



Failure prediction using machine learning and time series in optical network

ZHILONG WANG,¹ MIN ZHANG,^{1,*} DANSHI WANG,¹ CHUANG SONG,¹ MIN LIU,¹ JIN LI,¹ LIQI LOU,¹ AND ZHUO LIU²

¹State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

²China Mobile Communications Corporation, Beijing 100033, China

*mzhang@bupt.edu.cn

Abstract: In this paper, we propose a performance monitoring and failure prediction method in optical networks based on machine learning. The primary algorithms of this method are the support vector machine (SVM) and double exponential smoothing (DES). With a focus on risk-aware models in optical networks, the proposed protection plan primarily investigates how to predict the risk of an equipment failure. To the best of our knowledge, this important problem has not yet been fully considered. Experimental results showed that the average prediction accuracy of our method was 95% when predicting the optical equipment failure state. This finding means that our method can forecast an equipment failure risk with high accuracy. Therefore, our proposed DES-SVM method can effectively improve traditional risk-aware models to protect services from possible failures and enhance the optical network stability.

© 2017 Optical Society of America

OCIS codes: (060.0060) Fiber optics and optical communications; (060.4257) Networks, network survivability; (060.4261) Networks, protection and restoration.

References and links

1. S. Ramamurthy, L. Sahasrabudde, and B. Mukherjee, "Survivable WDM mesh networks," *J. Lightwave Technol.* **21**(4), 870–883 (2003).
2. X. Shao, Y. Bai, X. Cheng, Y.-K. Yeo, L. Zhou, and L. H. Ngoh, "Best effort SRLG failure protection for optical WDM networks," *J. Opt. Commun. Netw.* **3**(9), 739–749 (2011).
3. F. Dikbiyik, M. Tornatore, and B. Mukherjee, "Minimizing the Risk From Disaster Failures in Optical Backbone Networks," *J. Lightwave Technol.* **32**(18), 3175–3183 (2014).
4. H. Weigang, G. Lei, Y. Cunqian, and Z. Yue, "Risk-aware virtual network embedding in optical data center networks," in *Proceedings of OptoElectronics and Communications Conference (OECC 2016)*, pp. 1–3.
5. S. Huang, B. Guo, X. Li, J. Zhang, Y. Zhao, and W. Gu, "Pre-configured polyhedron based protection against multi-link failures in optical mesh networks," *Opt. Express* **22**(3), 2386–2402 (2014).
6. X. Li, S. Huang, S. Yin, B. Guo, Y. Zhao, J. Zhang, M. Zhang, and W. Gu, "Shared end-to-content backup path protection in k-node (edge) content connected elastic optical datacenter networks," *Opt. Express* **24**(9), 9446–9464 (2016).
7. D. Wang, M. Zhang, Z. Li, Y. Cui, J. Liu, Y. Yang, and H. Wang, "Nonlinear decision boundary created by a machine learning-based classifier to mitigate nonlinear phase noise," in *Proceedings of Optical Communication (ECOC 2015)*, pp. 1–3.
8. D. Wang, M. Zhang, Z. Li, C. Song, M. Fu, J. Li, and X. Chen, "System impairment compensation in coherent optical communications by using a bio-inspired detector based on artificial neural network and genetic algorithm," *Opt. Commun.* **399**, 1–12 (2017).
9. D. Wang, M. Zhang, M. Fu, Z. Cai, Z. Li, H. Han, Y. Cui, and B. Luo, "Nonlinearity Mitigation Using a Machine Learning Detector Based on k-Nearest Neighbors," *IEEE Photonics Technol. Lett.* **28**(19), 2102–2105 (2016).
10. D. Wang, M. Zhang, J. Li, Z. Li, J. Li, C. Song, and X. Chen, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Express* **25**(15), 17150–17166 (2017).
11. Y. Yuan, M. Zhang, P. Luo, Z. Ghassemlooy, D. Wang, X. Tang, and D. Han, "SVM detection for superposed pulse amplitude modulation in visible light communications," in *Proceedings of Communication Systems, Networks and Digital Signal Processing (CSNDSP 2016)*, pp. 1–5.
12. E. Osuna, R. Freund, and F. Girosi, "Support vector machines: Training and applications," *AI Memo*, AIM-1602 (1997).

13. S. Suthaharan, "Support Vector Machine," in *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, S. Suthaharan, (Springer US, 2016).
14. D. Wang, M. Zhang, Z. Cai, Y. Cui, Z. Li, H. Han, M. Fu, and B. Luo, "Combating nonlinear phase noise in coherent optical systems with an optimized decision processor based on machine learning," *Opt. Commun.* **369**, 199–208 (2016).
15. T. Knebel, S. Hochreiter, and K. Obermayer, "An SMO algorithm for the potential support vector machine," *Neural Comput.* **20**(1), 271–287 (2008).
16. A. C. Adamuthe, R. A. Gage, G. T. Thampi, and Ieee, "Forecasting Cloud Computing using Double Exponential Smoothing Methods," in *Proc. Advanced Computing and Communication Systems (ACCS 2015)*, pp. 5.
17. C. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification." <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
18. S. S. Keerthi and C. J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.* **15**(7), 1667–1689 (2003).

1. Introduction

With the development of the Internet, an increasing number of services involve massive data transfer in optical networks. When an optical network suffers a failure, an immense loss of data will occur. To reduce the damage, many optical network protection algorithms have been proposed, including shared-path protection (SPP) [1], best-effort shared risk link group (SRLG) failure protection [2], and others. However, these algorithms passively protect the optical network and reduce damage only after a failure occurs, which means the data are still lost on account of the time delay of protection and recovery.

Therefore, early-warning and proactive protection is required. In [3], risk models are proposed in which high-risk services are switched to a low-risk path to prevent damage from disaster failures in optical backbone networks. In [4], risk-aware models are presented to prevent data loss in data center networks. In [5] and [6], k -edge and k -node models are proposed to protect optical mesh networks and data center networks from multi-failures (e.g., disasters, massive power outages or mass destruction attacks). The above works provide a means of switching the services or backing up the data when the risk exists for each link or node (mainly in disaster/attack scenarios); however, they do not consider how to forecast the risk. In fact, a means of predicting an equipment failure in an optical network and providing protective action before a failure occurs remain inadequately investigated. By predicting equipment failures in daily use, the aforementioned protection algorithms based on risk-aware models could thus be extended to daily equipment fault scenarios. Accordingly, the optical network would be more robust and the user quality of experience (QoE) would be greatly improved.

Machine learning can be applied to advance the above efforts. Machine learning is a series of intelligent algorithms that can learn the inherent information of the training data. The inherent information is then abstracted into a decision model that provides guidance for further work. These algorithms can perform detection and decision-making in optical communications and improve the system performance.

The present authors recently demonstrated their means of reducing nonlinear phase noise [7,8], overcoming system impairments in fiber communications [9], optical performance monitoring [10] and performing data detection in visible light communications [11]. Because failure prediction is an estimation problem, and the operating data contain internal relations, machine learning is thus suited to this problem. These authors believe that this advanced technology can achieve good results in optical network failure prediction. Nevertheless, to the best of our knowledge, the machine learning algorithm has not been used in optical network equipment failure prediction.

In this paper, a method of predicting optical network equipment failure based on a method combining double-exponential smoothing (DES) and a specific support vector machine (SVM) is therefore proposed. The proposed DES-SVM method can be applied in a software-defined network. The main algorithm is not complex; therefore, the presented method is easy to establish in the network controller. Equipment information from the monitoring record of

the network management system in an actual wavelength division multiplex (WDM) network was used to prove the validity of our method.

2. Principles and procedures

A. Principle of the DES-SVM prediction method

A typical machine learning algorithm generally contends with the problem of an immense amount of prior data (e.g., typically more than 100,000 items) being required for training the decision model. With this scale, the controller should have a strong storage capacity and powerful calculating ability. In a real network, determining whether a node is available is performed by many different kinds of equipment. If the controller monitors and predicts the state of each piece of equipment with that scale, it will suffer a seriously heavy burden. It is necessary to choose a machine learning algorithm that can train a highly accurate model with fewer data. For the SVM algorithm, it has a high efficiency and high accuracy at a data amount less than 5,000, which is suitable for the practical application in the WDM optical network.

SVM is essentially a binary classification algorithm that screens the support vectors from the training data and uses them to establish a decision function [12,13]. The main principles of the approach are shown in Fig. 1. The data from the record of the optical network operator are divided into two groups, namely, the equipment failure data and the normal data, represented as the triangles and circles respectively in Fig. 1(a). These data can be separated by drawing a boundary line. In this case, there are many possible boundary lines, but only one optimal decision boundary has the maximum margin. This optimal decision boundary is called a hyperplane, and finding this line is the main purpose of the SVM algorithm. It is mathematically expressed as Eq. (1).

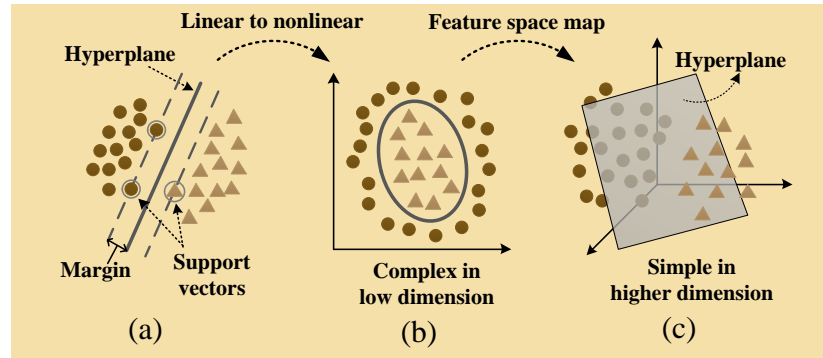


Fig. 1. The main principles of SVM.

$$y(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + b \quad (1)$$

where the vector \mathbf{w} and scalar b represent the undetermined coefficients of the hyperplane. The input parameter \mathbf{x} represents the data items in Fig. 1(a), called feature vector. We assume there are N feature vectors in the data set, expressed as $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$. In order to obtain \mathbf{w} and b , the SVM first finds the points closest to the hyperplane (called support vectors represented in Fig. 1(a)), and then verifies that they are the farthest away from the hyperplane (i.e., the margin is maximum). As we can see, the components of the feature vectors indicate the positions in the input space, which means their belonging group is dependent on these components. We assume that the failure of optical network equipment is closely related to the equipment states, which can be quantified by the indicators collected during the equipment runtimes by the network operators. The indicators here represent the physical parameters of the equipment such as the optical power, the laser current, the environmental temperature, the

module internal temperature, the central wavelength, the optical signal to noise ratio (OSNR), the power consumption, and the other parameters that the network operators record in daily operations. In general, these data will be recorded in the network management logs by operators. It is easy to export enough data to use. So, in the optical network equipment failure prediction, we use the value of indicators as the components to compose the feature vector. Each feature vector \mathbf{x}_n ($n = 1, 2, \dots, N$) has a corresponding label $l_n \in \{-1, +1\}$ depending on its group (failure or normal). The data with label $l_n = +1$ will satisfy $y(\mathbf{x}) > 0$ and the data with label $l_n = -1$ will satisfy $y(\mathbf{x}) < 0$. Thus, the perpendicular distance from \mathbf{x}_n to the hyperplane is given by $|\mathbf{w}^T \cdot \mathbf{x}_n + b| / \|\mathbf{w}\| = l_n (\mathbf{w}^T \cdot \mathbf{x}_n + b) / \|\mathbf{w}\|$. The goal of SVM algorithm is to obtain the values of \mathbf{w} and b which maximize the distance described by Eq. (2).

$$\arg \max_{\mathbf{w}, b} \left\{ \frac{2}{\|\mathbf{w}\|} \min_n \left[l_n \cdot (\mathbf{w}^T \cdot \mathbf{x}_n + b) \right] \right\} \quad (2)$$

This problem is complex and difficult to solve directly. A commonly used method is to convert this problem into a more tractable form [14]. In the problem discussed above, if we rescale \mathbf{w} to $k\mathbf{w}$ and b to kb , then the distance from any \mathbf{x}_n to the hyperplane remains unchanged. Thus, we can set $l_n y(\mathbf{x}_n) = 1$ for the feature vectors and the farthest data will become $l_n y(\mathbf{x}_n) > 1$. On the other hand, maximizing $2/\|\mathbf{w}\|$ is equivalent to minimizing $\|\mathbf{w}\|^2/2$. Then, the problem can be converted to an optimization problem with constraints given as follow,

$$\arg \min_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\} \quad (3a)$$

$$s.t. \quad l_n y(\mathbf{x}) \geq 1, \quad n = 1, 2, \dots, N \quad (3b)$$

To solve this kind of constrained optimization problem, a method of Lagrange multipliers is very useful [15]. With the help of Lagrange multipliers $a_n \geq 0$, we can get the dual problem as follows,

$$\arg \max_{\mathbf{w}, b} \left\{ \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N l_i \cdot l_j \cdot a_i \cdot a_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right\}, \quad (4a)$$

$$s.t. \quad a_i \geq 0, \quad i = 1, 2, \dots, N, \quad (4b)$$

$$\sum_{i=1}^N a_i l_i = 0. \quad (4c)$$

So, when we put in all the feature vectors into Eq. (4a-c), we can solve for values of a_i , namely the solution to the dual problem. With this result, we first obtain \mathbf{w} and then choose an $a_j > 0$ to obtain b as follows,

$$\mathbf{w} = \sum_{i=1}^N a_i l_i \mathbf{x}_i \quad (5a)$$

$$b = l_j - \sum_{i=1}^N a_i l_i \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (5b)$$

In this step, these feature vectors are called the training data and the calculation process is called the training process. As we can see, the selection of training data will affect the solution of \mathbf{w} and b directly, then further affect the classification accuracy. If we want to diagnose the equipment failure in a high accuracy, enough failure data and normal data are needed.

In practice, the data groups may overlap, which will result in overfitting and poor generalization if we separate the data exactly. In this case, SVM uses slack variables which allow the training data to be misclassified. A punishment factor C is used to control the balance between the exactness and generalization of the hyperplane. With the introduction of C , the constraint Eq. (4b) is modified as $C \geq a_i \geq 0, i = 1, \dots, N$.

The discussion above is based on the premise that the data groups are linearly separable; however, this is not common in practical problems. To solve a nonlinear problem, a kernel function is introduced to map the complex linearly inseparable data from a low dimension to a higher dimension feature space where the data is linearly separable as shown in Fig. 1(b) and (c). After substitution, we can use the solution to the linear problem to solve the nonlinear problem in higher dimension.

In order to illustrate this function more clearly, we assume that \mathcal{X} is the input space (original space in which the problem is nonlinear) and \mathcal{H} is the feature space (higher dimension than input space). If a mapping $\phi(x): \mathcal{X} \rightarrow \mathcal{H}$ exists, and for all $x, z \in \mathcal{X}$, the function $K(x, z)$ meets the conditions $K(x, z) = \phi(x) \cdot \phi(z)$, then $K(x, z)$ is called the kernel function. We observe that, in the dual problem, both the target function and decision function involve only inner products. Therefore, the $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ in Eq. (4a) and Eq. (5b) can be replaced by $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$. Then, the dual problem is changed into:

$$\arg \max_{\mathbf{w}, b} \left\{ \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N l_i \cdot l_j \cdot a_i \cdot a_j K(\mathbf{x}_i, \mathbf{x}_j) \right\}, \quad (6a)$$

$$s.t. \ C \geq a_i \geq 0, i = 1, 2, \dots, N, \quad (6b)$$

$$\sum_{i=1}^N a_i l_i = 0. \quad (6c)$$

Equation (5b) is changed into:

$$b = l_j - \sum_{i=1}^N a_i l_i K(\mathbf{x}_i, \mathbf{x}_j). \quad (7)$$

This change is equivalent to using the mapping function ϕ to convert the input space into a new feature space; the inner product in the input space is converted into an inner product in a new feature space. In this new feature space, the problem is a linear problem and can be solved using Eq. (4a)-(4c). Hence, if we find an appropriate kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$, the nonlinear problem could be solved by Eq. (6a)-(6c).

As talked above, with punishment factor and kernel function, SVM can handle most binary classification problem. We can regard the judgment of an equipment failure state as a classification problem. That is, if we only intend to evaluate whether the equipment is usable, the problem actually is a binary classification problem. Thus, SVM can implement the intelligent recognition function in this situation.

The SVM algorithm can extract the characteristics of equipment state to judge whether an equipment is failed from multiple indicators intelligently. Compared with the traditional method, the biggest advantage of SVM is that it can build an accuracy mathematical model for fault diagnosis without knowing the inner relation between indicators. This method will identify the fault more accuracy than the traditional method. In traditional method, people use the threshold value to judge whether an equipment is fault, but this method consider each indicator independently. The equipment failure might be related to optical power, temperature and current at the same time, for each of them, it is within the threshold value, but the combine of them reflects the fault characteristics. The model trained by SVM algorithm can diagnose these faults.

As discussed above, SVM algorithm provides an effective method to use indicators to estimate the equipment state. However, it cannot independently perform prediction. As shown in Fig. 2, the historical data are collected from time $t-n$ to time $t-1$, and we set the indicator data at this point to be time t . However, this is not prediction because the equipment state at time t is also known. A reasonable approach is to use another algorithm to predict the indicators at time $t+T$, and then to predict the equipment failure state at time $t+T$.

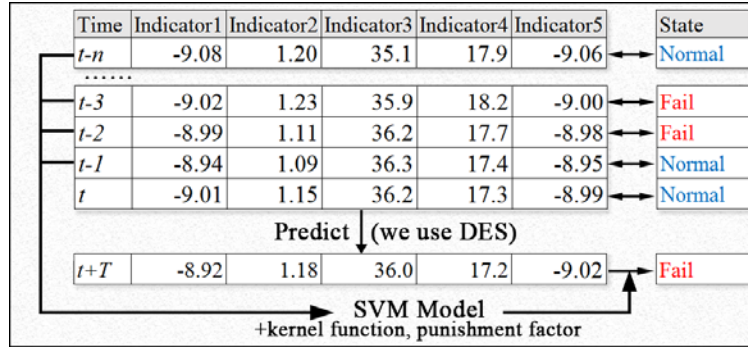


Fig. 2. Combination of the DES and SVM methods.

The indicators we use are measured physical data. Therefore, their changing tendency is a continuous curve and the short-term tendency can be predicted. On the other hand, when a piece of equipment is close to failing, its indicators will have a larger fluctuation. It is thus suitable for the DES algorithm to perform this work. DES is a kind of time-series prediction algorithm that is mainly used in short-time prediction. It is an improved exponential smoothing algorithm that can more sensitively identify the data change, which is suitable for a time series with drastic changes. The main characteristic of the DES algorithm is that it performs exponential smoothing on a single exponential smoothing result, as shown by Eqs. (8a) and (8b) [16]:

$$S_t^{(1)} = ay_t + (1-a)S_{t-1}^{(1)} \quad (8a)$$

$$S_t^{(2)} = aS_t^{(1)} + (1-a)S_{t-1}^{(2)} \quad (8b)$$

where $S_t^{(1)}$ is the exponential smoothing value at time t , $S_t^{(2)}$ denotes the DES value at time t , y_t is the actual value at time t , and $a \in (0,1)$ represents the smoothing constant. $S_t^{(1)}$ and $S_t^{(2)}$ are used to calculate the prediction value \hat{Y}_{t+T} at time $t+T$ by Eqs. (9a) to (9c).

$$\hat{Y}_{t+T} = a_t + b_t \cdot T \quad (9a)$$

$$a_t = 2S_t^{(1)} - S_t^{(2)} \quad (9b)$$

$$b_t = \frac{a}{1-a} (S_t^{(1)} - S_t^{(2)}) \quad (9c)$$

We choose DES to assist SVM in conducting equipment failure prediction. We refer to this combination as the DES-SVM prediction method. In this method, we firstly use DES to predict \hat{Y}_{t+T} at time $t+T$ for each indicator. From that point, we use these prediction values with SVM to judge the equipment failure state at time $t+T$.

B. Procedure of the DES-SVM prediction method

The main application scenario of the proposed method is a software-defined metropolitan area network (SDMAN), as shown in Fig. 3. In the physical plane, WDM nodes constitute a

metropolitan area network in a mesh topology. In the control plane, a central controller collects the operation and maintenance data from all WDM nodes and then analyzes the data before providing instructions.

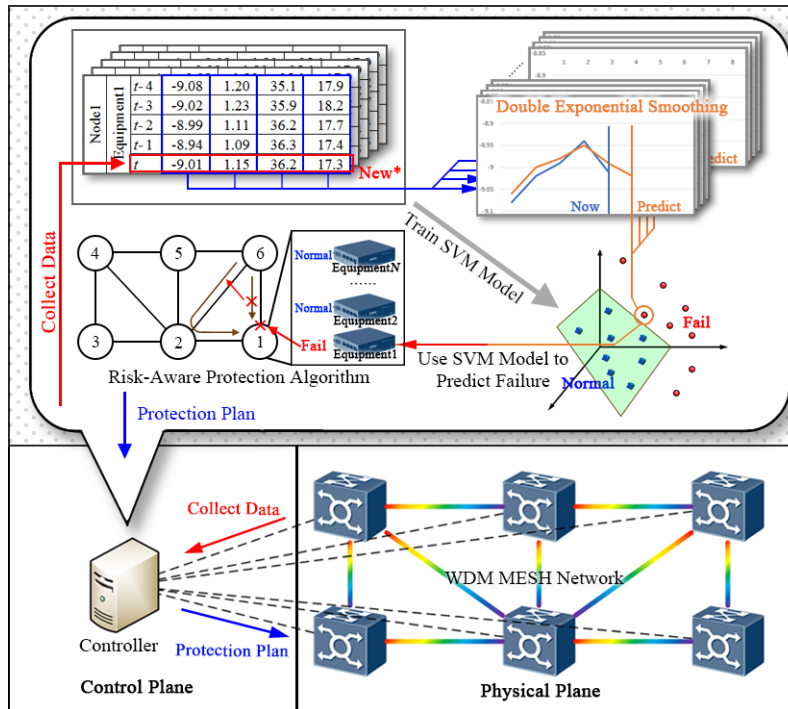


Fig. 3. DES-SVM method predicting an equipment failure and the risk-aware protection algorithm preventing data loss in SDMAN.

The proposed method operates in the controller using the operation data to train the prediction model and predict an equipment failure. If the controller identifies a potential equipment failure by the DES-SVM prediction method, it initiates protection measures in advance to calculate the best approach to protecting the services. Next, the controller sends control messages to all WDM nodes. These nodes switch the services to a safe path to prevent data loss. The main procedure of the DES-SVM prediction method is shown below.

Step A. The controller collects data and selects the indicators. In the SVM algorithm, we use features to form vector \mathbf{x} in each data item. These features are the value of the indicators. Note that each indicator might have multiple records, thus, each indicator might correspond to multiple features. According to the basic principle of SVM, features should be related to the equipment state in order to distinguish between equipment failure and normal. Feature selection directly influences the model accuracy, so it is important to select most related indicators to ensure the accuracy of the final result. As Fig. 4 describes, we can use the classification accuracy trained by a single indicator to get the relationship between this indicator and equipment failure. If an indicator is closely related to the equipment failure, the change of this indicator will affect the equipment failure state obviously.

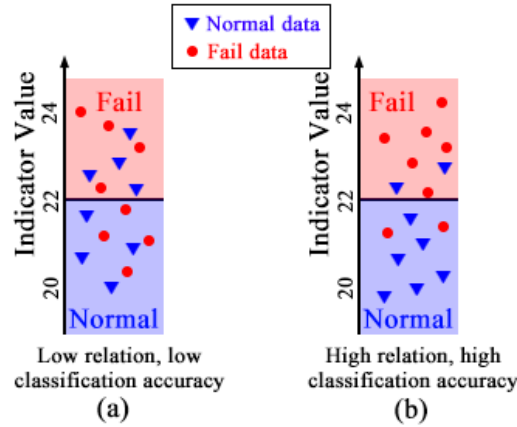


Fig. 4. Classification accuracy according to a single indicator reflex the related degree between this indicator and equipment failure

The present authors suggest using ten-fold cross-validation [17] to test the classification accuracy. The main steps of N -fold cross-validation is as follows:

Algorithm 1: N -fold cross-validation algorithm

Input: data set, N , punishment factor C and kernel function $K(x, z)$.

Output: average classification accuracy \bar{a} .

```

1:   Average divide data set into  $N$  parts as  $d_1, d_2, \dots, d_N$ .
2:   for  $i$  in  $1, 2, \dots, N$  do
3:       Choose  $d_i$  as test data and the rest data as
           training data  $d_i'$ .
4:       Use  $C$  and  $K(x, z)$  on SVM, train  $d_i'$  to get
           model  $M_i$ .
5:       Use  $M_i$  to classify the test data  $d_i$ .
6:       Compare the classify result with the actual
           class, calculate accuracy  $a_i$ .
7:   end for
8:   Calculate the average classification accuracy
       by  $\bar{a} = \sum_{i=1}^N a_i / N$ 

```

The higher is the classification accuracy, the stronger is the relationship between this indicator and an equipment failure, which means they are useful in the model. In this step, we suggest collect indicators as much as possible. For each of them, rule out other indicators and use only this indicator itself to create vector \mathbf{x} , then combine with the corresponding labels l_n , which is the failure state of the equipment to establish an exclusive data set. Do ten-fold cross-validation on this data set to test classification accuracy. Take punishment factor $C = 10$ and radial basis function (RBF) as kernel function in ten-fold cross-validation to ensure the effectiveness (RBF performs well in most problems [18]). The classification accuracy is used as the relation between this indicator and the equipment failure. Then, the collected data are thus screened according to the result.

Step B. Train the best failure diagnostic model. When using the SVM algorithm, two parameters—the kernel function and punishment factor—should be chosen to fit the characteristics of the problem data. The choice of kernel function and the punishment factor directly affect the model accuracy.

Therefore, to train an equipment failure diagnostic model with the highest accuracy, it is necessary to find the most appropriate kernel function and punishment factor. Owing to the uncertain relationship between the features and the equipment failure state, the most effective means of finding the appropriate kernel function and punishment factor is to set different combinations of them and employ ten-fold cross-validation to test the model accuracy (i.e. an

exhaustive method). Some commonly used kernel function are RBF, linear function, polynomial function and sigmoid function, each of them is suitable for a kind of specific problems. Punishment factor C controls the precision of the classification, its value should not be too big or small. If C is too small, the SVM model will lose its precision. If C is too big, the classification model will be too conforming to the training data so that lost its generality. In this step, we suggest try all the kernel function and different punishment factor values on SVM and test the classification accuracy by using ten-fold cross-validation. The kernel function and punishment factor C with the highest accuracy comprise the best choice.

Step C. Predict the indicator values. In this step, we use DES on each feature, thereby establishing and updating the prediction curve according to the latest monitoring data. To this end, we set $S_0^{(1)} = S_0^{(2)} = y_1$, $a = 0.5$, and we use Eq. (8a-b) to calculate $S_t^{(1)}$ and $S_t^{(2)}$ from $t = 1$ to N on each feature. Time $t = N$ is representative of the newest monitoring data. With $S_N^{(1)}$ and $S_N^{(2)}$ on each feature, we can predict the feature value \hat{Y}_{N+1} at time $t = N + 1$ by Eq. (9a-c).

Step D. Predict whether the equipment will fail. We obtain the equipment failure diagnosis model at Steps A and B. Then, we obtain the prediction value \hat{Y}_{N+1} of each feature in Step C. In this step, we use \hat{Y}_{N+1} of these features in the SVM model to predict the equipment failure state at time $t = N + 1$, which is the result of our prediction method.

3. Experimental setup and results

To verify the feasibility of our scheme, we collected real data in the WDM network from a telecommunications operator. Because we were proving the feasibility of our proposed method, large amount of failure data is needed. On the other hand, if the amount of normal boards is much larger than the failure boards, the prediction accuracy would be too high and could not tell the real effect of our method. So, we screened out most of the redundant normal boards and focusing on the boards that easy to fail. As a result, the number of normal data and fail data were balanced. After the screening, the operation state information of 320 boards from 18 nodes in the actual WDM network were recorded. These boards are access and convergence boards of the same type. The observation period was 44 days. The time unit was a day; hence, there were 14,080 data items in total.

Data in the first 35 days were used to create the SVM model by Steps A and B. In step C, we obtained the prediction value \hat{Y}_{N+1} for each indicator in each board for each day. The prediction values in the last nine days were used in Step D to judge the failure state. The indicators in each data sample are shown in Table 1.

Table 1. Indicators In Board Data

Indicator Name	Units
Input Optical Power (IOP)	dBm
Laser Bias Current (LBC)	mA
Laser Temperature Offset (LTO)	°C
Output Optical Power (OOP)	dBm
Environmental Temperature (ET)	°C
Unusable Time	s

In these indicators, “unusable time” represents the failure state of the equipment. We assumed a board as “failed” when the value of “unusable time” was larger than 40,000. The other five indicators were used to diagnose the board failure. Each of these included the maximum, minimum, and average values of the given day. Therefore, there were 15 features in each data sample. According to the main principle of SVM, each data sample is composed of characteristic vector \mathbf{x} and classify label l . $\mathbf{x} = (x_1, x_2, \dots, x_{15})$, x_1 - x_3 corresponds to the three values (i.e. the maximum, minimum, and average value of the given day) of input optical power, x_4 - x_6 corresponds to the three values of laser bias current and so on. For each \mathbf{x}_i , its corresponding label $l_i = -1$ if the unusable time $t_i < 40000$, otherwise $l_i = +1$. These data

would be further screening in the Step A, and would be used to select kernel function and punishment factor in Step B.

We first calculated the relation between the indicators and board failure by Step A, as shown in Fig. 5. It was determined that the laser bias current predominantly related to the board failure (approximately 83.6%). Environmental temperature data were next (about 74.48%). These data suggested that, when the board was close to failing, it was first reflected in the laser bias current. Moreover, environmental temperature showed a strong relationship with board failure.

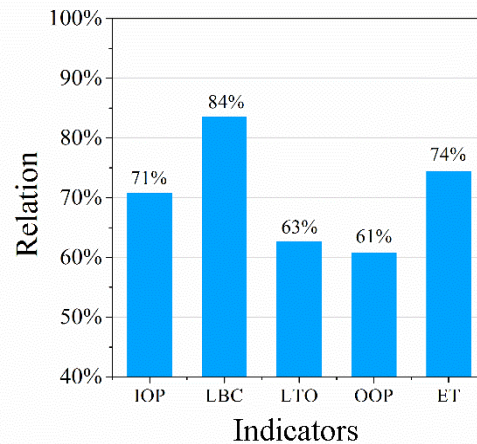


Fig. 5. Relation between indicators and board failure.

We did the following experiment to find out the relation between the number of indicators and the SVM fault diagnosis efficiency, based on the result in Fig. 5. First, we sorted all these indicators according to the relation degree from high to low. The first is LBC, ET comes next, then IOP, then LTO, OOP is the last. Then, we selected and added the number of indicators in accordance with the order and used them to build specific data sets. For one indicator, we selected the first indicator from the sorted list (i.e. LBC), for two indicators, we selected first two from the sorted list (i.e. LBC and ET), and so on. Finally, we used ten-fold cross-validation to test model accuracy for each data set. Besides, we did a same experiment according to the reverse order to select indicators, the results are shown in Fig. 6.

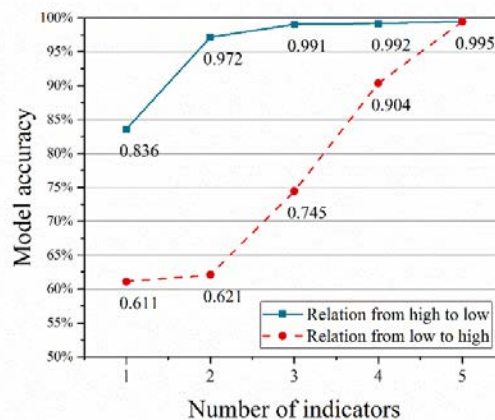


Fig. 6. The affection of number of indicators to model accuracy.

As we can see, with the number of indicator increases, the model accuracy improves obviously. Combine two indicators that most related to the board failure can improve the fault diagnosis accuracy (i.e. model accuracy or classification accuracy) to 97.2%. And with the increase of indicator number, model accuracy further improves. On the other hand, we can see that the combine of indicators that have low relationship to the board failure does not have a huge improvement on model accuracy. The introduction of these indicators can improve accuracy, but they are not the main basis to diagnose board failure.

The results indicate that these indicators reflect the state of the equipment from different aspects, its number increase can make the fault diagnosis more accurate. However, some indicators have low relationship to the board failure and they could not improve model accuracy obviously. When the computational burden is high or model accuracy is high enough, these indicators can be abandoned.

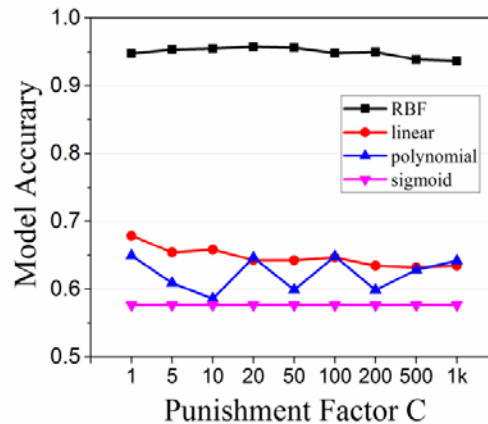


Fig. 7. SVM model accuracy with different kernel function and punishment factor C .

Figure 7 depicts the SVM model accuracy (i.e. classification accuracy calculated by ten-fold cross validation) when using different kernel functions and punishment factor C in Step B. It is evident that the RBF kernel has the best accuracy among these commonly used kernel functions. This is because the RBF kernel can map the data to a high dimension in which the data items are easy to classify. It is thus suitable for complex nonlinear problems, which conforms to the nonlinear relationship between the indicators and board failure. Furthermore, it is evident in Fig. 5 that punishment factor C has a minimal effect on the model accuracy: it arrives at the top at $C = 20$ in the RBF kernel. This phenomenon shows that most of the data can be separated with the board state; only a small amount of data is mixed together on the border. In sum, the following experiments employ RBF kernel with punishment factor $C = 20$ to train SVM fault diagnosis model.

Figure 8 depicts the result of Step C, thereby showing the DES prediction accuracy. These data comprise minimum input optical power from one board in 44 days. As the figure shows, when the indicator smoothly changes, the prediction result has a high accuracy. However, when the data trend changes, it suffers a brief period of low accuracy. The main cause of this result is that the number of data points is not adequate and the algorithm cannot smoothly match the change. Nevertheless, DES still shows high prediction accuracy.

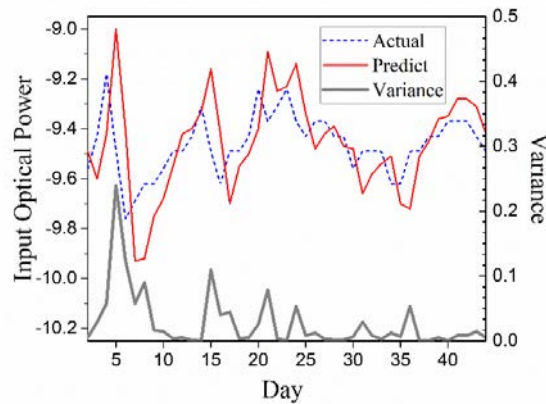


Fig. 8. Discrepancy between the actual and predicted values when using DES.

Furthermore, we employed the data in the first 35 days to train the model. The RBF kernel was selected, and punishment factor C was 20, as in the discussion above. From Day 35 to 44, we predicted the board failure of the next day in each day compared with the actual data. The result is shown in Fig. 9. Because half of the data were collected from boards that had faults in the observation period, the number of actual fault boards was maintained at approximately 180. With our DES-SVM prediction method, most of these fault boards were correctly predicted; the prediction accuracy was greater than 90%.

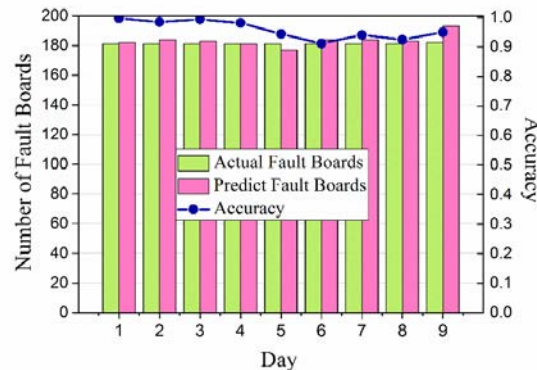


Fig. 9. Number of predicted fault boards compared with actual fault boards and prediction accuracy of DES-SVM.

According to the experiment data, the accuracy of the DES-SVM prediction method was 95.59% on average, which meant the failure state of 95% of the boards could be correctly predicted. Moreover, we selected for prediction approximately ten boards that usually incurred failure. Although their indicators suffered very large fluctuations, the prediction accuracy reached 86.37%. These results show that our method most often correctly predicted the equipment state.

In order to find out the affection of the training data size, we take first 5, 10, 15, ..., 35 days of data as the training data set respectively. DES algorithm is used and predict the data forward one day in each day. The prediction result of the last 9 days is used to test the accuracy, compared with the real data in last 9 days. The result is shown in Fig. 10.

As we can see, when we use the board data in 5 days as training data, the equipment failure prediction accuracy reached to 92%. In first 5 days, the number of training data is 1600 in total, which means we can get a good result by using less than two thousand data items on training. On the other hand, with the number increase of training days, prediction

accuracy is on the rise as a whole. Besides, in the last 9 days, there are 1784 fault board data items in total and we can see that most of these board failures could be predicted successfully. With the number increase of training days, more and more failure could be predicted successfully, which suggest that the failure prediction model becomes more accurate and effective. This is because when the number of training days increases, the data set will be more abundant so that the failure predication will be more accurate. This result proved that with the increase of the training data size, the trained SVM model become more precision, and the failure prediction accuracy become higher. So, under the premise that the controller is not overload, we can increase the training data size as much as possible to enhance the accuracy of the model.

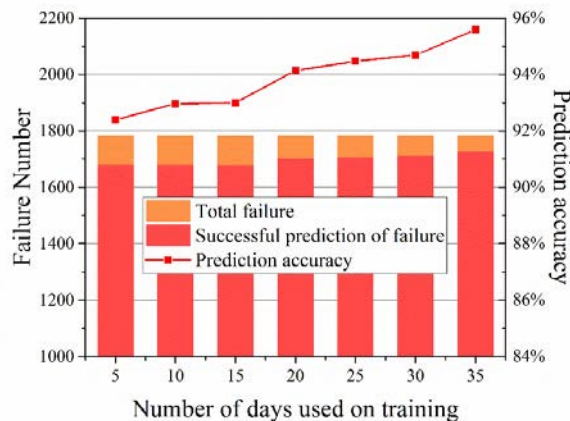


Fig. 10. Prediction result based on different number of training days.

According to the experimental results, the DES-SVM prediction method offers the following key benefits. It can predict the board failure in a WDM network, which means services can be protected from data loss before a network failure occurs. In addition, SVM and DES are not complex algorithms. Thus, the proposed combined method is easy to establish in the controller system.

Nonetheless, owing to the poor reaction of DES during the period when the data trend notably changed, our method could not precisely predict the inflection point. If the data were more refined (e.g., data record hourly, not per day), a more effective means of predicting the indicator values could be identified.

4. Conclusion

In this paper, we proposed the DES-SVM prediction method to predict equipment failure in an optical network based on SVM and DES. Experimental results demonstrated that the DES-SVM prediction method had a high accuracy (95%) in optical equipment failure prediction. Therefore, it is possible to back up data and switch services to safe links before a network failure occurs. It can thereby protect the network services from suffering a data loss while improving the user QoE.

Funding

National Natural Science Foundation of China (NSFC) (61372119).

Acknowledgments

We gratefully acknowledge China Mobile Communications Corporation (CMCC) for permission to use their operation data in this work.