

Sustainable Development Evaluation Based on Support Vector Machine and Principle Component Analysis

Wei Hua WANG^{1, a *}

¹North China Electric Power University Baoding 071000, China

^aemail:15175289765@163.com

Keywords: Support Vector Machine Model; PCA Model; Evaluation system; Sustainable development.

Abstract. Sustainable development evaluation is an important theme in sustainable development research. Aim to accurately measure the sustainability of a country, this paper constructs the comprehensive evaluation system of sustainable development for a country using the Support Vector Machine(SVM) and Principle Component Analysis(PCA) On the basis of analyzing the evaluation indexes. According to this system it evaluates the sustainable development level of eighteen typical countries over the world. Comparing the evaluation result with the conclusion of United Nations shows that this model is effective.

Introduction

The concept of sustainable development was introduced by the 1987 Brundtland Report[1] as “development that meets the needs of the present without compromising the ability of future generations to meet their needs.” The resources which are available to humanity are finite. Hence, it is important to increase equity and eradicate poverty on the basis of balancing increasing population and consumption with the earth’s finite resources. The sustainable development evaluation is really significant to the future development of country. A variety of factors such as human health, food security and access to clean water which impacts on the sustainability of a country should be considered in the evaluation system[2].

Sustainable development evaluation is a high dimensional complex system which has a large number of indicators that rely on with each other. Support Vector Machine (SVM) is a new technology in data mining. It was invented originally by Vladimir N. Vapnik and Alexey Ya. Chervonenkis and based on the Structural risk minimization principle. It has obvious advantages in the study of small sample、nonlinear and high dimension space-problems and can be used widely for classification, regression, or other tasks[3]. Principle Component Analysis (PCA) is a good method of dimension reduction with strict mathematical theory. Under the principle of trying to secure data information loss least, PCA simplify the multivariable cross-section data comprehensively.

Therefore, this paper makes a comprehensive evaluation model which combines SVM with PCA to evaluate the sustainability of eighteen typical countries over the word.

The Sustainable Development Evaluation Using AVM and AHP

To distinguish whether a country achieve sustainable development relatively, a sequence of indicators[4]is needed to selected firstly to comprehensively evaluate the sustainability of a country. And we find out the data of the respective microcosmic indicators that can reflect its characteristic from the macroscopic indicators. Divide the indicators into four sorts. By these data, two sub-models are built to distinguish more sustainable countries from less ones, and these two sub-models are built on the basis of SVM and PCA respectively.

The Evaluation Index Set.

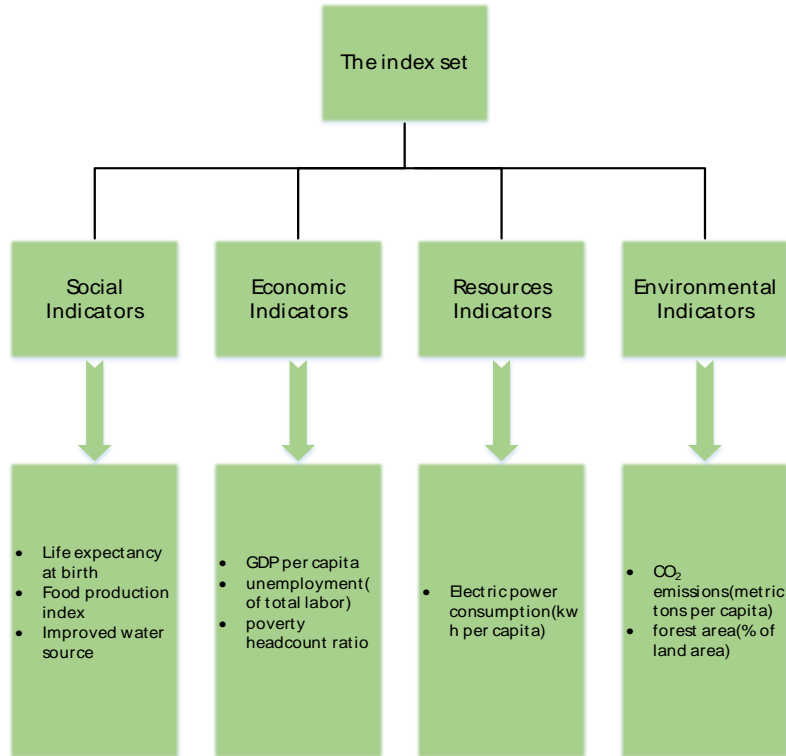


Fig. 1 The evaluation index set

Sub-Model 1: Support Vector Machine. Support vector machine (SVM) classification model[5] is built to distinguish between unsustainable countries and sustainable countries. SVM is a pattern recognition method based on statistical learning theory, its main idea is to find a hyperplane, enables it to divide the two sorts of data points correctly as many as possible, and at the same time makes the two divided sorts of data points be away from the separating surface to the furthest.

(1) Objects of classification should be chosen including Australia, Canada, Luxembourg, Switzerland, Ethiopia, Burundi, Malawi, Uganda[6], and divide these eight countries into two kinds according to the nine indices above. The eight known observation samples are signed as $(t_1, g_1), (t_2, g_2), \dots, (t_8, g_8)$ and t_i is composed of nine attributes.

(2) There is a best clarification, it is known that $w^T x + b = 0$, $w, x \in R^9, b \in R, w, b$ is uncertain, and they satisfy the following relations:

$$\begin{cases} w^T t_i + b \geq 1, g_i = 1 \\ w^T t_i + b \leq -1, g_i = -1 \end{cases} \quad (1)$$

Then it could be see that

$$g_i (wt_i - b) \geq 1 \quad (2)$$

To make the two kinds of clarifications farthest away from the hyperplane, the following equation can be got:

$$\max \frac{2}{\|w\|} \Rightarrow \min \frac{1}{2} \|w\|^2 \quad (3)$$

Then the SVM model is as follows:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & g_i(w^T t_i + b) \geq 1, i = 1, 2, \dots, n \end{aligned} \quad (4)$$

By obtaining the optimal value, the clarification function is got:

$$g(x) = \text{sgn}(w^{*T}x + b^*) \quad (5)$$

The above model is a quadratic programming one. When indexes of these eight countries are brought into $g(x)$, if the result is -1, then the country is sustainable, or is an unsustainable country.

Sub-Model 2: Principle Component Analysis. Principle component analysis[7] is used to get the futher rankings of countries according to the sustainability. We use the selected eight countries above which include nine indexes respectively to carry on principle component analysis.

(1) Normalize the original data

There are nine indicators x_1, \dots, x_9 , that need to be analyzed. Then normalize the indexes from x_{ij} to \bar{x}_{ij} :

$$\bar{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, (i = 1, 2, \dots, 8; j = 1, 2, \dots, 9) \quad (6)$$

The equation of \bar{x} is as follows:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, s_j = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, (j = 1, 2, \dots, 9) \quad (7)$$

Correspondingly, the formula is claimed

$$\bar{x}_i = \frac{x_i - x_j}{s_j}, (i = 1, 2, \dots, 8) \quad (8)$$

as standardized index vector.

(2) Calculate the correlation coefficient matrix R :

$$\begin{aligned} R &= (r_{ij})_{9 \times 9} \\ r_{ij} &= \frac{\sum_{k=1}^n \bar{x}_{ki} \times \bar{x}_{kj}}{n-1} (i, j = 1, 2, \dots, 9) \end{aligned} \quad (9)$$

$r_{ij} = 1, r_{ij} = r_{ji}, r_{ij}$ the correlation coefficient between two indexes.

(3) Calculate the eigenvalue and eigenvector:

Calculate the eigenvalue $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$, of correlation coefficient matrix and corresponding eigenvector u_1, u_2, \dots, u_m , $u_j = (u_{1j}, u_{2j}, \dots, u_{nj})^T$, Made up three new indicator variable by using characteristic vectors:

$$\begin{cases} y_1 = u_{11} \bar{x}_1 + \dots + u_{n1} \bar{x}_n \\ \dots \\ y_m = u_{1m} \bar{x}_1 + \dots + u_{nm} \bar{x}_n \end{cases} \quad (10)$$

y_1 is the first principle component, ..., y_m is the m th, there are three principle components y_1, y_2, y_3 in this topic.

(4) Choose p ($p \leq m$) principle components and calculate comprehensive evaluation value.

Calculate the Contribution rate and the cumulative contribution rate of information of eigenvalue λ . We claim

$$b_j = \frac{\lambda_j}{\sum_{k=1}^m \lambda_k} \quad (11)$$

as the contribution rate of information of y_i ;

$$\alpha = \sum_{j=1}^3 b_j \quad (12)$$

as the cumulative contribution of information of y_1, y_2, y_3 . The results are as follows:

Table 1 The contribution rate of principle component

three principle components	Information contribution rate	Cumulative contribution rate
y_1	0.6062	0.6062
y_2	0.1419	0.7481
y_3	1.106	0.8541

From the table, it is obvious to see that the cumulative contribution rate exceeds 0.85, thus three principle components are used to represent the original nine indexes.

Z is defined as composite scores:

$$Z = \sum_{j=1}^3 b_j y_j \quad (12)$$

Comparing the composite scores based on the three principle components, the final rankings can be got[8].

Experimental Analysis

Eighteen countries which could stand for different degrees[9] in development around the world are used to study the evaluation performance of the model.

First the SVM evaluation is used to get the result below:

Table 2 The evaluation result of SVM

Sustainable	Unsustainable
America	China
Japan	Nepal
New Zealand	India
Singapore	South Africa,
Switzerland	Burma,
Luxembourg	Russia
Canada	Uganda
Australia	Malawi
	Burundi
	Abyssinia

Second the PCA evaluation is used to get the result, there we take the results of 8 samples of countries as an example. The composite scores of the eight countries are as follows:

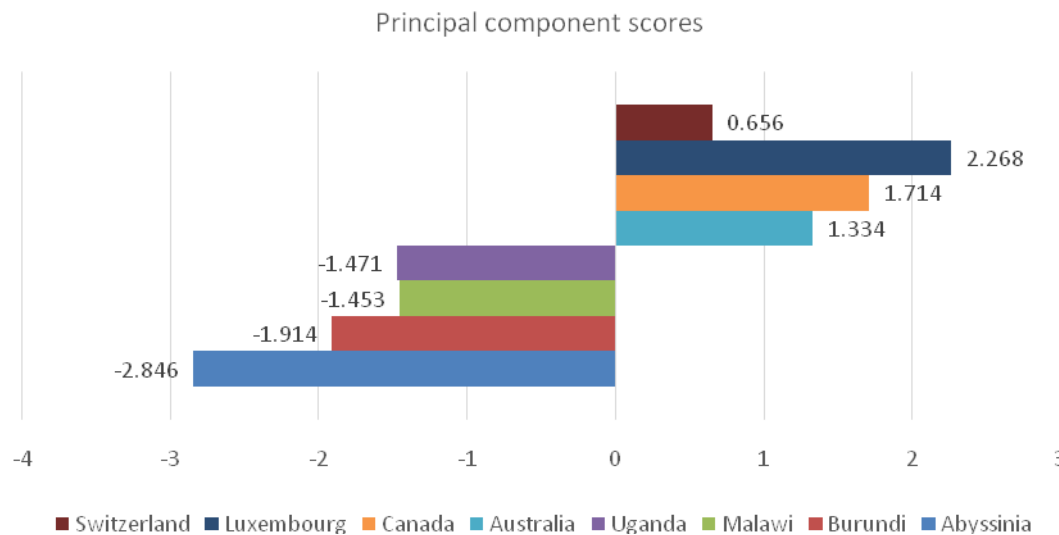


Fig. 2 principal component score

Summary

In the study, SVM and PCA are applied comprehensively to evaluate sustainable development of countries. Support Vector Machine (AVM) has obvious advantages in the study of small sample, nonlinear and high dimension space-problems and Principle Component Analysis (PCA) is a good method of dimension reduction. Through analyzing the evaluation indexes, evaluation system of sustainable development is constructed using AVM and PCA comprehensively. Eighteen typical countries are used to study the performance of the evaluation system. Comparing the evaluation results with the conclusion of United Nations[10] shows that the model is effective to evaluate the sustainability of countries.

References

- [1] Information on http://en.wikipedia.org/wiki/Our_Common_Future
- [2] Z.W. Yuan: Bulletin of Chinese Academy of Sciences, Vol. 27 (2012) No.3, p.280
- [3] Information on http://en.wikipedia.org/wiki/Support_vector_machine
- [4] Information on <http://blogs.worldbank.org/>
- [5] S.K. Si, X.J. Sun: *Mathematical Modeling*(4th ed)(National Defense Industry Press, China 2014),p.178.
- [6] H.Y. Wang: China Population Resources and Environment, Vol. 6 (1996).No.1, p.39(In Chinese)
- [7] S.K. Si, X.J. Sun: *Mathematical Modeling*(4th ed)(National Defense Industry Press, China 2014),p.207.
- [8] S.K. Si, X.J. Sun: *Mathematical Modeling*(4th ed)(National Defense Industry Press, China 2014),p.361.
- [9] Information on <http://blogs.worldbank.org/>
- [10] Information on <http://blogs.worldbank.org/>