

Research Article

A Novel Power System Reliability Predicting Model Based on PCA and RVM

Yuping Zheng,¹ Guoqiang Sun,^{1,2} Zhinong Wei,² Feifei Zhao,² and Yonghui Sun²

¹ NARI Technology Development Co., Ltd., Nanjing 210061, China

² College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China

Correspondence should be addressed to Yonghui Sun; sunyonghui168@gmail.com

Received 9 January 2013; Accepted 4 February 2013

Academic Editor: Yang Tang

Copyright © 2013 Yuping Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The power system reliability is an important index to evaluate the ability of power supply. According to the characteristics of the practical grid operation, this paper trains and sets up power grid reliability predicting model, based on relevance vector machine, taking the load supplying capacity of power grid and natural calamities as input variables, and the outage time of power grid failure affecting the reliability of the power supply as output variables. In the modeling process, through principal component analysis of the training sample set of relevance vector machine, the input factor number of sample is improved, the input number of network is reduced, the network structure is simplified, and the predicting accuracy is increased. Simulation results are provided to verify the effectiveness of the proposed algorithm, which show that it provides a new way for power system reliability predicting.

1. Introduction

The power system reliability is the ability of power system to provide continuous power supply for all the consumers. With the development of society and the increasing of people's living level, the requirement for power system reliability is becoming more and more important in recent years. The higher reliability of power system is not only the need of the consumers but also of importance for the development of power companies. Therefore, in the past few decades, the power system reliability has attracted an increasing attention by lots of researchers inland or abroad; several different ways have been proposed to improve the power system reliability by analyzing affecting factors [1].

The traditional predicting assessment methods for power system reliability require the historical data with accurate network structure and the reliable elements [2]. In [3], the authors applied the Monte Carlo method in the assessment of the power system reliability, where the calculation structure was simple, but the calculation error was inversely proportional to the square root of experiment frequency, which reduced errors at the expense of computing time. In [4], a grid reliability algorithm based on the sensitivity

analysis was proposed, in which a more important key element information was obtained through the sensitivity analysis of element reliability parameters. In [5], a power grid reliability evaluation model based on radical basis function (RBF) neural network has been proposed, where the states of power system could be firstly classified through the adaptive algorithm of RBF neural network, then the speed of identifying grid failure state could be improved by computing the power grid reliability. In [6], the BP neural network was used to predict the power supply reliability of a city grid; however, due to the less influential factors, the predicting performance was not good and the speed was slow, and sometimes it even fall into local extremism. The authors in [7] proposed a power system reliability predicting assessment algorithm based on the actual operation constraint, which used the actual power grid operation parameters to predict the reliability index of planned distribution network, so that the reliability index could be decomposed and the reliability management could be arranged. In [8], the Markov cut-set algorithm was employed to evaluate power system reliability.

However, the grid structure is too complex and the amount of data is too huge, and it is often difficult to determine the structure, so the traditional predicting methods

cannot be used to predict the power system reliability. On the other hand, in practical applications, if all the influencing factors are taken as the inputs into the sample set, it will inevitably lead to dimension expanding of the sample set and then affecting the predicting accuracy and the generalization capability.

Based on the above discussion, this paper presents power system reliability predicting model based on relevance vector machine (RVM); the principle component analysis (PCA) method is used to extract sample sets, and then it is taken into the RVM model after eliminating the correlation between variables. The proposed approach not only considers the feature extracting ability of PCA but also takes advantage of the nonlinear approximating ability of RVM, and thus the predicting accuracy of the predicting model and the generalization ability can be improved greatly.

The rest of this paper is organized as follows. In Sections 2 and 3, the theory of PCA and the principle of RVM are introduced, respectively. In Section 4, the reliability predicting model is proposed and the corresponding algorithm is presented. In Section 4, an illustrative example about a regional grid is provided to illustrate the effectiveness of the developed results. At last, this paper completes with a conclusion.

2. The Theory of PCA

In this section, the theory of PCA is briefly introduced. It firstly calculates the correlation matrix of the data matrix derived from many sample data, then gets the accumulated variance contribution according to the eigenvalues of the correlation matrix, and finally determines the principal components from the eigenvectors of the correlation matrix. The main results can be divided into the following five steps [9–11].

Step 1 (standardization of the original data). The main purpose is to eliminate the effects of different dimensions of the original variables and large numerical difference; that is,

$$x_j^* = \frac{x_j - E(x_j)}{\sqrt{\text{Var}(x_j)}}, \quad (1)$$

where x_j is the j th column of the original data $\mathbf{X}_{n \times p}$, n is the class number of the sample data, p is the number of evaluation indexes, and $E(x_j)$ and $\text{Var}(x_j)$ stand for the mean and covariance of x_j , respectively.

Step 2. Calculation of the correlation coefficient matrix $\mathbf{R} = (r_{ij})_{p \times p}$, where r_{ij} ($i, j = 1, 2, \dots, p$) is the correlation coefficient of variables x_i and x_j , $r_{ij} = r_{ji}$, and the formula is given by

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}. \quad (2)$$

Step 3 (calculation of the eigenvalues and eigenvectors). Firstly deriving the characteristic values λ_i ($i = 1, 2, \dots, p$) by solving the characteristic equation $|\lambda I - R| = 0$, then arranging them in order of size, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, and finally calculating the eigenvectors e_i ($i = 1, 2, \dots, p$) corresponding to the eigenvalues λ_i .

Step 4. Calculation of the principal component contribution rate and cumulative contribution rate. The principal component contribution rate (%) is derived by

$$\eta_i = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100\%, \quad (3)$$

and the cumulative contribution rate (%) is derived by

$$\eta_\Sigma(k) = \sum_{i=1}^k \eta_i. \quad (4)$$

The number of principle component is selected depending on the cumulative contribution, the cumulative contribution rate is usually set more than 85%–90%, and then the corresponding first k main ingredients comprise most of the information provided by the p original variables. The number of principle component is k .

Step 5. Calculation of the principle component load by

$$l_{ij} = \sqrt{\lambda_j} e_{ij} \quad (i = 1, 2, \dots, p, j = 1, 2, \dots, k). \quad (5)$$

Thus the principle component score is further obtained by

$$Z = (z_{ij})_{n \times k}. \quad (6)$$

3. The Principle of RVM

RVM is a learning algorithm derived from the theory of Bayesian learning algorithm, which is based on the support vector machine [12–14]. This algorithm combines several theories, such as the Markov chain, the Bayesian theorem, automatic related decision prior, and the maximum likelihood, and thus it has the following advantages: (1) the high sparsity; (2) the shorter training time due to only the setting of the kernel function; (3) the flexible choosing of the the kernel function for it does not need to satisfy the Mercer condition.

For a given training sample set $\{x_n\}_{n=1}^N$, the output set is defined as $\{t_n\}_{n=1}^N$, and the RVM regression model can be obtained by

$$t_i = \sum_{i=1}^N w_i K(x, x_i) + w_0 + \epsilon, \quad (7)$$

where $\epsilon \sim N(0, \sigma^2)$ is the independent sample error, w_i ($i = 0, 1, \dots, N$) is the weight coefficient, $K(x, x_i)$ is kernel function, and N is the sample quantity.

For the independent output set, the likelihood function of the whole sample is

$$p(t | w, \sigma^2) = \prod_{i=1}^N \widehat{N}(t_i | y(x_i; w), \sigma^2) \quad (8)$$

$$= (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{\|t - \Phi(x)w\|^2}{2\sigma^2}\right),$$

where $t = (t_1, t_2, \dots, t_N)$, $w = [w_0, w_1, \dots, w_N]^T$, $\Phi = [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]^T$, $\phi(x_N) = [1, K(x_1, x_N), k(x_2, x_N), \dots, K(x_n, x_N)]^T$, and $\widehat{N}(t_i | y(x_i; w), \sigma^2)$ is the Gaussian distribution function.

When employing the maximum likelihood method to solve w and σ^2 directly, it usually results in serious over-fitting problem, and in order to avoid this phenomenon, w is set as the zero-mean Gaussian prior distribution by using the sparse Bayesian principle

$$p(w | \alpha) = \prod_{i=0}^N \widehat{N}(w_i | 0, \alpha_i^{-1}), \quad (9)$$

where α is the corresponding hyper parameter of weight w ; if every weight is corresponding to one hyper parameter, then it can control the influences of the prior distribution on each parameter, and thus the sparse characteristics of the RVM can be realized.

After defining the prior probability distribution and the likelihood distribution, according to the Bayesian theorem, the posterior probability distribution of all the unknown parameters can be derived by

$$p(w | t, \alpha, \sigma^2) = (2\pi)^{-(N+1)/2} |\Psi|^{-1/2} \exp\left\{-\frac{1}{2}(w - \mu)^T \Psi (w - \mu)\right\}, \quad (10)$$

where the posterior covariance matrix is $\Psi = (\sigma^{-2}\Phi^T\Phi + A)^{-1}$, $\mu = \sigma^{-2}\Psi\Phi^T t$, and $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$.

Finally, the maximum likelihood method is used to estimate the hyper parameter α and variance σ^2 . Given the new input sample x , the corresponding output probability distribution will follow Gaussian distribution, and the corresponding predicting value is derived by $y = \mu^T \phi(x)$.

4. The Reliability Predicting Assessment

4.1. Selection of the Input Variables. It is well known that power system reliability highly depend on the grid supply capacity and the natural environment of the grid; therefore, in this paper, some main factors reflecting supply capability of the grid itself, such as the available efficient of the grid equipment, the ratio of the power system reliability, power supply radius, capacity-load ratio, the loss of electricity as for power rationing, the unit of the new substation capacity, and high temperature, lightning strikes, strong winds, and heavy

rain which has a larger impact on the safe operation of power system, are all taken as the input of the model. The available efficient of the grid equipment expresses the ratio of the available hours of major equipment to hours during survey period. The ratio of the power system reliability stands for the ratio of the power supply total hours to hours during survey period. Power supply radius means the physical distances of power supply circuit between power point and the farthest point of the power supply. Capacity-load ratio expresses the ratio of the city substation capacity to the corresponding load on the basis of which meet with the power supply reliability. The loss of electricity as for power rationing shows the power supply loss value caused by power rationing during survey period. The unit new substation capacity illustrates the corresponding values to the new substation capacity of different voltage levels when increasing the unit load during survey period.

4.2. Selection of the Output Variables. In the daily production, there are two main aspects affecting the power system reliability: one is the fault power outages, and the other is the preliminary arrangement power. It is known that grid system is made up of a large number of transmission lines, transformers, switch equipment, sorts of reactive power compensation equipment, and power points. The final direct representation is mainly the grid blackout due to these factors on the grid. Therefore, in this paper, the grid average fault power failure time is taken as the output of the power reliability predicting model.

4.3. RVM Modeling and Structure Model. The basic idea of the hybrid kernel function is briefly introduced in advance; firstly, the principal component analysis is performed on the original input variables, then adopting RVM to model the new matrix, and finally, we get the result of predicting. The selection of kernel function of RVM does not need to satisfy Mercer condition; therefore, there have a certain degree of freedom in the selection of the kernel function, and integrating the different characteristics of the kernel function, we will get the nuclear function which has better properties.

In this paper, the linear combination of the RBF kernel function and polynomial function is chosen as the kernel function of RVM, which can be derived by

$$K(x_i, x_j) = \lambda G(x_i, x_j) + (1 - \lambda) P(x_i, x_j),$$

$$G(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right), \quad (11)$$

$$P(x_i, x_j) = [(x_i \cdot x_j) + 1]^2,$$

where $G(x_i, x_j)$ is the radial basis function, $P(x_i, x_j)$ is the binomial kernel function, and λ is the weight of kernel function. It is a single kernel function if $\lambda = 0$ or $\lambda = 1$. σ is the width of kernel function. λ and σ are the parameters to be optimized, where the grid search algorithm is employed.

Figure 1 is the structure diagram of the power supply reliability predicting model. The unit load new substation

TABLE 1: The Eigenvalue and principal component contribution.

Principal component	Eigenvalue	Contribution (%)	Cumulative contribution (%)
1	9.458 8	40.90	40.90
2	7.216 5	26.25	67.15
3	5.795 5	14.37	81.52
4	3.551 6	9.66	91.18
5	1.254 6	7.06	98.24
6	0.753 0	1.76	100.00
7	0	0	100.00

TABLE 2: Principal component.

Number	Category	Component				
		Z_1	Z_2	Z_3	Z_4	Z_5
1	Overhead available coefficient of 220 kV	0.100 7	-0.606 6	-0.074 7	-0.774 2	-0.037 0
2	Overhead available coefficient of 110 kV	0.166 4	0.709 5	0.571 8	-0.068 7	0.326 8
3	Transformer available coefficient of 220 kV	-0.573 5	0.494 5	0.602 5	0.150 2	0.180 8
4	Transformer available coefficient of 110 kV	-0.886 7	-0.011 0	0.444 6	-0.088 3	-0.089 2
5	Circuit breaker available coefficient of 220 kV	-0.876 3	0.137 1	-0.437 6	0.139 8	-0.020 6
6	Circuit breaker available coefficient of 110 kV	0.125 5	0.760 8	-0.606 0	-0.080 0	-0.160 5
7	Capacity-load ratio of 220 kV	0.807 5	0.017 5	-0.332 3	0.331 2	0.303 7
8	Capacity-load ratio of 110 kV	0.871 4	0.370 4	-0.050 7	0.271 6	0.158 5
9	Supply radius of 220 kV	0.980 8	0.026 6	0.165 9	0.084 0	-0.040 2
10	Supply radius of 110 kV	-0.341 3	0.508 3	-0.554 2	0.200 5	0.504 2
11	Substation capacity of 220 kV supply load per unit	-0.612 8	0.521 2	-0.532 1	0.104 8	0.173 5
12	Substation capacity of 110 kV supply load per unit	0.323 5	0.763 6	0.149 8	0.374 7	-0.369 5
13	Rural power supply stability	-0.966 7	-0.170 4	0.016 7	-0.142 2	-0.007 6
14	City power supply stability	-0.978 6	-0.151 2	-0.019 2	-0.120 9	0.012 7
15	The loss of switch out limit electric power	0.024 4	0.620 0	-0.027 2	0.237 3	-0.746 1
16	Lightning days	-0.324 9	-0.426 4	0.442 9	0.697 4	0.088 4
17	Rainstorm days	0.284 7	-0.839 5	-0.445 8	0.095 9	0.078 1
18	Storm days	0.340 6	-0.826 3	0.174 2	-0.190 1	-0.146 4
19	High temperature days	0.828 9	0.066 5	0.312 5	-0.413 3	0.186 0

capacity, power supply radius, and capacity-load ratio are divided into two levels of 220 kV and 110 kV. The rate of power supply reliability is divided into rural and urban. The available coefficients of the equipment are divided into 220 kV and 110 kV overhead line available coefficient, 220 kV and 110 kV transformer available coefficient, and 220 kV and 110 kV breaker available coefficient.

5. Numerical Analysis

In this section, a regional grid is taken as an example to illustrate the effectiveness of the proposed algorithm. This region mainly relies on the 500 kV and 220 kV grid, the transformer substations are mainly 220 kV and 110 kV, and there are more 220 kV and 110 kV lines. The average fault power failure time from year 1998 to year 2008 and 19 influence factors of historical data are taken as the training sample, and the actual data of year 2009 (true value is 12.13465 per house) is taken as the testing sample. The power supply

reliability predicting model of this region can be established in the MATLAB environment.

By using the PCA discussed in Section 3, for the original input sample $X_{12 \times 19}$, the eigenvalues and principal component contribution of principal component can be derived, and the results are provided in Table 1. It follows from Table 1 that the cumulative contribution rate of the first five principal components already achieves 98.24%, which is larger than 95%; that is, the first five principal components can provide sufficient information, so these five irrelevant new variables can replace the original nineteen variables. Firstly, we get the eigenvectors of the five new principal components and then calculate the load of variables in the principal component, and the calculation results are given in Table 2.

Finally, based on the five principle components, the RVM approach is used to predict the new sample space, and some predicting results are shown in Table 3. It can be seen from Table 3 that the predicting accuracy of RVM method is better than that used by the BP neural network and has a shorter training time. The new approach proposed combines the

TABLE 3: Prediction result analysis.

Predicting method	Predict values/h/household	Relative error	Variable dimension	Training time (s)
BP	12.983 16	0.069 9	19	4.12
RVM	12.742 68	0.050 1	19	2.43
PCA + RVM	12.413 04	0.022 9	5	1.24

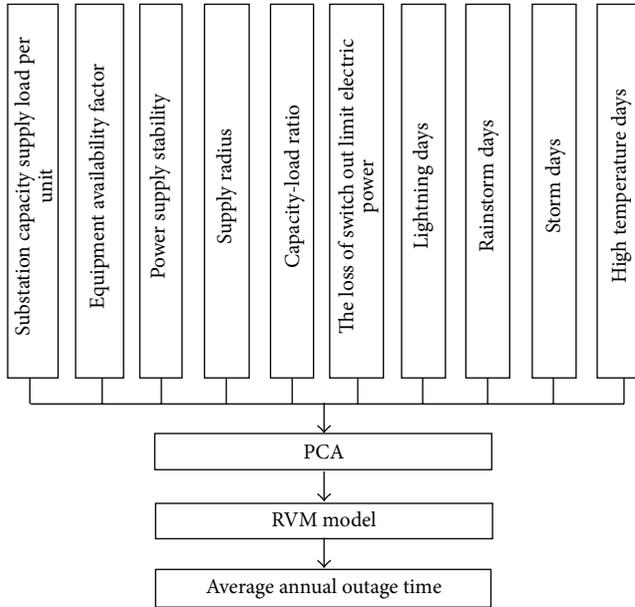


FIGURE 1: Schematic diagram of power supply reliability structure based on PCA and RVM.

RVM method and the PCA method together, which does not change the structure of the sample data but can reduce the number of input variables and simplify the network structure, thus the learning rate and performance of the network can be improved due to the elimination of the correlation of the network input factors; therefore, the predicting performance can be improved greatly and can get a shorter training time.

6. Conclusions

In this paper, a novel power system reliability predicting model based on PCA and RVM has been proposed. By using the PCA method, a variety of factors affecting the reliability of power supply were analyzed, and the new derived matrix has been taken as the input of RVM for training and predicting. This lower-dimensional matrix could eliminate the redundant information and reduce the number of the dimensions of the sample space, thus the predicting accuracy of the network could be greatly improved. At last, simulation results on a regional grid were provided to show the usefulness and effectiveness of the proposed predicting approach. The proposed approach not only provides a scientific reference for improving power system reliability but also provides an effective method for the grid investment modeling.

Furthermore, for testing the robustness of the proposed method, it would be better to consider the impact of noise on prediction for future research [15–18].

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 51277052, 51107032, and 61104045 and in part by the Fundamental Research Funds for the Central Universities of China under Grant 2012B03514.

References

- [1] Y. M. Liu and J. Liang, “Influence of planned interruption on power supply reliability of distribution system,” *North China Electric Power*, vol. 2, pp. 52–542, 2011.
- [2] Y. T. Song, Y. J. Guo, and L. Cheng, “Statistical analysis of reliability data for power system units,” *Relay*, vol. 30, no. 7, pp. 14–16, 2002.
- [3] Z. Bie and X. Wang, “The application of Monte Carlo method to reliability evaluation of power systems,” *Automation of Electric Power Systems*, vol. 21, no. 6, pp. 68–73, 1997.
- [4] K. G. Wu and Z. F. Wu, “Reliability evaluation algorithm of electrical power systems using sensitivity analysis,” *Proceedings of the CSEE*, vol. 23, no. 4, pp. 53–56, 2003.
- [5] K. G. Wu, S. Wang, A. B. Zhang, and J. Q. Zhou, “The study on reliability assessment of electrical power systems using RBF neural network,” *Proceedings of the CSEE*, vol. 20, no. 6, pp. 9–12, 2000.
- [6] Y. T. Song, J.-L. Wu, D. Peng et al., “A BP neural network based method to predict supply reliability of urban power network,” *Power System Technology*, vol. 10, no. 32, pp. 56–59, 2008.
- [7] Q. Gu, A. Luo, J. Wang, and M.-J. Qi, “A practicable algorithm to forecast and evaluate reliability of power supply,” *Power System Technology*, vol. 12, no. 27, pp. 76–79, 2003.
- [8] Y. Liu and C. Singh, “Reliability evaluation of composite power systems using markov cut-set method,” *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 777–785, 2010.
- [9] Z. P. Wang and X. Wang, “Inrush current recognition in power transformer based on modified principal component analysis,” *Power System Protection and Control*, vol. 39, no. 22, pp. 1–6, 2011.
- [10] B. Liu and R. Yang, “Short-term load forecasting model based on LS-SVM with PCA,” *Electric Power Automation Equipment*, vol. 28, no. 11, pp. 13–17, 2008.
- [11] J. C. Li, D. X. Niu, J. Y. Li, and Z. D. Wang, “Evaluation of the power grid development coordination based on the GRA and PCA method,” *Power System Protection and Control*, vol. 38, no. 18, pp. 49–53, 2010.
- [12] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.

- [13] Z. G. Sun, W.-X. Zhai, W.-L. Li, and Z.-N. Wei, "Short-term load forecasting based on EMD and RVM," *Proceedings of the CSU-EPSCA*, vol. 23, no. 1, pp. 92–97, 2011.
- [14] G. Y. Lü, C. M. Wu, W. M. Sun, X. D. Wang, and X. S. Cai, "Detection and identification of voltage sags based on RVM and ANF," *Power System Protection and Control*, vol. 39, no. 12, pp. 65–68, 2011.
- [15] E. Tian, D. Yue, and C. Peng, "Quantized output feedback control for networked control systems," *Information Sciences*, vol. 178, no. 12, pp. 2734–2749, 2008.
- [16] E. Tian, D. Yue, and C. Peng, "Reliable control for networked control systems with probabilistic actuator fault and random delays," *Journal of the Franklin Institute*, vol. 347, no. 10, pp. 1907–1926, 2010.
- [17] Y. Tang, W. Zou, J. Q. Lu, and J. Kurth, "Stochastic resonance in an ensemble of bistable systems under stable distribution noises and heterogeneous coupling," *Physical Review E*, vol. 85, Article ID 046207, 2012.
- [18] Y. Tang, H. J. Gao, W. Zou, and J. Kurths, "Distributed synchronization in networks of agent systems with nonlinearities and random switchings," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, no. 1, pp. 358–370, 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

