Applied Soft Computing 12 (2012) 1801–1809

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A new method of soft computing to estimate the contribution rate of S&T progress on economic growth

Yong He*, Nuo Liao, Hongwei Liu, Yinghong Zhong

School of Management, Guangdong University of Technology, Guangzhou, Guangdong 510520, China

ARTICLE INFO

Article history:
Received 23 January 2011
Received in revised form
17 November 2011
Accepted 10 January 2012
Available online 10 February 2012

Keywords:
Economic contribution rate of S&T progress (ECRST)
Genetic iterative self-organizing data analysis technique algorithm (GA-ISODATA)
Fuzzy artificial neural network (FANN)
Science and technology (S&T) progress

ABSTRACT

In this paper, soft computing is applied to estimate the contribution rate of science and technology (S&T) progress on economic growth. First, the main influence factors of economic growth are defined, consisting of capital assets, labor force, human capital and research and development (R&D), and the human capital is calculated by improved labor-payment method. Second, target system is categorized by genetic iterative self-organizing data analysis technique algorithm (GA-ISODATA). Then, we set up the I/O model by fuzzy artificial neural network (FANN), with the capital assets, labor force, human capital and R&D as input variables, and the corresponding gross domestic product (GDP) as the output, to extract several fuzzy rules. Last, from the obtained fuzzy rules, we can get the effect of influence factors on economic growth, and calculate the economic contribution rate of S&T progress (ECRST). Take Guangdong province of China as an example, the result indicates that: during the year 2000–2008, Guangdong province (contains 21 cities) could be classified into three clusters according to the S&T progress. The first cluster (High S&T) has an ECRST of 47.52%, and contains 4 cities; the second cluster (Medium S&T) has an ECRST of 42.74%, and contains 4 cities; the third cluster (Low S&T) has an ECRST of 39.96%, and contains 13 cities; the average ECRST of Guangdong province is 44.02%. The result is accordance with the economic reality of Guangdong province, and demonstrates the validity of the proposed method.

1. Introduction

Economic contribution rate of science and technology progress (ECRST) is defined as the contribution share of science and technology (S&T) progress, as an influence factor in economic growth. Many studies have been reported on the contribution rate of S&T progress on economic growth [1–8], including Solow residual value, production function method, etc. But, the economy system is a complex nonlinear system, so the contribution rate of S&T on economic growth differently, depending on the region under study, the period selected and the relative technology level. Many problems of the traditional computing methods remain: (1) it is not easy to quantify the value of human capital, S&T level, economic institution and economic structure of a region; (2) the traditional methods ignore the complexity of target system under study and the possible incompletion of statistical data; (3) the traditional methods ignore both the long-term effect and the lagging effect of S&T on economic growth. Therefore, it is necessary to search for a new computing method that could not only improve over the traditional computing methods but also help to reveal the true relationship between S&T and economic growth.

Soft computing (SC) [9] is a newly developed computing method that combines various knowledge and methods to set up an integrated system to solve the complicated actual problems under complex circumstances. In order to solve complex problems, SC uses multiple techniques simultaneously in the computation in a harmonious manner. Generally, soft computing technique consists of artificial neural networks (ANN) [10,11] for pattern recognition and self-adjustment to evolving environment, fuzzy systems (FS) [12,13] for reasoning and decision-making and genetic algorithm (GA) [14] for the optimization of systems. These three techniques not only function independently but also effectively integrate together into the soft computing.

Considering that the real economic system is complex and nonlinear, we introduce soft computing into the field of complex economic system, to estimate ECRST. The basic idea of soft computing applied to estimate ECRST is as follows. First, the target system (countries or regions) is classified in a fuzzy fashion according to technology progress. Second, we set up the fuzzy mapping relation from influence factors (capital assets, labor force, human capital and research and development (R&D) expenditure) to economic output, and calculate the ECRST in each cluster. Third, we calculate the average ECRST of the target system.

* Corresponding author.
E-mail address: jiabeisheng@sina.com (Y. He).

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2012.01.007

The reminder of this paper is organized as follows. In Section 2, we propose the genetic iterative self-organizing data analysis technique algorithm (GA-ISODATA), describe the fuzzy artificial neural network (FANN), and provide the procedure to estimate ECRST by GA-ISODATA and FANN. Section 3 is case study. We compute the ECRST of Guangdong province in China during the period of 2000–2008. Section 4 is conclusions.

2. The method of soft computing to estimate ECRST

2.1. GA-ISODATA algorithm and the optimal clustering number

2.1.1. Fuzzy c-clustering

Definition 1. Given $X = \{X_1, X_2, \ldots, X_N\} \subset R^n$, where R^n denotes the Euclid n dimensional vector space of real numbers. For $\forall j$, $1 \le j \le N$, $X_j = (x_{j1}, x_{j2}, \ldots, x_{jn})^{\mathrm{T}} \in R^n$, where $x_{jk} (k = 1, 2, \ldots, n)$ is the k-th value of vector $X_j (j = 1, 2, \ldots, N)$. Let $V^{\mathrm{T}} = (V_1, V_2, \ldots, V_c)$ $(V_i \in R^n, \quad i = 1, 2, \ldots, c)$ is the matrix of central vectors of clustering, the fuzzy c-clustering for X can be expressed by the following mathematics programming:

$$\min J_h(U, V, c) = \sum_{i=1}^{c}\sum_{j=1}^{N} \mu_{ij}^h d_{ij}^2 = \sum_{i=1}^{c}\sum_{j=1}^{N} \mu_{ij}^h \left\| X_j - V_i \right\|^2, \quad 1 \le h \le \infty \tag{1}$$

such that

$$s.t \begin{cases} \sum_{i=1}^{c} \mu_{ij} = 1, \quad 1 \le j \le N \\ 0 \le \mu_{ij} \le 1, \quad 1 \le i \le c; \quad 1 \le j \le N \\ 0 < \sum_{j=1}^{N} \mu_{ij} < N, \quad 1 \le i \le c \end{cases} \tag{2}$$

where $d_{ij}^2 = \left\| X_j - V_i \right\|^2 = (X_j - V)^{\mathrm{T}}(X_j - V_i)$.

The iterative self-organizing data analysis technique algorithm (ISODATA) [15] is an iterative operation, which iterates after the initialization of matrix $U^0 = \left(\mu_{ij}^0\right)_{c \times N}$ for

$$V_i = \frac{\sum_{j=1}^{N} (\mu_{ij})^h X_j}{\sum_{j=1}^{N} (\mu_{ij})^h} \quad (i = 1, 2, \ldots, c) \tag{3}$$

and

$$\mu_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(h-1)} \right)^{-1} \quad (i = 1, 2, \ldots, c, \quad j = 1, 2, \ldots, N) \tag{4}$$

It can be demonstrated that the algorithm converges if $h > 1$ [16]. It is also clear that this algorithm depends on a pre-determined number of clusters, and therefore is unable to provide the optimal number of clusters [17].

2.1.2. GA-ISODATA algorithm

Genetic iterative self-organizing data analysis technique algorithm (GA-ISODATA) embeds ISODATA into GA. In other words, it consists of a GA outer layer and an ISODATA inner layer. ISODATA conducts the best approximation to realize fuzzy c-partition, while GA searches for the best number of clustering in the entire space. At each step GA provides the preset partitioning number for ISODATA, while the optimal partitioning number depends on the results of approximation in ISODATA. Both of the methods work together to realize the best partitioning and produce the number of optimal clustering according to the following procedure of computation:

(1) Random generation of m-species $\vec{Y_s} = \{Y_1^s, Y_2^s, \ldots, Y_m^s\}$, $s = 0$ (start with $s = 0$) in a binary system with length l.
(2) Computation of integer Y_k^s $(k = 1, 2, \ldots, m)$ corresponding to individual c_k^s $(k = 1, 2, \ldots, m)$.
(3) Random generation of initial partitioning matrix $U^s(0) = \left(\mu_{ij}^s(0)\right)_{c_k^s \times N}$ $(i = 1, 2, \ldots, c_k^s, \quad j = 1, 2, \ldots, N)$ in Definition 1, with initial step $t = 0$.
(4) Computation of cluster center matrix $V_k^s(t)$ $(k = 1, 2, \ldots, m)$ by Eq. (3).
(5) Iterative computation of a new fuzzy matrix $U^s_{c_k^s \times N}(t+1)$ $(k = 1, 2, \ldots, m)$ by Eq. (4). For preset ε, if $\max_{i,j} \left| \mu_{ij}^s(t+1) - \mu_{ij}^s(t) \right| < \varepsilon$ $(i = 1, 2, \ldots, c_k^s, \quad j = 1, 2, \ldots, N)$, let $U_k^* = \left(\mu_{ij}^s(t)\right)$ $(k = 1, 2, \ldots, m)$ and compute V_k^* $(k = 1, 2, \ldots, m)$ by Eq. (3) and then go to Step (6). Otherwise return Step (4).
(6) Computation of fitness function $G\left(U_k^*, V_k^*, c_k\right)$ $(k = 1, 2, \ldots, m)$.
(7) Let $G(U^*, V^*, c^*) = G\left(U_k^*, V_k^*, c_k\right)$ $(k = 1, 2, \ldots, m)$, the binary system string Y_k^s corresponding to c^* is the best individual of the current species. Keep the best individual of the current species, and perform the GA operation based on the proportion of fitness degree, the crossover probability $p_c \in [0, 1]$, and the variation probability $p_m \in (0, 1)$, which leads to a new species; then return $\vec{Y_{s+1}} = \{Y_1^{s+1}, Y_2^{s+1}, \ldots, Y_m^{s+1}\}$ to Step (2) until the GA converges.

2.2. Introduction of fuzzy artificial neural network

2.2.1. Fuzzy system model

A fuzzy system can be expressed by following fuzzy rules:

$$R_i : \text{if } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \ \ldots \text{ and } x_n \text{ is } A_{in} \text{ then } y_i = z_i(X) \quad (i = 1, 2, \ldots, c) \tag{5}$$

here $X = (x_1, x_2, \ldots, x_n) \in U_1 \times U_2 \times \cdots \times U_n$ is a linguistic variable (input vector), while n is the number of linguistic variables. A_{ij} is the fuzzy set on the universe of discourse of U_j $(j = 1, 2, \ldots, n)$, and R_i $(i = 1, 2, \ldots, c)$ represents the i-th rule. y_i is the output of the i-th rule, and $y_i = z_i(x)$ $(i = 1, 2, \ldots, c)$ could be a number or a linear equation.

If the learning data contain N samples with n inputs and one output, given $\varepsilon > 0$, in order to make

$$\mathrm{MSE} = \frac{1}{N}\sum_{j=1}^{N} (o_j - y_j)^2 < \varepsilon \tag{6}$$

where MSE is the abbreviation of mean square error, $o_j (j = 1, 2, \ldots, N)$ is the model output value of the j-th sample, $y_j (j = 1, 2, \ldots, N)$ the real output value of the j-th sample. Let A_{ij} $(i = 1, 2, \ldots, c, \quad j = 1, 2, \ldots, n)$ be Gaussian membership function, $y_i = z_i(x)$ $(i = 1, 2, \ldots, c)$ be linear equations. When Eq. (6) is satisfied, the optimal parameters $a = (a_{ij})_{c \times n}$, $\sigma = (\sigma_{ij})_{c \times n}$ and $B = (b_{ij})_{c \times (n+1)}$ could be obtained. a and σ are the parameter matrices of the rule premise (the 'if' part). B is the parameter matrix of the rule consequent (the 'then' part). c is the optimal number of cluster from GA-ISODATA. Given input $X = (x_1, x_2, \ldots, x_n)$, the output of fuzzy model could be expressed as follows [18]:

$$y = \frac{\sum_{i=1}^{c} z_i \left[\prod_{j=1}^{n} u_{ij}(x_j)\right]}{\sum_{i=1}^{c} \prod_{j=1}^{n} u_{ij}(x_j)} \tag{7}$$