltd.

References

1. Kulkarni P, Tzanakaki A, Machuka C M, Tomkos I. Benefits of *Q*-factor based routing in WDM metro networks. In: Proceedings of European Conference on Optical Communication (ECOC). Glasgow: IET, 2005, Th3.5.7
2. Feng Q, Li W, Zheng Q, Wang Y, Hu C, Xie Y, Lian W. The OTDR with high dynamic range based on LFM signal and FDM. IEEE Photonics Technology Letters, 2020, 32(7): 359–362
3. Hu C, Li W, Zheng H, Feng Q, Zheng Q, Wang Y. A novel cost-effective and distributed in-band OSNR monitoring method using Gaussian process regression. IEEE Photonics Journal, 2019, 11(4): 720431
4. Pachnicke S, Paschenda T, Krummrich P M. Physical impairment based regenerator placement and routing in translucent optical networks. In: Proceedings of Optical Fiber Communication Conference. San Diego: IEEE, 2008, OWA2
5. Markidis G, Sygletos S, Tzanakaki A, Tomkos I. Impairment-constraint-based routing in ultralong-haul optical networks with 2R regeneration. IEEE Photonics Technology Letters, 2007, 19(6): 420–422
6. Politi C T, Anagnostopoulos V, Matrakidis C, Stavdas A. Physical layer impairment aware routing algorithms based on analytically calculated *Q*-factor. In: Proceedings of Optical Fiber Communication Conference. Anaheim: IEEE, 2006, OFG1
7. Barletta L, Giusti A, Rottondi C, Tornatore M. QoT estimation for unestablished lighpaths using machine learning. In: Proceedings of Optical Fiber Communication Conference. Los Angeles: IEEE, 2017, Th1J.1
8. Zheng Q, Li W, Yan R, Feng Q, Xie Y, Wang Y. XPM mitigation in WDM systems using split nonlinearity compensation. IEEE Photonics Journal, 2019, 11(6): 7205411
9. Zheng Q, Yuan Z, Li Y, Li W. Optimization of split transmitter-receiver digital nonlinearity compensation in Bi-directional Raman unrepeatered system. Applied Sciences (Basel, Switzerland), 2018, 8(6): 972
10. Shao J, Liang X, Kumar S. Comparison of split-step Fourier schemes for simulating fiber optic communication systems. IEEE Photonics Technology Letters, 2014, 6(4): 7200515
11. Poggiolini P, Bosco G, Carena A, Curri V, Jiang Y, Forghieri F. The GN-model of fiber non-linear propagation and its applications. Journal of Lightwave Technology, 2014, 32(4): 694–721
12. Mo W, Huang Y, Zhang S, Ip E, Kilper D C, Aono Y, Tajima T.

ANN-based transfer learning for QoT prediction in real-time mixed line-rate systems. In: Proceedings of Optical Fiber Communication Conference. San Diego: IEEE, 2018, W4F.3

13. Mata J, Miguel I d, Durán R J, Aguado J C, Merayo N, Ruiz L, Fernández P, Lorenzo R M, Abril E J, Tomkos I. Supervised machine learning techniques for quality of transmission assessment in optical networks. In: Proceedings of International Conference on Transparent Optical Networks (ICTON). Bucharest: IEEE, 2018, We.B3.5
14. Morais R M, Pedro J. Evaluating machine learning models for QoT estimation. In: Proceedings of International Conference on Transparent Optical Networks (ICTON). Bucharest: IEEE, 2018, Tu.A3.4
15. Bouda M, Oda S, Akiyama Y, Paunovic D, Hoshida T, Palacharla P, Ikeuchi T. Demonstration of continuous improvement in open optical network design by QoT prediction using machine learning. In: Proceedings of Optical Fiber Communication Conference. San Diego: IEEE, 2019, M3Z.2
16. Panayiotou T, Savva G, Shariati B, Tomkos I, Ellinas G. Machine learning for QoT estimation of unseen optical network states. In: Proceedings of Optical Fiber Communication Conference. San Diego: IEEE, 2019, Tu2E.2
17. Azzimonti D, Rottondi C, Tornatore M. Using active learning to decrease probes for QoT estimation in optical networks. In: Proceedings of Optical Fiber Communication Conference. San Diego: IEEE, 2019, Th1H.1
18. Aladin S, Tremblay C. Cognitive tool for estimating the QoT of new lightpaths. In: Proceedings of Optical Fiber Communication Conference. San Diego: IEEE, 2018, M3A.3
19. Morais R M, Pedro J. Machine learning models for estimating quality of transmission in DWDM networks. Journal of Optical Communications and Networking, 2018, 10(10): D84–D99
20. Dar R, Feder M, Mecozzi A, Shtaif M. Accumulation of nonlinear interference noise in fiber-optic systems. Optics Express, 2014, 22 (12): 14199–14211