Pol. J. Environ. Stud. Vol. 27, No. 2 (2018), 927-938
DOI: *10.15244/pjoes/76407* ONLINE PUBLICATION DATE: 2018-01-15

Original Research

Analyzing and Predicting CO_2 Emissions in China Based on the LMDI and GA-SVM Model

Jianguo Zhou[1], Xuechao Yu[1]*, Fengtao Guang[2], Wei Li[1]

[1]Department of Economics and Management, North China Electric Power University, Baoding 071000, China
[2]School of Economics and Management, North China Electric Power University, Beijing 102206, China

Received: 20 July 2017
Accepted: 15 August 2017

Abstract

With the effect of CO_2 emissions being the primary cause of the greenhouse effect, a selection and analysis study of driving factors of CO_2 emissions is vital to controlling growth from the source. This paper decomposes CO_2 emissions based on the logarithmic mean division index (LMDI) from three industries and residential consumption in China during the period 2000-14. A genetic algorithm-support vector machine (GA-SVM) was established. The eight driving factors as input variables have been innovated to apply the forecasting model. In the case study, the data set of driving factors from 2000 to 2009 is selected as training samples, and the other data set of driving factors from 2010 to 2014 is regarded as test samples. The results show that the factor decomposed based on the LMDI method of CO_2 emissions is very rational and can greatly improve forecast accuracy. The effectiveness of the GA-SVM model has been proven by the final simulation, which indicates that the proposed model outperforms a back propagation neural network (BPNN) model and a single SVM model in forecasting CO_2 emissions.

Keywords: CO_2 emissions, driving factors, LMDI, GA-SVM

Introduction

Global warming, which is caused by greenhouse gas emissions, has brought severe challenges to human survival and development [1-2]. Given our reliance on fossil fuels for energy, high CO_2 emissions have become a hot issue around the world [3]. The rapid development of China's economy has been accompanied by a tremendous increase in energy consumption, and China's coal-based fossil fuel structure would be hard to change in the short term, and in the meantime there has been an unavoidable increase of CO_2 emissions. China's government is actively exploring a large number of ways to reduce energy consumption and accelerate the transition to a low-carbon economy [4-5]. According to IEA statistics in 2007, China's CO_2 emissions exceeded the United States and China became the largest CO_2 emissions country in the world [6]. At the Copenhagen Climate Change Conference in 2009, the Chinese government promised that CO_2 emissions intensity would decline by 40~45% in 2020 compared with 2005 levels [7]. In 2015 the Chinese government made a commitment to reduce CO_2 emissions intensity by 60~65% in 2030 in contrast to 2005 levels, and planned to peak CO_2 emissions around 2030 and strove for an early peak [8]. There had been declared once again by the Chinese government in 2017 [9]. China planned to increase the value proportion of the GDP in the service sector by 4%, and to reduce energy consumption by 16% per unit of GDP and CO_2 emissions by 17% per

*e-mail: yuxuechao121@163.com

unit of GDP by 2015 compared with 2010 in the 12th Five-Year Plan Outline [10]. Therefore, actively seeking ways to reduce CO_2 emissions, developing a low-carbon economy, and achieving sustainable development are still current battles of the Chinese government. At the present stage more and more literature about CO_2 emissions is appearing. They mainly focus on two aspects: the analysis of CO_2 emissions factors and the prediction model of CO_2 emissions.

At the present stage, two common decomposition techniques have been applied to research methods for the scientific evaluation and quantitative analysis of driving factors of CO_2 emissions: structural decomposition analysis (SDA) and indexed composition analysis (IDA). SDA is based on the input-output model for the analysis of driving factors of CO_2 emissions [11-12], while SDA has strict regulatory requirements for data. IDA is based on the aggregate data of each department indicator for analyzing energy consumption with the characteristic of time series in major energy consuming sectors [13] under two types of methodology: Laspeyres index decomposition and Divisia index decomposition analyses proposed by Ang and Zhang [14]. A large number of studies about CO_2 emissions decomposition utilizing the IDA model have been reported. Ang explored the CO_2 emissions related to energy consumption in China's industrial sector by using the LMDI method for the first time in the analysis of different exponential factorization methods [15]. Wu applied the LMDI approach to studying energy-related CO_2 emissions [16]. Lee decomposed the changes of CO_2 emissions by LMDI method in APEC countries and the result revealed that per capita GDP and population are the two major factors in the increase of CO_2 emissions [17]. The change of industrial CO_2 emissions is analyzed from 36 industrial sectors and the result showed that industrial activity and energy intensity are the two major contributions [18]. Fan analyzed changes in China's carbon intensity based on adaptive the weighted Divisia index [19]. Variation characteristics in carbon intensity in Jilin Province were researched by LMDI based on IDA, and the main driving factors and their effects were analyzed [20]. Zhang improved the LMDI technique, which includes energy density and energy consumption intensity, to explore the driving factors of carbon emission intensity in 29 Chinese provinces from 1995 to 2012 [21]. In a word, the LMDI method can improve the accuracy of the calculation. It handles well zero problems in data and has the characteristics of no residual decomposition, a simple calculation process, and intuitive decomposition results.

So far, studies of the CO_2 emissions prediction method are gradually systematized and diversified. Chinese professor Deng put forward Grey system theory in 1980s. Through the processing of historical data, this method adopted certain a regularity to analyzing data and then forecast future indicators. Thus the classic G (1, 1) model came into being. Wang applied the GM (1, 1) model based on CO_2 emissions of Jiangsu Province from 2001 to 2010 and achieved the time of carbon reduction target [22]. Song applied a model based on support vector regression (SVR), which was established for predicting China's CO_2 emissions. The results showed that China can appropriately decrease the growth speed of GDP and constantly improve energy structure so as to achieve carbon reduction targets efficiently [23]. Zhou and Zhang established a combination forecasting model based on the rough set, the Grey system model (GM), and support vector machine (SVM), which predicted China's CO_2 emissions during 2012-17 [24]. Sun and Xu established an improved particle swarm optimization-back propagation algorithm (IPSO-BP). An empirical analysis of CO_2 emissions in Hebei Province of China verified the overall performance and effectiveness of the model [25]. Sun and Liu classified influence factors of carbon emissions from the three major industries and residential consumption. Least squares support vector machine (LSSVM) was applied to predict different types of CO_2 emissions. Case studies indicated that classification and prediction of CO_2 emissions can commendably promote forecast accuracy [26].

Given the above, although the LMDI method is widely applied to CO_2 emissions, it is confined to factor decomposition and policy analysis at present. No one adopts it to quantify the selection of indicators as an input index of prediction models. The major prediction models of CO_2 emissions are only confined to the single model: Grey Prediction, BPNN, and SVM. The index selection of the prediction model is generally qualitative analysis. In this paper, the principal purpose of the study decomposes the CO_2 emissions from the three industries and residential consumption during 2000-14 based on the LMDI method, which could obtain and analyze eight driving factors: energy structure, industrial structure, industrial energy intensity, life energy intensity, urbanization, population scale, economic output, and per capita income. These driving factors will be adopted as the input index of the SVM model. In order to improve the precision of the prediction, genetic algorithm (GA), which simulates the biological evolution process based on natural selection and genetic mechanism, is adopted to predict CO_2 emissions. The parameters in SVM are optimized by GA to accelerate the convergence speed and improve the accuracy of the model, which can guarantee the generalization and learning abilities of the SVM model. Meanwhile, it also verifies the feasibility of using the eight factors of LMDI decomposition as a prediction index. Through the case study, the proposed GA-SVM model based on LMDI has greater accuracy compared with contrast models in predicting CO_2 emissions in China.

Methods and Model

LMDI Decomposition Model

The LMDI decomposition model adopted in this paper is established based on the Kaya identity, which shows that CO_2 emissions are mainly determined by population size, economic level, and energy intensity [27] – as is shown in Equation (1):

$$C = \frac{C}{E} \times \frac{E}{GDP} \times \frac{GDP}{P} \times P \tag{1}$$

…where C represents total CO_2 emissions, E represents total energy consumption, GDP represents gross domestic product, and P represents population scale. The Kaya identity is very simple in structure and operation but it has a confined number of variables, and the results can only be confined to the quantitative relationship between CO_2 emissions and population (economy as well as energy). The LMDI decomposition method could effectively solve the residual problems in the decomposition and the zero value problems or the negative value problems in the data [28-29]. Therefore, the article employs LMDI to identify the driving factors of CO_2 emissions in China. According to Kaya identity, we can divide CO_2 emissions into several affecting variables from the three industries and residential consumption. This approach is used to identify and estimate the factors affecting CO_2 emissions and to provide a clear view of the factors influencing changes in CO_2 emissions. As shown in Eqs. (2-4).

$$C = C_1 + C_2 \tag{2}$$

$$\begin{aligned} C_1 &= \sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{C_{ij}}{E_{ij}} \times \frac{E_{ij}}{E_i} \times \frac{E_i}{GDP_i} \times \frac{GDP_i}{GDP} \times \frac{GDP}{P} \times P\right) \\ &= \sum_{i=1}^{3}\sum_{j=1}^{3}(CI_{ij} \times ES_{ij} \times EI_i \times IS_i \times GP \times P) \end{aligned} \tag{3}$$

$$\begin{aligned} C_2 &= \sum_{i=4}\sum_{j=1}^{3}\left(\frac{C_{ij}}{E_{ij}} \times \frac{E_{ij}}{E_i} \times \frac{E_i}{GNI} \times \frac{GNI}{P} \times \frac{UP+VP}{P} \times P\right) \\ &= \sum_{i=4}\sum_{j=1}^{3}(CI_{ij} \times ES_{ij} \times LEI_i \times PI \times UPR \times P) + \\ &\sum_{i=4}\sum_{j=1}^{3}(CI_{ij} \times ES_{ij} \times LEI_i \times PI \times VPR \times P) \end{aligned} \tag{4}$$

…where C_1 represents CO_2 emissions of the three major industries and C_2 represents CO_2 residential emissions. Subscript i = 1, 2, 3, 4 respectively represents the first industry, the second industry, the third industry, and residential life. j = 1, 2, 3 respectively represents coal, oil, and gas. The description of each index in the model is shown in Table 1.

Using LMDI, we divide the change of CO_2 emissions into 10 parts. The aggregated change in CO_2 emissions recorded in year T is calculated by subtracting CO_2 emissions in a base year. The corresponding formula is in equation (5):

Table 1. Description of each index in the model.

Index	Description	Index	Description
C	CO_2 emissions	E	Energy consumption
GDP	Gross domestic product	P	Total population
GNI	Gross national income	UP	Urban population
VP	Rural population	CI	Carbon intensity effect
ES	Energy structure	EI	Industrial energy intensity
IS	Industrial structure	GP	Per capita GDP
LEI	Life energy intensity	PI	Per capita income
UPR	Urbanization rate	VPR	Rural population ratio

$$\begin{aligned} \Delta C &= C^T - C^0 = \Delta C_{CI} + \Delta C_{ES} + \Delta C_{EI} + \Delta C_{IS} + \Delta C_{GP} \\ &+ \Delta C_P + \Delta C_{LEI} + \Delta C_{PI} + \Delta C_{UPR} + \Delta C_{VPR} \end{aligned} \tag{5}$$

The decomposition factors are expressed in Eqs. (6-16).

CO_2 emissions intensity effect

$$\Delta C_{CI} = \sum_i\sum_j L(C_{ij}^T, C_{ij}^0) \times Ln\frac{CI_{ij}^T}{CI_{ij}^0} \tag{6}$$

Energy structure effect

$$\Delta C_{ES} = \sum_i\sum_j L(C_{ij}^T, C_{ij}^0) \times Ln\frac{ES_{ij}^T}{ES_{ij}^0} \tag{7}$$

Energy intensity effect

$$\Delta C_{EI} = \sum_i\sum_j L(C_{ij}^T, C_{ij}^0) \times Ln\frac{EI_i^T}{EI_i^0} \tag{8}$$

Industrial structure effect

$$\Delta C_{IS} = \sum_i\sum_j L(C_{ij}^T, C_{ij}^0) \times Ln\frac{IS_i^T}{IS_i^0} \tag{9}$$

Economic output effect

$$\Delta C_{GP} = \sum_i\sum_j L(C_{ij}^T, C_{ij}^0) \times Ln\frac{GP^T}{GP^0} \tag{10}$$

Population scale effect

$$\Delta C_P = \sum_i\sum_j L(C_{ij}^T, C_{ij}^0) \times Ln\frac{P^T}{P^0} \tag{11}$$

Life energy intensity effect

$$\Delta C_{LEI} = \sum_i \sum_j L(C_{ij}^T, C_{ij}^0) \times Ln \frac{LEI_i^T}{LEI_i^0} \quad (12)$$

Per capita income effect

$$\Delta C_{PI} = \sum_i \sum_j L(C_{ij}^T, C_{ij}^0) \times Ln \frac{PI^T}{PI^0} \quad (13)$$

Urbanization rate

$$\Delta C_{UPR} = \sum_i \sum_j L(C_{ij}^T, C_{ij}^0) \times Ln \frac{UPR^T}{UPR^0} \quad (14)$$

Rural population ratio

$$\Delta C_{VPR} = \sum_i \sum_j L(C_{ij}^T, C_{ij}^0) \times Ln \frac{VPR^T}{VPR^0} \quad (15)$$

…wherein:

$$L(C_{ij}^T, C_{ij}^0) = \begin{cases} (C_{ij}^T - C_{ij}^0) / (\ln C_{ij}^T - \ln C_{ij}^0), C_{ij}^T \neq C_{ij}^0 \\ C_{ij}^T, C_{ij}^T = C_{ij}^0 \\ 0, C_{ij}^T = C_{ij}^0 = 0 \end{cases} \quad (16)$$

In the decomposition analysis, the CO_2 emissions intensity effect hardly has any impact on CO_2 emissions because the carbon emission coefficient of the fossil fuel is constant. In other words, ΔC_{CI} is equal to 0. As a result of *UPR* + *VPR* = 1, we could select one of them, which is between UPR and VPR. UPR is chosen in this article. The formula could be simplified in equation (17).

$$\Delta C = C^T - C^0 = \Delta C_{ES} + \Delta C_{EI} + \Delta C_{IS} + \Delta C_{GP} + \Delta C_P + \Delta C_{LEI} + \Delta C_{PI} + \Delta C_{UPR} \quad (17)$$

Support Vector Machine

Support vector machine (SVM), which was introduced by Cortes and Vapnik in 1995 [30], is a new sort of machine-learning algorithm to handle the problem of classification. The Vapnik-Chervonenkis dimension theory and structural risk minimization (SRM) were the basis of SVM [31]. In order to achieve better generalization ability, SVM explored the best tradeoff between model complexity and learning ability based on limited sample information [32]. Since SVM is a convex quadratic programming problem and the resulting solution is a global optimal solution, there are many distinctive advantages in solving small sample, non-linear, and high-dimensional pattern recognition. The algorithm has been widely applied to the domain of function fitting, regression, and prediction problems [33]. The mathematical expression could be in equation (18):

$$Min \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$s.t. \begin{cases} Y_i - w \cdot \varphi(X_i) - b \leq \varepsilon + \xi_i \\ -Y_i + w \cdot \varphi(X_i) + b \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2 \cdots N \end{cases} \quad (18)$$

Wherein w is a weight vector that defines a normal vector to the hyper-plane that defines the boundary and b represents the hyper-plane's distance from the origin. The optimal hyper-plane has maximum margin, and the distance is between this hyper-plane and its nearest data point of each class. ξ_i and ξ^*_i represent the non-negative slack variables, which is necessary to allow misclassification to the data set so as to drop the complexity of the calculation. **c** is on behalf of the generalization parameter or soft margin classifier, which is a trade-off between the misclassification and boundary complexity.

There are two essential parameters to solve the nonlinear programming problem. One is penalty parameter **c** and the other is kernel function parameter **g**. Parameter **c** controls penalty degree in the sample while parameter **g** is Gauss distribution width, which controls radial range of the function. It is the essence of nonlinear mapping. However, there is no obvious functional relationship between the learning performance of support vector machine and parameters **c** as well as **g**. The original parameters are often given at random or depend on experience, which leads to badly predicting SVM accuracy. Although the cross validation method for selecting the **c** and **g** could avoid over fitting and improve prediction accuracy, the convergence rate is very slow. The global search characteristics of the genetic algorithm can be effectively used to select the parameters of the SVM, which can speed up convergence and improve the accuracy of the prediction.

Genetic Algorithm

Genetic algorithm (GA) is a kind of adaptive probabilistic optimization technique based on the genetic and evolutionary mechanism proposed by Holland in 1975 [34]. The advantage of the algorithm is potential parallel, robust, and has excellent global searching ability [35]. Firstly, the genetic algorithm is used to carry out gene coding for a population-coding data of solution space. Before the search of the genetic algorithm, the data of solution space must be expressed as the structure data of genetic space. After gene coding, the initial population is generated, which is the process of generating random initial solution. And then it is necessary for the fitness value of selection. Due to the fitness value representing the merits of the individual solution, the selection process is to eliminate individuals with low fitness value and retain the individuals with high fitness value so as to ensure the existence of the optimal solution. Finally, the three main operators are selection, crossover, and mutation, which are generating new solutions through these three processes, iteration loop, and until gain to the optimal solution [36-38].

The selection is to pass the optimized individual directly to the next generation by the selection operator. The selection operation is based on the fitness assessment of the individual in the group, which is denoted by objective function value. The crossover operation refers to the partial structure of the two parent individuals to be replaced and reorganized to generate new individuals. Crossover is the core operator in GA, which can make the global search ability to leap to improve. For example, the main crossover methods are single-, double-, and multi-point crossover. The mutation operation is a random alteration on certain gene loci of the individual gene string in the population. The mutation operator is used as a supplementary operator for its local search ability. Compared with the traditional search algorithm, GA is not based on a single evaluation function of the gradient or higher statistics to produce a deterministic test solution sequence but search for global optimal solution by simulating the natural evolutionary process.

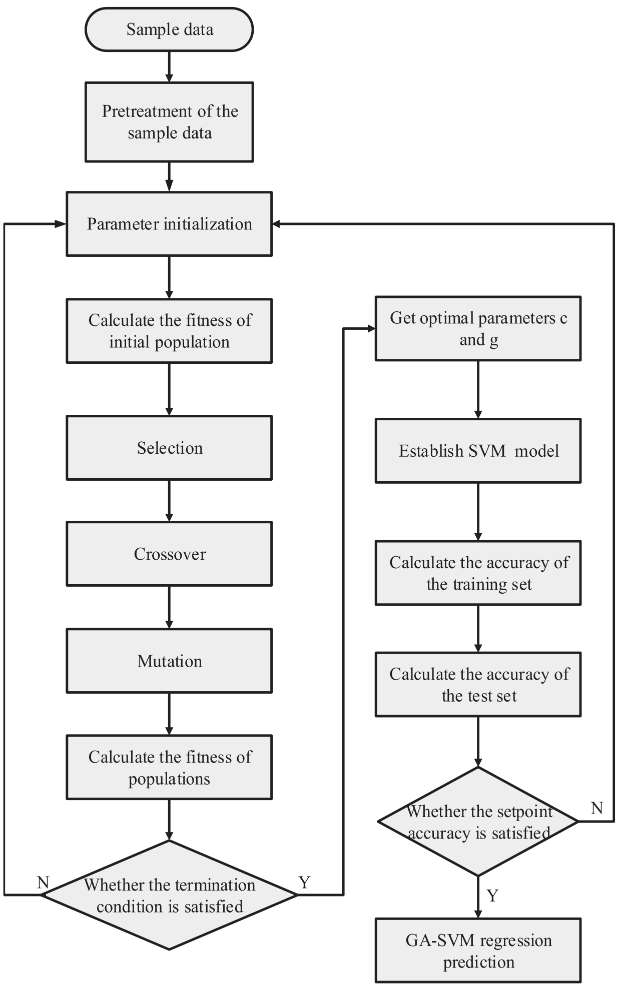

Fig. 1. Flowchart of the proposed model.

GA-SVM Model

In this section, the prediction model of CO_2 emissions based on GA and SVM is shown in Fig. 1. The data set during 2000-09 are selected as training samples, and the other data set (from 2010-14) are regarded as test samples. The flowchart of the proposed model can be acquired as per the following steps:

1) Collect related data sets needed for the study and pre-train them.
2) Initialize the parameters of GA: Set the initial population number to 40, the maximum value of iterations is 100, the crossover and mutation coefficients are considered as 0.8 and 0.5, the range of search space is [-10, 10], and velocity range is [-1, 1].
3) Select the fitness values: The genetic population will be encoded; adopt training error as the fitness function in order to distinguish the pros and cons of parameter selection.
4) Update parameters c and g: The penalty factor c and the kernel function g of the SVM are iterative calculation, and select the fitness value; if the current iteration number is equal to the maximum number of iterations, it could output optimal parameters c and g – otherwise repeat Step 2.
5) Establish SVM model: The radial basis function is selected and ε is set to 0.01; according to the training set data and getting parameters c and g, establish SVM model and calculate accuracy of the training and text sets.
6) Output the optimal solution: If the optimal parameters satisfy the set point accuracy then establish the GA-SVM model; otherwise repeat Step 2.

Results and Discussions

Data Analysis

Data Source and Conversion

This paper seeks energy consumption as well as other relevant data in China during 2000-14 for performing case research. Because statistical data of CO_2 emissions cannot be directly obtained, it can be converted according to the relevant data and proportionality coefficient. First, total energy consumption and component percentage of different sorts of energy, which are recorded in the "China Statistical Yearbook" and "China Energy Statistical Yearbook," which is finishing as shown in Table 2. Second, in order to eliminate the influence of price factors, GDP and Resident income are the base year 2000. The carbon emission coefficient of primary energy is provided by the Energy Research Institute of the National Development and Reform Commission (Table 3). Finally, the sample time is selected from 2000 to 2014 from the latest China Statistical Yearbook. The CO_2 emissions of China over the years can be obtained through the conversion method, and the calculation results are shown in Table 4.

Table 2. Primary energy consumption (unit: 10^4 tons of SCE).

Year	Primary energy consumption	Coal	Petroleum	Natural gas	Nuclear power
2000	130,297.00	89,253.45	28,665.34	2,866.53	9,511.68
2001	134,916.00	91,742.88	28,602.19	3,237.98	11,332.95
2002	148,222.03	101,532.10	31,126.62	3,409.11	12,154.20
2003	174,992.02	122,844.40	35,173.39	4,024.82	12,949.41
2004	203,227.04	142,665.40	40,442.17	4,674.22	15,445.25
2005	224,682.04	162,669.80	39,993.40	5,392.37	16,626.47
2006	246,270.02	178,299.50	43,097.25	6,649.29	18,223.98
2007	265,582.05	192,547.00	45,148.94	7,967.46	19,918.65
2008	291,447.98	208,385.30	48,671.82	9,909.23	24,481.63
2009	306,647.06	219,559.30	50,290.11	10,732.65	26,065.00
2010	324,939.02	224,857.80	56,539.39	12,997.56	30,544.27
2011	348,002.00	244,297.40	58,464.34	16,008.09	29,232.17
2012	402,138.98	275,465.20	68,363.63	19,302.67	39,007.48
2013	416,913.96	281,000.00	71,292.29	22,096.44	42,525.23
2014	425,805.01	281,031.30	72,812.66	24,270.89	47,690.16

Table 3. CO_2 coefficients of primary energy.

Project	Coal	Petroleum	Natural Gas	Nuclear, Power
Carbon emission coefficient	0.7476 kgce/kg	0.5825 kgce/kg	0.4435 kgce/m^3	0.0000 kgce/(kwh)

Table 4. Annual CO_2 emissions of China.

Year	CO_2 Emissions (100 megatons)	CO_2 Emission growth rate (%)
2000	31.60	0
2001	31.88	2.64
2002	35.03	9.88
2003	41.88	19.55
2004	48.51	15.83
2005	54.31	11.96
2006	59.17	8.95
2007	63.73	7.71
2008	69.13	8.47
2009	72.68	5.13
2010	75.83	4.33
2011	82.48	8.77
2012	91.25	10.63
2013	94.84	3.93
2014	96.54	1.79

A comparison of CO_2 emissions and its growth rate, as shown in Fig. 2, shows an upward trend in China. During 2000-14 China's total CO_2 emissions have tripled from 3,106 megatons to 9,654 megatons. China's CO_2 emissions growth has experienced a bi-modal trend, which is a rapid rise during 2000-03 and a slow decline 2003-10, and then a slow increase during 2010-12 and quick decrease 2010-12. It increases rapidly from 2.64% in 2001 to 19.55% in 2003 and CO_2 emissions average growth rate more than 10% by the end of 2008. The period belongs to the "high CO_2 emissions." Next the growth rate of CO_2 emissions is not very volatile during 2009-14. There is a small peak in 2012, but the average growth rate is not more than 10% during the period. This period belongs to the "low CO_2 emissions" and shows that China's energy-saving emission reduction measures have achieved remarkable results in the 12th Five-Year Plan.

Calculation Results and Analysis of LMDI

The eight driving factors of CO_2 emissions are obtained using the LMDI method. These driving factors include energy structure, energy intensity, industrial structure, economic output, population scale, energy intensity, per capita income, and urbanization rate. We have calculated

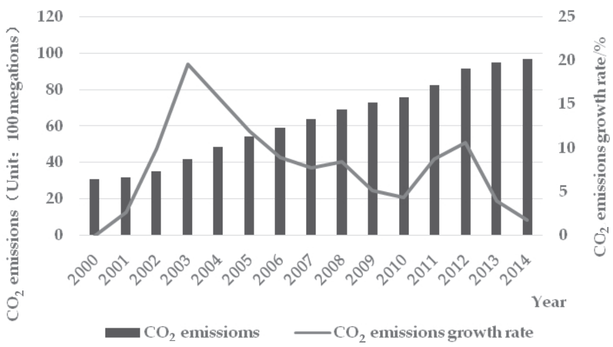

Fig. 2. CO_2 emissions and growth rate in China, 2000-14.

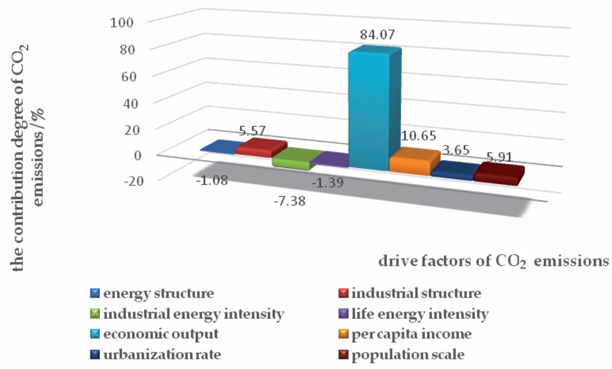

Fig. 3. Contribution degree of CO_2 emissions in China.

the contribution of each factor according to the base year of 2000, which is shown in Table 5.

The contribution of each driving factor to China's CO_2 emissions is shown. As can be seen from Table 5, the influence of different driving factors on CO_2 emissions has a significant difference, and the same factors have different effects on CO_2 emissions in different years. This is mainly determined by economic factors and policy guidelines. The trend of China's total CO_2 emissions showed a steady upward trend during 2001-14. Among them, the positive contribution of economic effect is the largest, which shows that the increase of CO_2 emissions is closely related to economic growth. The contribution value of the energy intensity effect is negative, indicating that China's energy utilization rate is gradually increasing and low-carbon emission reduction technology continues to improve. According to sorting data in Table 5 and calculating the effect of CO_2 emissions factor contribution, the results obtained are shown in Fig. 3.

The positive driving factors are industrial structure, economic output, per capita income, population scale, and urbanization level, and cumulative contributions are 5.57%, 84.07%, 10.65%, 3.65%, and 5.91%, respectively. The greatest contribution is economic output, mainly because China is a developing country, which implements industrialization as the current stage in the country's economic development. The negative driving factors include energy structure, industrial energy intensity, and life energy intensity. The cumulative contributions are

Table 5. LMDI decomposition results of CO_2 emissions in China.

Year	Structure effect	Energy intensity effect			Economic effect		Population effect		Total effect
	Energy structure	Industrial structure	Industrial energy intensity	Life energy intensity	Economic output	Per capita income	Urbanization rate	Population scale	
2001	0.16	0.04	-1.29	-0.16	2.03	0.25	0.14	0.22	1.39
2002	0.27	0.19	-1.31	-0.14	4.48	0.59	0.29	0.44	4.81
2003	0.12	0.71	3.85	0.06	7.78	1.03	0.46	0.70	14.71
2004	0.16	0.97	2.42	-0.06	11.61	1.47	0.61	0.99	18.17
2005	0.36	1.18	2.41	-0.09	16.07	1.99	0.77	1.29	23.98
2006	0.37	1.43	1.04	-0.31	21.34	2.74	0.95	1.58	29.14
2007	0.34	1.73	-1.36	-0.72	27.43	3.56	1.15	1.89	34.02
2008	-0.17	1.98	-1.95	-0.37	32.21	4.16	1.38	2.22	39.46
2009	-0.21	2.27	-3.96	-0.51	36.79	4.72	1.58	2.51	43.19
2010	-0.39	2.88	-6.32	-0.94	42.08	5.25	1.77	2.82	47.15
2011	-0.46	3.36	-8.03	-1.02	47.97	5.99	2.01	3.20	53.02
2012	-1.42	3.78	-7.24	-0.83	55.35	6.85	2.29	3.72	62.50
2013	-1.91	3.88	-7.57	-0.84	59.65	7.51	2.52	4.07	67.31
2014	-2.70	3.93	-8.23	-1.13	63.14	8.09	2.67	4.38	70.15
Total	-5.48	28.33	-37.54	-7.06	427.93	54.20	18.59	30.03	509

-1.08%, -7.38%, and -1.39%, respectively. The greatest contribution is industrial energy intensity followed by life energy intensity. With the progress of science and technology in China, energy-saving and emission-reduction have improved significantly. The result inhibits the growth of CO_2 emissions, which promotes the transition to a low-carbon economy in China. In order to study the rising trend of CO_2 emissions, our paper mainly analyzes the impact of driving factors on CO_2 emissions.

Structural Effect

Structural effect includes the energy structure effect and the industrial structure effect. The energy structure effect refers to the impact of the proportion of applied energy types on the change of CO_2 emissions. From the decomposition results, the energy structure effect on CO_2 emissions in China is positive during 2001-07 and negative during 2008-14. Although it shows that China's energy structure has a certain effect on CO_2 emissions in recent years, the impact on CO_2 emissions is relatively limited. China's primary energy consumption is dominated by coal, and coal energy consumption is the largest impact factor. The change rate of coal energy consumption is very similar to that of energy structure, as proven by our calculations. Thus, the consumption of coal directly determines the contribution of energy structure and indirectly determines CO_2 emissions in China.

The industrial structure effect refers to the impacts of the proportion of energy consumption on total CO_2 emissions. From the decomposition results, industrial structure in China has played an important role in the growth of CO_2 emissions during 2001-14 as before, which demonstrates that the effect of past adjustments of our industrial structure is limited. Because this is due to China still being in its stage of industrialization, which the high-carbon industry accounts for a significant part. By analyzing the relationship between the change rate of energy consumption and that of industrial structure of the second industry, we find that there exists a similar development patern. Therefore, we conclude that the energy consumption of the second industry is an important factor to determine CO_2 emissions in China.

In terms of the energy structure effect, the government should develop renewable energy – including nuclear power, wind power, hydropower, solar energy, and biomass energy – in order to increase the proportion of non-fossil fuels. Scholars should conduct intensive research into distributed energy projects and energy storage technologies. In terms of the industrial structure effect, industrial structure ought to be optimized and upgraded in order to realize a low-carbon transition. On the one hand, the government should eliminate backward production capacity and limit the development of high-carbon industry, the integration of resources in the field of heavy chemical industry; on the other hand, the government should vigorously develop the modern service industry and high-tech industry as well as continuously increase the proportion of the third industry in the national economy.

Energy Intensity Effect

Energy intensity effect indicates the energy consumption of per unit GDP used to represent the input-output characteristics of the energy system and can reflect on overall efficiency of energy economic activities. In theory, if the energy intensity of an industry (department) decreases, the energy efficiency of the industry (department) will increase. Generally considered, the improvement of energy efficiency comes from the progress of science and technology. Energy intensity effect can be divided into industrial energy intensity effect and life energy intensity effect. As seen in Table 5, the negative effect of energy intensity effect on CO_2 emissions is the largest, which makes CO_2 emissions decreased significantly during 2000-14 and carbon emission driving effect to reach a total of 44.6 million tons. Among them, the negative driving effect of industrial energy intensity effect is the most obvious indicator, reaching -37.54 million tons. This shows that the industry (department) energy efficiency has improved, reflecting China's advocacy slogan of low carbon emission reduction and scientific emission reduction. These methods greatly restrict the growth of CO_2 emissions in China. As seen in Table 5, the contribution value of China's industrial energy intensity is positive during 2003-06.

Because of the 10th five-year plan period, a new round of heavy industry has developed rapidly again and energy consumption drove the fast development of economics. High energy consumption and high pollution returned, which led to an augmentation of CO_2 emissions in China. In the 11th Five-Year period the government committed to eliminating backward production capacity, improving technology, and optimizing the industrial structure. Therefore, China's industrial energy intensity contribution value is negative, which has accelerated the following declination.

In terms of the industrial energy intensity effect, the government should increase investment in advanced energy-saving technologies and promote innovation in energy extraction, conversion, and utilization. Scholars and technicians should devote themselves to the development and promotion of energy-saving technologies. In terms of the life energy intensity effect, the nation ought to call on the public to reduce waste and conserve resources. As far as possible from the moral, legal, regulatory, and other aspects, the relevant departments must impose restrictions and penalties on those who violate the regulations.

Economic Effect

Economic development can be seen from the growth of economic output and per capita income. Generally speaking, when output and income levels are increasing, fossil energy consumption and CO_2 emissions will

Table 6. Grey correlation degree between the eight driving factors and CO_2 emissions.

Driving factors	Energy structure	Industrial structure	Industrial energy intensity	Life energy intensity	Economic output	Per capita income	Urbanization rate	Population scale
GRD	0.69	0.68	0.68	0.67	0.74	0.70	0.94	0.94

be a rising trend in the production process. From the decomposed results, the economic effect is the most significant economic output effect among all the positive driving effects, and it is obviously higher than the effect of per capita income, which is closely related to China's basic national conditions. Economic development will inevitably lead to fossil energy consumption and a large amount of CO_2 emissions. With the development of science and technology, the growth trend of China's economic effect tends to be stable and CO_2 emissions in China can be expected to grow slowly.

The nation should accelerate transformation of the economic growth modalities, which means from relying mainly on speed to relying on efficiency. The driving force of economic growth should be transformed from investment factors to innovation. And the government departments should deepen the all-round reform and eliminate the institutional obstacles to realize this transformation.

Population Effect

Population effect in this paper includes two parts: the population scale effect and the urbanization effect. China's population-scale effect has been steadily increasing, which is earlier and larger than the urbanization effect. Because of population growth and industrial development, the urbanization process is accelerated now and may lead to a huge increase on fossil energy increasing, deforestation, and changing the method of utilizing land. All the factors above can have positive impacts on our CO_2 emissions growth. In terms of the population scale effect, the government should appropriately control population growth, reduce per-capita energy consumption and the level of per capita CO_2 emissions, and enhance the national quality and low-carbon consciousness. In terms of the urbanization effect, on the one hand the government must rationally plan urban structures to reduce crowded areas and vehicle congestion; on the other hand, the nation should balance the urban level and control the population flow to the first-tier cities, and properly guide the flow to the surrounding cities – thus mitigating the greenhouse effect.

Case Analysis

Model Performance Evaluation

To quantitatively examine the performance of a model, this paper selected coefficient of determination (R^2), mean square error (MSE), and mean absolute percentage error (MAPE) to measure forecast accuracy. R^2 tests the goodness-of-fit of model, wherein the higher R^2, the better the goodness-of-fit of a model. MSE measures the deviation between forecasted values and actual values. MAPE measures the forecast capacity of the model at each data point. The lower the MSE and MAPE, the higher the model precision. The algorithms of three criteria are shown in Eqs. (19-21):

$$R^2 = 1 - \sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{(y_t - \bar{y})^2} \tag{19}$$

$$MSE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|^2 \tag{20}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| * 100\% \tag{21}$$

…where n represents the forecasted number of CO_2 emissions, y_t is the t-th actual CO_2 emissions, and $\bar{y}$ is the average of actual CO_2 emissions. $\hat{y}_t$ is the corresponding forecasted carbon emission.

Data and Model Selection

To verify the correlation between the eight driving factors based on LMDI and CO_2 emissions, the method of Grey relativity analysis (GRA) is applied. The results of Grey correlation degree (GRD) are shown in Table 6.

Intuitively, the correlation degrees are mainly divided into three levels. The highest level is up to 0.94, which includes population scale and urbanization rate. It illustrates that the two driving factors have the greatest impact on CO_2 emissions. The medium level includes economic output and per capita income, whose values are 0.74 and 0.70, respectively. This explains that the influence of the two factors on carbon emissions is considerable. The lowest level is from 0.67 to 0.69, which includes energy structure, industrial structure, industrial energy intensity, and life energy intensity. It gives the facts of the four driving factors having a strong correlation with CO_2 emissions. Therefore, the Grey relational degrees are all above the threshold of 0.6. In other words, the eight driving factors have great influence on CO_2 emissions and become input variables of the forecasting model. Wherein energy structure accounted for coal proportion, industrial structure accounted for the second industry GDP proportion. The sample data set is shown in Table 7.

Table 7. Sample data set.

Year	Energy structure (%)	Industrial structure (%)	Industrial energy intensity (hundred t/million yuan)	Life energy intensity (hundred t/million yuan)	Urbanization (%)	Population scale (billion people)	Economic output (yuan/people)	Per capita income (yuan/people)
2000	68.5	45.537	1.151	0.150	36.219	1.267	7,912.082	7,816.297
2001	68.0	44.794	1.100	0.144	37.659	1.276	8,509.434	8,390.893
2002	68.5	44.452	1.107	0.145	39.089	1.284	9,224.094	9,137.282
2003	70.2	45.624	1.191	0.152	40.530	1.292	10,085.731	10,036.223
2004	70.2	45.901	1.268	0.148	41.760	1.300	11,039.381	11,025.102
2005	72.4	47.024	1.259	0.147	42.990	1.307	12,225.638	12,155.023
2006	72.4	47.558	1.226	0.141	44.297	1.314	13,705.759	13,699.141
2007	72.5	46.861	1.161	0.130	45.889	1.321	15,571.305	15,631.929
2008	71.5	46.932	1.150	0.140	46.989	1.328	16,995.157	17,123.536
2009	71.6	45.884	1.105	0.137	48.342	1.334	18,502.421	18,488.821
2010	69.2	46.396	1.063	0.127	49.949	1.341	20,365.854	20,295.683
2011	70.2	46.401	1.039	0.126	51.270	1.347	22,194.019	22,016.556
2012	68.5	45.273	1.115	0.131	52.570	1.354	23,829.028	23,791.846
2013	67.4	44.008	1.068	0.132	53.729	1.361	25,561.587	25,355.976
2014	66.0	43.103	1.014	0.126	54.770	1.368	27,285.213	27,317.982

Analysis of Results

The procedure in this paper is implemented in MATLAB 2015a on a Windows 7 system. In addition, to verify rationality of the driving factors based on LMDI and whether the proposed GA-SVM model is suitable for the study of drive factors of CO_2 emissions, a back propagation neural network (BPNN) model and a single SVM model were applied for contrast. Model 1 is a BPNN prediction model, model 2 is a single SVM prediction model, and model 3 is hybrid GA-SVM prediction model. The forecast results are shown in Fig. 4 and Table 8.

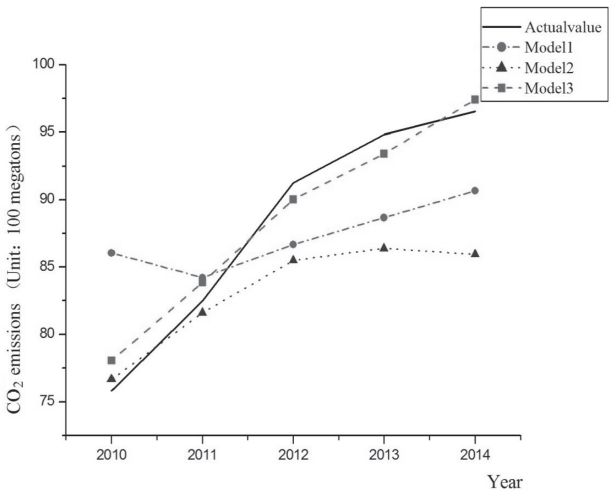

Fig. 4. Prediction of CO_2 emissions in China, 2010-14.

The prediction value of model 3 that is the GA-SVM model is closest to the actual value compared with models 1 and 2. Although the predictive trend of model 2 is similar to that of actual values, the predicted value is much lower than the actual value during the period 2012-14. The predicted value in model 1 has a wide range of changes, which is the best predictive value among plenty of tests. However, the gap between the forecast value and the actual value is too large in 2010; the prediction of model 1 is not good. R^2 is 0.97913 in model 3, which is much better than the goodness-of-fit of models 1 and 2. MSE and MAPE are 0.00514 and 1.67, respectively, in model 3, which is much smaller than the corresponding index values of models 1 and 2. It is clear that model 3 is better than models 1 and 2 for all indexes. It is clearly verified that the GA-SVM method, with a higher fitting degree and accuracy, is suitable for analyzing the relationship between CO_2 emissions and driving factors in China. Consequently, the GA-SVM method has the

Table 8. Error analysis of CO_2 emissions.

Index	Model 1	Model 2	Model 3
R^2	0.81428	0.94447	0.97913
MSE	0.12162	0.10081	0.00514
MAPE (%)	6.64	5.67	1.67

potential for practical application in similar studies about predicting CO_2 emissions.

Conclusions

This paper decomposes CO_2 emissions based on the LMDI method in China during 2000-14. The eight driving factors of carbon emission are obtained and analyzed using the LMDI method. Because the method has the advantage of no residual decomposition, the eight driving factors as input variables are innovated to apply a forecasting model of CO_2 emissions. To verify rationality of the driving factors based on the LMDI method and whether the proposed GA-SVM model is suitable for the study of drive factors of CO_2 emissions, a BPNN model and a single SVM model were applied for contrast. We can conclude that:

A) Through measuring the total amount of CO_2 emissions in China during 2000-14, its growth has experienced a bi-modal trend, which is a rapid rise during 2000-03 and slow decline in 2003-10, and then a slow increase during 2010-12 and quick decrease 2010-12. The period during 2001-08 belongs to "high CO_2 emissions." "Low CO_2 emissions" is from 2009 to 2014.

B) The greatest contribution is economic output among the eight driving factors, which implements industrialization as the current stage in the economic development of China. The greatest negative contribution is industrial energy intensity, whose energy-saving and emission-reduction have improved significantly because of the development of science and technology in China.

C) It is clearly verified that the eight driving factors based on the LMDI method is very rational. The GA-SVM method has a higher fitting degree and accuracy compared with the BPNN model and the single SVM model. The above conclusions can provide reference for the study of CO_2 emissions in China.

Acknowledgements

The current work is supported by the National Social Science Foundation of China (grant No. 15BGL145), the National Natural Science Foundation of China (No. 71471061), the Fundamental Research Funds for the Central Universities (No. 2016MS125), and Philosophy and Social Science Research Base of Hebei Province.

References

1. SPECHT E., REDEMANN T., LORENZ N. Simplified mathematical model for calculating global warming through anthropogenic CO_2 Int. J. Therm. Sci, **102**, 1, **2016**.
2. ZHANG B., PENG S., XU X., WANG L. Embodiment analysis for greenhouse gas emissions by Chinese economy based on global thermodynamic potentials. Energies, **4** (11), 1897, **2011**.
3. LIU Z., GUAN D., WEI W., DAVIS S.J., CIAIS P., BAI J. Reduced carbon emission estimates from fossil fuel combustion and cement production in china. Nature, **524** (7565), 335, **2015**.
4. YAN Z., ZHANG J., YANG Z., LI S. Regional differences in the factors that influence China's energy-related CO_2 emissions, and potential mitigation strategies. Energy Policy, **39** (12), 7712, **2011**.
5. DENG M. X., LI W., YAN H. Decomposing industrial energy-related CO_2 emissions in Yunnan province, China: switching to low-carbon economic growth. Energies, **9**, 23, **2016**.
6. IEA, 2007a. World Energy Outlook 2007: China and India Insights. International Energy Agency: Paris, **2007**.
7. CHENG K., PAN G., SMITH P., LUO T., LI L, ZHENG J. Carbon footprint of china's crop production – an estimation using agro-statistics data over 1993-2007.Agriculture Ecosystems & Environment, **142** (3-4), 231, **2011**.
8. Strengthening the Response to Climate Change – China National Autonomous Contribution. Available online: http://www.scio.gov.cn/xwfbh/xwbfbh/wqfbh/2015/20151119/xgbd33811/Document/1455864/1455864.htm (accessed on 19 November 2015).
9. Revolutionary strategy of energy production and consumption. Available online: http://www.ndrc.gov.cn/fzgggz/fzgh/ghwb/gjjgh/201705/t20170517_847664.html. (accessed on 25 April 2016).
10. The State Council of the People's Republic of China (SCPRC), The 12th Five-Year Plan Outline of National Economy and Social Development of People's Republic of China, **2011**.
11. WANG Y., ZHAO H., LI L., LIU Z., LIANG S. Carbon dioxide emission drivers for a typical metropolis using input – output structural decomposition analysis. Energy Policy, **58** (9), 312, **2013**.
12. CHANG Y.F., LEWIS C., LIN S.J. Comprehensive evaluation of industrial CO_2, emission (1989-2004) in Taiwan by input – output structural decomposition. Energy Policy, **36** (7), 2471, **2008**,
13. XU X.Y., ANG B.W. Analysing residential energy consumption using index decomposition analysis. Applied Energy, **113** (1), 342, **2014**.
14. ANG B.W., ZHANG F.Q. A survey of index decomposition analysis in energy and environmental studies. Energy, **25** (12), 11496, **2000**.
15. ANG B.W., ZHANG F.Q., CHOI K.H. Factorizing changes in energy and environmental indicators through decomposition. Energy, **23** (6), 489, **1998**.
16. WU L., KANEKO S., MATSUOKA S. Driving forces behind the stagnancy of China's energy-related CO_2, emissions from 1996 to 1999: the relative importance of structural change, intensity change and scale change. Energy Policy, **3** (3), 319, **2005**.
17. LEE K., OH W. Analysis of CO_2 emissions in APCE countries: a time-series and a cross-sectional decomposition using the log mean divisia method. Energy Policy, **34** (17), 2779, **2006**.
18. LIU L. C., FAN Y., WU G., WEI Y. M. Using lmdi method to analyze the change of China's industrial CO_2, emissions from final fuel use: an empirical analysis. Energy Policy, **35** (11), 5892, **2007**.
19. FAN Y., LIU L.C., WU G., TSAI H.T., WEI Y.M. Changes in carbon intensity in China: empirical findings from 1980-2003. Ecological Economics, **62** (3-4), 683, **2007**.

20. TIAN L., TANG J., HANG L., WANG B. Decomposition analysis of CO_2 emission intensity of Jilin industry using lmdi. Ecological Economy, **2014**.
21. ZHANG W., LI K., ZHOU D., ZHANG W., GAO H. Decomposition of intensity of energy-related CO_2 emission in Chinese provinces using the lmdi method. Energy Policy, **2**, 369, **2016**.
22. WANG Z., DANG Y. Research on carbon emission prediction in Jiangsu Province based on an improved GM (1, 1) model. IEEE International Conference on Grey Systems and Intelligent Services, 93-97, **2013**.
23. SONG J.K. China's CO_2 emissions prediction model based on support vector regression. Journal of China University of Petroleum, **36** (1), 182, **2012**.
24. ZHOU J.G., ZHANG X.G. Projections about Chinese CO_2 emissions based on rough sets and gray support vector machine. China Environmental Science, **33** (12), 2157, **2013**.
25. SUN W., XU Y. Using a back propagation neural network based on improved particle swarm optimization to study the influential factors of CO_2 emissions in Hebei province, China. Journal of Cleaner Production, **112**, 1282, **2016**.
26. SUN W., LIU M. Prediction and analysis of the three major industries and residential consumption CO_2 emissions based on least squares support vector machine in China. Journal of Cleaner Production, **122**, 144, **2016**.
27. KAYA Y. Impact of CO_2 emissions control on GNP growth: interpretation of proposed scenarios response strategies. Paris: Working Group, **1990**.
28. ANG B.W., LIU N. Handling zero values in the logarithmic mean divisia index decomposition approach. Energy Policy, **35** (1), 238-, **2007**.
29. ANG B.W., LIU N. Negative-value problems of the logarithmic mean divisia index decomposition approach. Energy Policy, **35** (1), 739, **2007**.
30. CORTES C., VAPNIK V. Support-vector networks. Machine Learning, **20** (3), 273. **1995**.
31. CHANG C.C., LIN C.J. LIBSVM: A library for support vector machines. ACM, 2011.
32. RABAOUI A., KADRI, H., LACHIRI, Z., ELLOUZE, N. One-class SVMs challenges in Audio Detection and Classification Applications. Eurasip Journal on Advances in Signal Processing, 1-14. **2008**.
33. BRERETON R.G., LLOYED G.R. Support vector machines for classification and regression. Analyst, **2010**.
34. WHITLEY D. A genetic algorithm tutorial. Statistics & Computing, **4** (2), 65, **1994**.
35. GOLDBERG D.E., GOLDBERG D.M., GOLDBERG D.E. Genetic algorithm is search optimization and machine learning. **XIII** (7), 2104, **1989**.
36. TIMOTHY B., EDWARD M., HARRY B., MICHAEL, ARLENE H., DENNIS M. Application of GA optimization for automatic generation control design in an interconnected power system. Energy Conversion & Management, **52** (5), 2247, **2011**.
37. ODA T., ELMAZI D., BAROLLI A., SAKAMOTO S., BAROLLI L., XHAFA F. A genetic algorithm-based system for wireless mesh networks: analysis of system data considering different routing protocols and architectures. Soft Computing, **20** (7), 2627, **2016**.
38. WEN J., YANG H., TONG X., LI K., WANG S., LI Y. Optimization investigation on configuration parameters of serrated fin in plate-fin heat exchanger using genetic algorithm. International Journal of Thermal Sciences, **101**, 116, **2016**.