# The Eco-Environmental Fragility Assessment of Ningdong Base

Feng-jun LI[1,a,*], Min HU[1,b] and Xiao-le GUO[1,c]

[1]School of Mathematics and Statistics, Ningxia University, Yinchuan *750021,* China

[a]muzi152148@163.com, [b]18295087693@163.com, [c]501875738@qq.com

*Corresponding author

*Abstract*—**We select the Ningdong Energy and Chemistry Industry Base (referred to as Ningdong Base) as research object and take the ecological environment as evaluating target. According to the subject and object factors, the SVM model of ecological environment vulnerability assessment is constructed by using principal component analysis method. We embed GA in SVM to optimize its structure and parameters. SVM model is simple, universal, and accurate, and it can be applied in evaluation of eco-environmental vulnerability in practice. Also, we compare our evaluation results with the the BP-ANN model's results. Then, we draw a conclusion that the ecological environment vulnerability of the Ningdong Base is the second level, which belongs to moderate vulnerability. Finally, the causes of vulnerability formation are analyzed.**

*Keywords- eco-environmental fragility; assessment; SVM; Ningdong Base*

## I. INTRODUCTION

The vulnerability assessment of ecological environment has received more and more attention in the energy sources and ecological environment fields. The weighted mapping relation need to be established between the main influence factors and the evaluation results in most of the existing methods[1-4]. The correlation studies[5-8] worked on grade evaluation of ecological environment by using fuzzy mathematics, gray theory, object-oriented theory and neural network methods. However, these methods are flawed, for example, low evaluation precision, especially it is difficult to determine a suitable network structure, and other drawbacks such as over learning and less learning problems, and local minimum problems[8]. Support vector machine (called "SVM") is a new general machine learning method, which based on VC dimension theory and structural risk minimization principle in statistical theory. SVM has advantages of being simple and the best generalization compared with the existing artificial neural networks (called "ANN"). It also has great ability to avoid the local extreme, and exhibits many unique advantages in many applications, such as nonlinear problems, small sample problems, and high dimension pattern recognition problems[8]. In this paper, the optimized SVM by genetic algorithm (called "GA") and error back propagation algorithm (called "BP") are applied respectively to assess the ecological environment fragility in Ningdong Energy and Chemical Base (called "Ningdong base"). In the end the paper, we collect the evaluation results and analyze the evaluation conclusion.

## II. GENERAL SITUATIONS OF STUDY REGION

The Ningdong Energy and Chemistry Industry Base (referred to as Ningdong Base), in the east-central of Ningxia, situated in the south-west fringe of Mu Us sandy land. Hills and mountainous regions are major components of the area. The geological structure is complex, mainly includes the migratory dune, semi-fixed dune and flat sandy land. Its terrain is obviously higher in the West and lower in the North, which has an average elevation of 1330 meters. Ningdong base is the one of important production base of coal industry in our country, which has long history that the coals exploited. The base is named as "coal with good quality" with huge reserve and perfect coal quality, mostly with limestone and dolostone. The whole study area is 3484km2, it covers Lingwu city, Yanchi county, Tongxin county, Hongsipu development zone etc. The study region belongs to arid temperate climate zone, which has a typical continental climate such as drying, less rain, windiness, cold, sandy, long sunshine hours and so on. The main soil resources in the study area are Sierozem, aeolian sandy soil, and a small amount of saline soil. The water-saving capability of soil is an important factor influencing vegetable growth because the soil layer is too thin and the concentration of soluble salts topsoil contained in topsoil is very high. The region is short of soil organic matter for the average value is less than 1%. Vegetation coverage of the planning area is 10%-30%, the distribution of natural vegetation is scarce and uneven, and the typical species of vegetation mainly include annual plants, xeric shrubs, subshrub and herbaceous plants. Compared with other vegetation, they are characterized by strong resistance to drought, cold, barren and salt alkalescence. There is a severe shortage of water resources in Ningdong base. The seasonal variation and inter-annual variation of precipitation in the study area are significant. On average, its annual precipitation is less than 300 mm in recent years.

Ningdong base is "No. 1 project" of Ningxia Hui Autonomous Region in "Eleventh Five-Year Plan" project. Ningdong base is not only the solid foundation to realize leaping development, but also the fundamental way and hope to shake off poverty, backwardness and keep pace with the nation to realize the grand goal of constructing an affluent society. According to the requirement of "State Council on several opinions related to further promotion of economic and social development in Ningxia", our purpose is to promote development of large coalmining bases and

carry out project to send electricity from west to the east. Meanwhile the base should be further strengthened to ensure the achievement of the circular economy demonstration area. Ningdong base, with harsh natural conditions, the serious degeneration of ecosystem, a fragile ecological environment and severe pollution, is especially vulnerable to climate change. A fundamental problem about it is how to deal with and balance the relationship between ecological and economic.

### III. THE BASIC PRINCIPLE OF SVM CLASSIFICATION

SVM is generated from the optimal sorting line of seeking two types of linearly separable sample sets $(x_i, y_i), i = 1, 2, \cdots, n, x_i \in R^d$, $y_j \in \{1, -1\}$, the basic idea can be illustrated by Fig. 1.
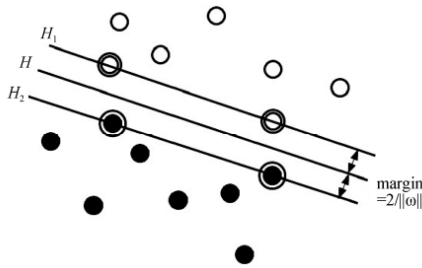


Figure 1. The optimal sorting line under condition of linear and divisible

The classification interval at this time is $2 / \|\omega\|$, which satisfies

$$y_i(\omega \cdot x_i + b) - 1 \geq 0, i = 1, 2, \cdots, n. \qquad (1)$$

And the classification hyper-plane is translated into optimal classification hyper-plane when $\|\omega\|^2$ is at its minimum. The sample points satisfied equation (1) are called support vector (The points on the $H_1, H_2$). Solving the optimal hyper-plane can be transformed into the following constrained optimization problem:

$$\begin{cases} \min & \frac{1}{2}\|\omega\|^2 \\ sub & y_i(\omega \cdot x_i + b) - 1 \geq 0, i = 1, 2, \cdots, n. \end{cases}$$

This is a convex quadratic programming problem since the objective function and constraints are all convex. The problem exist a unique global minimum solution according to optimization theory. By using Lagrange multiplier method and KKT conditions:

$$\alpha_i[y_i(\omega \cdot x_i + b) - 1] = 0. \qquad (2)$$

Finally, the optimal classification function takes the following form:

$$f(x) = \text{sgn}[(\omega^* \cdot x) + b^*] = \text{sgn}[\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*], \qquad (3)$$

where $\alpha^*, b^*$ are parameters to determine the optimal separating hyper-plane. The $\alpha_i = 0$ to non-support vector can be known from equation (2), and so the summation in equation(3) is only for a small number of support vectors.

The nonlinear problems can be converted to linear problems by nonlinear transformation in the certain high-dimensional feature space. The key is to find the optimal classification hyper-plane in the transformation space. This transformation may be more complex, so the idea is difficult to achieve in the general case. Thankfully, we find that Kernel function $K(x_i, x_j)$ satisfied mercer condition is inner product in the transformation space according to the theory of Functional[14]. Therefore, the basic idea is to adoption of appropriate inner product function in the optimal classification hyper-plane, so that after nonlinear transformation the characteristics of the sample space become linearly separable while not increasing the computational complexities. The classification function will be as below[14]:

$$f(x) = \text{sgn}[\sum_{i=1}^{n} \alpha_i^* y_i K(x_i \cdot x) + b^*]. \qquad (4)$$

There are three major types of SVM kernel functions:

1）The Polynomial function: $K(x, x_i) = (1 + x \cdot x_i)^d$;

2）The Gaussian radial basis function:
$$K(x, x_i) = \exp(-1\|x - x_i\|^2 / \sigma^2);$$

3）The neural network function:
$$K(x, x_i) = \tanh(v(x \cdot x_i) + c).$$

The form of SVM classification algorithm is similar to ANN, which hidden nodes are corresponding to the integral operator kernel function between input sample and support vector, while the output nodes correspond to a weighted linear combination of hidden and output layer. The SVM network structure diagram is shown in Fig.2. We only need to select the kernel function satisfies Mercer condition as we design unfold.
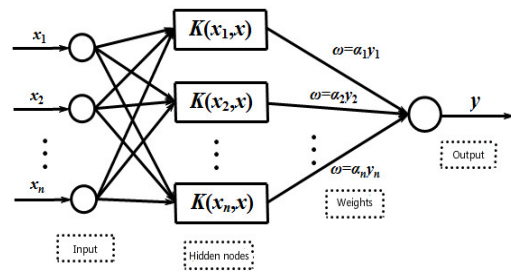


Figure 2. The topological diagram of SVM

### IV. EVALUATION MODEL

#### A. Sample Source

At present, there is no uniform evaluation criterion of ecological vulnerability and related evaluation indicators. This paper analyzes the influencing factors of regional

ecological environment, and then 11 factors are identified as Ningdong ecological vulnerability evaluation indicators by using principal component analysis according to the scope of all indicators and the previous research, see Table I. And three grade evaluation criterion for Ningdong ecological vulnerability are constructed, which grade I represents slight vulnerability, grade II represents moderate vulnerability, grade III represents severe vulnerability. The higher the level is, the greater vulnerability, detailed in Table II.

The different samples between each level are generated randomly according to index interval value of ecological environment vulnerability assessment, so that samples belong to the different indicators level. The 60 groups sample datum are generated from each grade by using random matrix functions of MATLAB, so a total of 180 groups sample datum are generated, which can be used to train and test SVM.

TABLE I. THE EVALUATION INDICATORS AND MEANING

| Evaluation index | Measurement and statistical method | Index meanings |
|---|---|---|
| $T_1$ average annual rainfall/mm | The sum of the average monthly rainfall | The basic situation of rainfall |
| $T_2$ evaporation coefficient | The average annual evaporation / average annual rainfall | The degree of drought |
| $T_3$ Forest coverage/% | Forest area/total land area | The regulation and anti-jamming capability of vegetation |
| $T_4$ The proportion of water area/% | Water area/total land area | The degree of air-dried |
| $T_5$ The proportion of dune area/% | Dune area/total land area | The indicators of land quality |
| $T_6$ The proportion of the total population of college/% | The number of college/total population | The status of population quality |
| $T_7$ Per capita GDP/yuan | GDP/total population | The overall level of comprehensive economic development |
| $T_8$ Rural per capita net income/yuan | Farmers' net income / total population | The overall level of economic development |
| $T_9$ Grain yield | Food production / cultivated area | The index ability of grain production |
| $T_{10}$ Per capita total power of agricultural mechanization/ kW | The total power of agricultural mechanization / total population | The level of agricultural modernization |
| $T_{11}$ The proportion of the second industrial output in GDP /% (kW) | The second industrial output value / GDP value | The level of modernization |

TABLE II. THE EVALUATION GRADE STANDARDS

| The evaluation index | evaluation standards | | |
|---|---|---|---|
| | I(mild vulnerable) | II(moderately vulnerable) | III(severe vulnerable) |
| $T_1$ | [400,500] | [300,400) | <300 |
| $T_2$ | [1.00,4.00] | (4.00,7.00] | >7.00 |
| $T_3$ | [15.00,20.00] | [10.00,15.00) | <10.00 |
| $T_4$ | [5.00,6.00] | [4.00,5.00) | <4.00 |
| $T_5$ | [2.00,3.00] | (3.00,4.00] | >4.00 |
| $T_6$ | [30.00,40.00] | [20.00,30.00) | <20.00 |
| $T_7$ | [17000,22000] | [12000,17000) | <12000 |
| $T_8$ | [7000,10000] | [4000,7000) | <4000 |
| $T_9$ | [2000,2500] | [1500,2000) | <1500 |
| $T_{10}$ | [1.00,1.20] | [0.80,1.00) | <0.80 |
| $T_{11}$ | [40,50] | (50,60] | >60 |

### B. Optimized Parameters and Structure of SVM by GA

In the case of linear and inseparable, a relaxation $\xi_i (\geq 0)$ can be added into condition (1), which becomes:

$$y_i(\omega \cdot x_i + b) - 1 + \xi_i \geq 0. \tag{5}$$

The objective function will be modified to:

$$\begin{cases} \min & \frac{1}{2} \|\omega\|^2 + C(\sum_{i=1}^{n} \xi_i) \\ sub & y_i(\omega \cdot x_i + b) - 1 + \xi_i \geq 0, i = 1, 2, \cdots, n. \end{cases}$$

It should be compromised that the least error-classifying samples and the maximum class interval, so we can get the generalized optimal classification hyper-plane. Where constant $C > 0$, it controls the degree of punishment to error-classifying samples.

Low-dimensional space sample sets are often difficult to classify. At present, it is a nice solution to transform them into a higher-dimensional space. Unfortunately, it usually brings computational burden, while Kernel function is catered to the need of this problem. In this sense, the essence of SVM is to select the Kernel function. The classification function of high-dimensional space can be received if there is proper Kernel function found[9].

Overall, the penalty factor $C$ and Kernel function $K$ are two main parameters that affect the classification accuracy. The optimized Gaussian radial basis function is selected in this paper. The GA, a newly developed optimization algorithm, is based on the theory of fittest, natural selection and genetic inheritance. The GA applies important mechanism of natural biological systems into the design of artificial system, there are essentially different compared with other optimization algorithms. The structure and

parameters of SVM evaluation model are optimized by GA in the paper, and algorithm flow chart is shown in Fig.3. Specific optimization process is shown below:

The structure model of SVM is encoded into individual represented by code string, and a number of different initial parameters are used to train population structure. The fitness of each individual is determined according to the training results or other strategies, and the individuals with high fitness are directly chosen into the next generation, the rest inherit based on probability determined by the adaptation value. The next population is generated by genetic manipulation such as crossover, mutation and so on. After the evolution of several generations, finally the structural model of an individual will meet the requirements in current population, and then all parameters of support vector are encoded into individual represented by the binary string. Generating this initial population of code string randomly, namely, the computation can be optimized by using the conventional genetic algorithms.
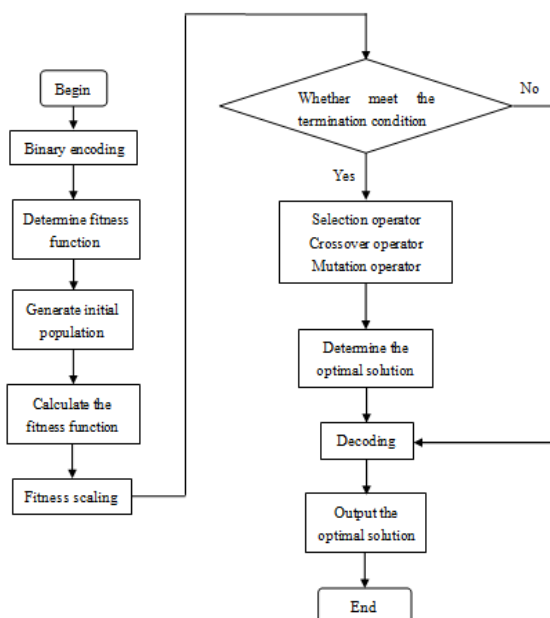


Figure 3. Optimized SVM by GA.

After each generation computation, the code strings are decoded as parameters to constitute new SVM. The mean square error of SVM output can be obtained to determine the fitness of each individual. After several computations, SVM will evolve into the global minimum error. The 40 groups datum are drawn from each level of samples datum generated randomly, 120 groups are taken as training samples in total, and the remaining 60 groups are taken as detection samples. After samples are normalized by using Libsvm toolbox of literature[10], training with 120 sets of training samples and then testing with 60 test samples. The process of optimizing structure and parameters by GA is shown in Fig.4. The classification accuracy is the highest, up to 99.6% when parameter C=0.8272 and Kernel function K=2.2418, specific results are shown in Fig.5.

Next, the BP-ANN model with the same sample is applied to evaluate and results are as follows: The classification accuracy is only 87.6% when selecting 400 training samples. The classification accuracy is 94.5% when selecting 1600 training samples. The classification accuracy reaches to 99.6% when choosing 2400 training samples. It shows that BP-ANN model can't produce accurate results in small sample training. Compared with BP-ANN model, the GA-SVM can accurately evaluate the vulnerability in the case of small sample training.
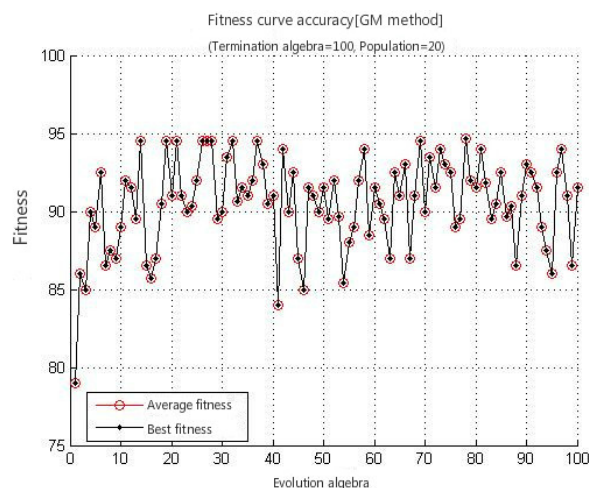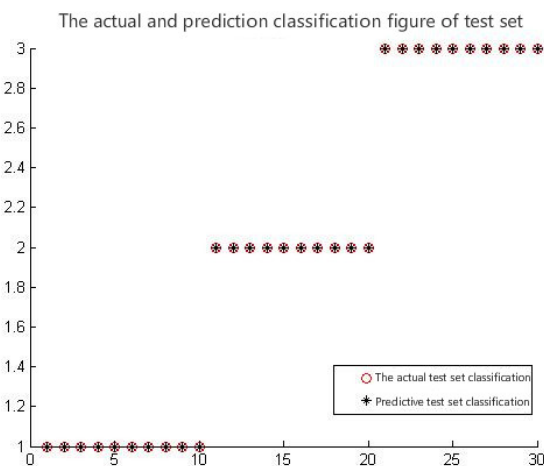


Figure 4. The fitness curve of optimized SVM



Figure 5. The comparison results between actual and prediction

## V. EVALUATION RESULTS

The paper evaluates eco-environment vulnerability of Ningdong base by using trained GA-SVM model and BP-ANN model respectively. Data are from Development and Reform Bureau, Economic Development Board and Environmental Protection Agency of Ningdong Energy and Chemical Base Management Committee and so on, the results are shown in Table III.

Table III shows that final evaluation grade is consistent by using GA-SVM model and BP-ANN model, but the grade of each evaluation index is slightly different. There are six indicators belong to II grade in the former evaluation results, however, there are only four indicators belong to II grade in the latter evaluation results. Although there are eight indicators belong to the same level in both evaluation results, but the specific values are different, the former is better than the latter. In the past two years, ecological vulnerability of Ningdong Base maintained at II grade (moderate vulnerability). Currently, the ecological environment is being improved gradually on the whole. This reflects that the evaluation result of GA-SVM model is more in line with current ecological environment of Ningdong base. Compared with BP-ANN model, the GA-SVM model has a series of advantages, such as small sample training, high classification accuracy, excellent generalization performance and so on.

TABLE III.    THE EVALUATION NDEX ANDGRADE

| Evaluation model | Evaluation samples and grade | | | | | | | | | | | Assessment level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | $T_7$ | $T_8$ | $T_9$ | $T_{10}$ | $T_{11}$ | |
| GA-SVM | 237 | 6.86 | 11.4 | 3.47 | 3.77 | 30.96 | 1597 | 6828 | 1529 | 0.79 | 71 | II |
| | III | II | II | III | II | I | II | II | II | II | III | |
| BP-ANN | 269 | 7.29 | 11.27 | 2.91 | 4.09 | 31.95 | 1729 | 5536 | 1737 | 0.90 | 77 | II |
| | III | II | II | III | III | I | I | II | II | II | III | |

## VI.    DISCUSSION AND CONCLUSIONS

The parameters and structure of SVM are optimized by GA in this paper, so that it has higher classification precision. At the same time, the evaluation model of ecological environment is established based on GA-SVM to evaluate ecological environmental vulnerability of Ningdong base. The assessment results are consistent with BP-ANN model, but GA-SVM model has more advantages and practical value.

By analyzing the causes of ecological environmental vulnerability of the Ningdong Base, the main factors leading to the ecological vulnerability are reflected in two aspects: (1) objective factors. Ningdong Base belongs to typical continental arid climate, which has low humidity and rainfall. The water resources are very deficient because most of the base area is seriously degraded grassland and desert. Less precipitation and large evaporation are the potential factors that cause the ecological environment vulnerability of Ningdong Base. (2) Subjective factors. Several years ago, ecological environment was disconnected with the economic construction. The ecological environment was deteriorated seriously, as was excessive pursuit of economic growth target. Furthermore, unreasonable industry exploitation for a long time has led to many outstanding issues as surface subsidence, groundwater recession, vegetation degradation and so on.

The good news is that over the years, the government pays close attention to the relationship between resources and environment, and it constructs a highly efficient, coordinated and sustainable relationship. According to cooperation monitoring research of Ningxia Water Resources Department with Ningxia Institute of Remote Sensing, the ecological environment is being good development. It also verifies the evaluation results of GA-SVM model used in this paper.

### REFERENCES

[1]  T. Moiseenko, N. Gashkina, and Y. Sharova, "Ecotoxicological assessment of water quality and ecosystem health: A case study of the Volga River," Ecotoxicology and Environmental Safety, vol. 71, pp. 837-850, April   2008.

[2]  D. Vignati, D. Secrieru, and Y. Bogatova, "Trace element contamination in the arms of the Danube Delta: Current state of knowledge and future needs," Journal of Environmental Management, vol. 125, pp. 169-178, August   2013.

[3]  Z. B. Wang, C. L. Fang, and S. W. Cheng, "Evolution of coordination degree of eco-economic system and early-warning in the Yangtze R iver Delta," Journal of Geographical Sciences, vol. 23, pp. 147-162, September 2013.

[4]  H. R. Safavi, E. M. Khoshoei, and A. R. Zamani, "Integrated index for assessment of vulnerability to drought," Water Resources Management, vol. 28,   pp. 1671-1688, August   2014.

[5]  S. Sahoo, A. Dhar, and A. Kar, "Environmental vulnerability assessment using Gray Analytic Hierarchy Process based model," Environmental Impact Assessment Review, vol. 56, pp. 145-154, January 2016.

[6]  S. B. Liu, C. Y. Wei, and Y. Q. Guo, "Exploring harmonious development between urbanization and eco-environmentbased on climate analysis-A study in Changsha, China," Journal of central South University of Technology,   vol. 18, pp. 101-107, June 2011.

[7]  G. M. Grossman, A. B. Kreuger, "Economic growth and the environment," Quarterly Journal of Economics, vol. 110, pp. 353-377, August, 1995.

[8]  M. Kijima, K. Nishide, and A. Ohyama, "Economic models for the environmental Kuznets curve: a survey," Journal of Economic Dynamics and Control, vol. 34, pp. 1187-1201, December 2010.

[9]  N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press, 2000.

[10] C. C. Chang, C. J. Lin, LIBSVM-A Library for Support Vector Machines. http: //www. kernel-machines.org/.