RESEARCH ON THE DRIVING FACTORS OF ENERGY CARBON FOOTPRINT IN LIAONING PROVINCE USING RANDOM FOREST ALGORITHM

SUN, Q.

Business School, Shandong University of Technology
No.88, Gongqingtuan West Road, Zibo City, Shandong Province, China
(e-mail: yssunqiang@163.com; phone: +86-151-6608-3805)

(Received 7th Mar 2019; accepted 3rd May 2019)

Abstract. Rapid development of economy has obviously promoted the growth of carbon emission, which causes the ecological environment more fragile. How to calculate the energy carbon footprint and explore tradeoff between energy and environment becomes further significance. This study examines energy carbon footprint in Liaoning Province. The main goal of this study is to calculate the energy carbon footprint using random forest algorithm and verify the performance of random forest algorithm. The study first examines factors of energy carbon emissions and implicates calculating method of energy carbon footprint. Subsequently random forest algorithm is developed in calculating energy carbon footprint. The results show that the calculated results of random forest algorithm is close to the actual value and the trend of change is basically similar to the actual value. Furthermore, the performance of random forest algorithm is compared with Partial Least Squares algorithm. The relative error and average relative error of random forest algorithm are both less than Partial Least Squares algorithm.

Keywords: carbon footprint, LMDI, energy efficiency, low carbon economy, energy structure, strategies

Introduction

An ecosystem is an open system, in which many basic substances are continuously cycled. Ecosystem has the capacity to maintain stability via self-regulation. However, when strong pressures on the ecosystem go beyond its threshold level, the ecosystem loses self-regulation ability and the ecosystem structure will be damaged (Ginestet et al., 2013; Tao et al., 2014; Chen et al., 2018). In an age of limited resources and destructive ecology, the idea of ecological civilization should give close attention to build a safe ecological environment (Yasar and Kalfa, 2012).

Carbon footprint which refers to the overall emissions of greenhouse gases can be analyzed by life cycle assessment and comes from the concept of ecological footprint (An and Xue, 2017; Sun, 2017). It is applied to measure the energy using and reveal the trends on climate change and atmospheric environment (Hertzberg and Schreuder, 2016). Current research is focused on the assessment of the carbon footprint model and the approach to reduce carbon footprint (Mancini et al., 2016; Qi et al., 2018; Long et al., 2018). Also many factors are mentioned to reduce the carbon footprint (David et al., 2009; Sovacool and Brown, 2010). Besides, some methods such as input-output model, theoretical modeling, objective program method, profit-maximized carbon cycle model (Dominic et al., 2008; Kenny and Gray, 2009). Studies have shown that carbon accounting standards for the production and consumption-based so that newly industrialized countries face a serious carbon leakage (Sun and Zhou, 2016). The carbon footprint of the developed countries has been seriously underestimated. Therefore, it is recommended to consume carbon footprint accounting system as a benchmark by the rules of regional trade agreements, eliminate carbon leakage and promote equitable global range of carbon emission reduction targets.
The aim of this paper is to develop a random forest algorithm model for calculating the energy carbon footprint in Liaoning Province and identify the main driving factors of carbon emission. The reminder of this paper is structured as follows: after the introduction section, section 2 describes the method of energy carbon footprint, random forest algorithm and model fitting evaluation model. Results are given in Section 3 and discussion is given in Section 4. Finally, the conclusions are summarized in Section 5.

Material and methods

Energy carbon footprint

Based on the actual situation of energy consumption and economic development and taking into account the availability of data, the following categories were chosen for the calculation of carbon footprints: raw coal, coke, crude oil, petrol, diesel oil, black mineral oil and natural gas. Raw coal referred to the coal obtained after the removal of dirt and iron pyrite from the run-of-mine (rom) coal. Coke referred to the solid substance obtained after carbonation of coking coal at a high temperature (Chelgani et al., 2016; Berghorson, 2018). Crude oil was the mineral oil comprised of the polymer from natural hydrocarbon source. The black mineral oil was the black sticky residue obtained from the distillation of crude oil, or its mixture with lighter components. It was mainly used as the fuel for the steam boiler and other heating furnaces, large slow diesel oil fuel, or other industrial fuels (Romijn, 2011; Pragya et al., 2017; Dalir et al., 2018).

According to the standard coal, the content of carbon emitted from the energy can be calculated clearly. Given the 2006 IPCC Guidelines for National Greenhouse Gas Inventories published by Intergovernmental Panel on Climate Change (IPCC) advice, the carbon footprint of energy consumption in China was computed (Druckman and Jackson, 2009; Ohimain, 2012; Affuso and Hite, 2013). In addition, the coefficient of each fuel conversion, the standard coal identified in China 2015 List of Greenhouse Gases, as well as the way to compute carbon footprint and carbon emission are as follows:

$$f_{CE} = \sum \frac{44}{12} \cdot \beta_i \cdot e_i \cdot Q_i \cdot V_i \cdot q_i \cdot 10^{-3}$$  \hspace{1cm} (Eq.1)

where

- $f_{CE}$ is carbon footprint emitted from all kinds of energy.
- $\beta_i$ is the conversion factor of $i$ type of energy which was achieved by transforming the fuel into energy (TJ) in accordance with the net calorific value.
- $e_i$ is the transforming coefficient for $i$ kind of energy into standard coal Which was the benchmarking for transforming a variety of energies into standard coal.
- $Q_i$ is the consumed amount of $i$ type energy.
- $V_i$ is carbon emission factor of $i$ type energy.
- $q_i$ is the carbon oxidation ratio of $i$ type energy.
- $10^{-3}$ is the unit conversion coefficient.
- $\frac{44}{12}$ is the conversion coefficient between carbon and CO$_2$. 
The conversion factor of standard coal was identified in the General Principle for Computing Comprehensive Energy Consumption (GB/T2589-2008). The carbon emission factor, carbon oxidation ratio and energy conversion factor were identified in IPCC National List of Greenhouse Gases (2006). More detailed information of the conversion parameters is shown in Table 1.

<table>
<thead>
<tr>
<th>Energy</th>
<th>$\varepsilon_i$ (kg kg$^{-1}$)</th>
<th>$V_i$ (kg J$^{-1}$)</th>
<th>$q_i$ (%)</th>
<th>$\beta_i$ (TJ g$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw coal</td>
<td>0.7143</td>
<td>25.8</td>
<td>0.90</td>
<td>26.7</td>
</tr>
<tr>
<td>Coke</td>
<td>0.9714</td>
<td>29.2</td>
<td>0.90</td>
<td>28.2</td>
</tr>
<tr>
<td>Crude oil</td>
<td>1.4286</td>
<td>20.0</td>
<td>0.98</td>
<td>42.3</td>
</tr>
<tr>
<td>Petrol</td>
<td>1.4714</td>
<td>20.2</td>
<td>0.98</td>
<td>43.0</td>
</tr>
<tr>
<td>Diesel oil</td>
<td>1.4571</td>
<td>20.2</td>
<td>0.98</td>
<td>43.0</td>
</tr>
<tr>
<td>Black mineral oil</td>
<td>1.4286</td>
<td>21.1</td>
<td>0.98</td>
<td>40.4</td>
</tr>
<tr>
<td>Natural gas</td>
<td>12.1430</td>
<td>15.3</td>
<td>0.99</td>
<td>48.0</td>
</tr>
</tbody>
</table>

**Random forest algorithm**

Random forest (RF) algorithm is a nonlinear modeling tool which proposed by Leo Breiman (Zhou and Qiu, 2018). It is a useful method to prevent the nonlinear approximation of statistical relationships problems among different variables (Abellán et al., 2018). This method is widely used in cancer analysis, data biology, astronomy, and other fields of discriminate analysis and feature selection (Dalir et al., 2018; Kamińska, 2018). But it is nearly used in calculating energy carbon footprint and few studies reported the use of the RF algorithm in energy carbon footprint applications.

RF algorithm begins with bootstrap sampling method, random vector $\theta$ comprise the combined model:

$$h(X, \theta_m), m = 1, 2, \ldots, M$$

(Eq.2)

$X$ is the input variable, $\theta_m$ is the independent and identically distributed random vector. If the output variable is $Y$, then the original is $(X, Y)$. The predicted sequence $\{h(X, \theta_1), h(X, \theta_2), \ldots, h(X, \theta_m)\}$ will be obtained by training of $m$ times and this result can be constructed predictive model prediction and eventually results:

$$\overline{h}(X) = (1/M) \cdot \sum_{m=1}^{M} h(X, \theta_m)$$

(Eq.3)

The prediction can be obtained vector $\overline{h}(X)$ of the extrapolation error mean:

$$EMSE = EX,Y(Y - \overline{h}(X))$$

(Eq.4)

It has its inherent characteristics of random forests. When the number of trees in the forest tends to infinity, there is:

$$E_{X,Y} \left( Y - \alpha v_k h(X, \theta_k) \right)^2 \to E_{X,Y} \left( Y - E_0 h(X, \theta_0) \right)^2$$

(Eq.5)
For all $\theta$, then:

$$PE^*(\text{forset}) \leq \rho PE^*(\text{tree})$$  \hspace{1cm} (Eq.6)

$$PE^*(\text{tree}) = E_y E_{y,Y} (Y - h(X, \theta))^2$$  \hspace{1cm} (Eq.7)

\(\rho\) is weight-related between the \(Y - h(X, \theta)\) and reaming \(Y - h(X, \theta)\), \(\theta\) is independent.

Realization of random forests is as follows:

Drawn \(n_{\text{tree}}\) from the original sample by bootstrap. Each sample has not been able to get a sample consisting out-of-bag (OOB) as the test sample.

Set of raw data variables \(P\), in each tree each node randomly selects \(M_{\text{try}}\) feature, and then select the best branches.

In random forest, the choice of the parameters \(M_{\text{try}}\) is in accordance with \(M_{\text{try}} = P/3\).

Using the effect of the model to evaluate the residual mean square of OOB prediction. As shown in the following formula:

$$MSE_{\text{OOB}} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i^{\text{OOB}})^2}{n}$$  \hspace{1cm} (Eq.8)

$$R_{\text{RF}}^2 = 1 - MSE_{\text{OOB}} / \hat{\sigma}_y^2$$  \hspace{1cm} (Eq.9)

- \(y_i\) is the actual value of the output variable of the OOB.
- \(\hat{y}_i^{\text{OOB}}\) is the corresponding forecast value.
- \(\hat{\sigma}_y^2\) is the variance of the predictive value of the OOB.

**Study area**

Liaoning Province is located in the south of Northeast China, bordering on the Huanghai Sea and Bohai Sea in the south, and crossing the Yangtze River from the North Korea, as shown in Fig. 1. It is the gateway to the northeast and the eastern part of Inner Mongolia Autonomous Region of China. Liaoning Province is an important old industrial base in China and is one of the provinces with a relatively complete industrial category. The land area of Liaoning Province is 148,000 square kilometers, the mainland coastline is 2,292 kilometers long and the offshore waters cover an area of 68,000 square kilometers. It is a key area for grain production, animal husbandry, fisheries, high-quality fruits and a variety of special products. At present, there are 39 major categories, 197 medium-sized and 500 small-sized industries in Liaoning Province. With superior location and convenient transportation, Liaoning Province is an important gateway to the northeast area and an important gateway to the Eurasian Continental Bridge. It is an area with relatively developed infrastructure such as transportation and electric power.
Liaoning Province is located in the northern margin of the Pacific Rim. It has excellent geological mineralization conditions and is rich in mineral resources. More than 110 minerals have been discovered, of which 66 have been proved. The magnesite of Liaoning Province is an advantageous mineral in the world, with a reserve of 2.56 billion tons and about 25% of the world. Minerals with advantages in the country include boron, iron, diamond, talc, jade and so on. The minerals with comparative advantages are mainly coal, coalbed methane, natural gas, gold and silver.

In this study, Liaoning Province is chosen to analyze the energy carbon footprint characteristics. First, Liaoning Province is listed among the circular economy in 2008. The energy carbon emission should be slowed down dramatically to fulfill the development of circular economy. The circular economy is to reach the harmonious circulation of the economic system and the natural ecosystem through the waste reduction and resource harmlessness in the economic development. The basic principles of circular economy are reduction, reuse, and resource utilization. It is the fundamental change in the traditional growth model of mass production, mass consumption and mass abandonment. Circular economy is a production activity process that comprehensively utilizes energy and waste in accordance with the method of sustainable development under the guidance of sustainable development.

Liaoning Province is a major province of heavy chemical industry with large resource consumption and environmental pollution. Under the traditional economic development mode, the excessive development of resources has increasingly affected economic development and social progress. The depletion of resources is to force transformation to circular economy in Liaoning Province.

Second, as a province depended on nature resources, Liaoning Province is highly similar to other resource-based provinces in industrial transition. Third, because of the vulnerable environments pollution, location and regional characteristics, Liaoning Province deserve to be studied.
Data acquisition

Liaoning is a resource-dependent province, while the most common used are raw coal, coke, crude oil, petrol, diesel oil, black mineral oil and natural gas. In this section, these seven kinds of different energy are chosen to calculate the energy carbon footprint. Raw coal refers to the coal after screening to remove the meteorite, pyrite and so on from the ground or underground. Coke is a kind of solid fuel which is made by high temperature dry distillation of coal, which has high calorific value and is mostly used for iron making. Crude oil is crude oil that is directly extracted from oil wells. It is a dark brown or dark green viscous liquid or semi-solid combustible material composed of various hydrocarbons. Petrol, diesel oil and black mineral oil are liquid oil produced by crude oil.

The energy consumption data used in this section was mainly retrieved from China Statistical Yearbook, China Energy Statistical Yearbook, China Forestry Statistical Yearbook, China Agriculture Statistical Yearbook, and Liaoning Statistical Yearbook published from 1998 to 2016. Part of the data was obtained from the Liaoning provincial government website and Liaoning Statistical Bureau website. The part of the data, which was missing, was compensated by the method of moving averages. Year 1998 was considered as the benchmark, to avoid the influence of price level variations.

Fig. 2 shows energy consumption in Liaoning Province from 1998 to 2016. It is shown that the energy consumption is a increasing trence.

![Figure 2. Energy consumption in Liaoning Province from 1998 to 2016 (unit: million tons)](image-url)

Results

Evaluation results of carbon footprint

According to the calculating method of energy carbon footprint mentioned above, the normalized energy carbon footprint and energy carbon capacity as well as energy carbon surplus in Liaoning Province from 1998 to 2016 are figured up. Fig. 3 shows the carbon footprint (10^4 ton) in Liaoning from 1998 to 2016. As shown in Fig. 3, the overall carbon footprint in Liaoning Province increased by 236.768% with an annual rate of 13.154% from 410.164 million tons in 1998 to 971.139 million tons in 2016. Energy carbon footprint caused by coal has been the dominant part of energy carbon footprint. Since 2003, the carbon footprint of energy consumption has started a new round of rapid growth. In the period of national economy rapid developing, while...
accelerating the process of urbanization and a series of external factors driven by the energy consumption continues to rise. The promotion and use of natural gas as well as the rapid growth of private cars has further increased the energy requirements and consumption in Liaoning Province. The carbon footprint of energy consumption in Liaoning Province is mainly influenced by the social and economic factors. To grasp the dynamic change trend of future energy consumption is of great significance to the implementation of sustainable development in Liaoning Province.

![Energy carbon footprint](image1)

*Figure 3. Carbon footprint in Liaoning from 1998 to 2016 (unit: million tons)*

Fig. 4 shows the each energy carbon footprint in Liaoning from 1998 to 2016.

![Energy footprint of each energy](image2)

*Figure 4. Carbon footprint of each energy in Liaoning from 1998 to 2016 (unit: million tons)*

It is clear that energy structure of carbon footprint in Liaoning Province is relative stable. The proportions of the raw coal and crude oil have decreased and the proportion of the natural gas increased gradually. The main sources of greenhouse gases on a global approach include consumption of fossil fuels and primary energy. Coal combustion is one of the most important ways of pollution. Primary energy structure dominated by coal, which is impossible to get a complete change in the short term coupled with relatively outdated coal mining technology and refining our processing methods. It has increased the pressure on the government for environmental protection. This will give our global greenhouse gas control and a tremendous challenge.
Evaluation results of random forest algorithm

The major application of the random forest algorithm in this study was to calculate carbon footprint. Thus, the training dataset was selected from the statistical report from Chinese official bulletin. Then the sampling size and feature number were determined based on the principle of minimal error. And the results were generated with the training dataset with specific sampling size and feature. After a comparison of parameters and multiple simulation experiments, a set of optimal parameters is obtained. Finally, according to the simulation process, the simulation experiment is carried out with this set of parameters: $M_{try}$ is 4 and $n_{tree}$ is 600. The result of random forest algorithm is shown in Fig. 5.

![Carbon footprint calculated by random forest algorithm](image)

Fig. 5. Carbon footprint calculated by random forest algorithm

Fig. 6 shows the calculation results of carbon footprint of Liaoning Province by each algorithm.

![Comparison of algorithm accuracy](image)

Fig. 6. Comparison of algorithm accuracy

It can be seen from Fig. 6 that the calculated results of random forest and PLS are both close to the actual value, and the trend of change is basically similar. However, the value calculated by PLS algorithm is higher than random forest algorithm and the actual
value, which indicating that random forest algorithm has higher accuracy. Besides, random forest algorithm has a high accuracy comparing to the linear regression or logistic regression. The main reason of the high accuracy is generated by the random nature in both sampling size and feature. The other reason of the high accuracy is the application of bootstrap method in the algorithm.

Fig. 7 shows the absolute error of each algorithm and Fig. 8 shows the relative error of each algorithm in calculating the carbon footprint of Liaoning Province.

![Figure 7. Absolute error of each algorithm](image1)

![Figure 8. Relative error rate of each algorithm](image2)

It can be seen from Fig. 8 that the random forest model has little relative error with the actual value of energy carbon footprint. The relative error of the PLS algorithm is larger and the relative error is at most 0.1704. The relative error of random forest algorithm is less than PLS algorithm, with the largest relative error being 0.1101 and the smallest being only 0.0204. In addition, from the aspect of average relative error, the average relative error of random forest algorithm is 0.0643, while the average relative error of PLS algorithm is 0.1097, which also verifies the superiority of random forest algorithm in reducing the computational error. Thus the results of Fig. 7 and Fig. 8 are consistent with the accuracy evaluation of the reflected algorithm.
Sensitive analysis of random forest algorithm

Random forest algorithm is mainly to select the appropriate parameter values by OOB errors. The relationship between $M_{try}$ and OOB error was shown in Fig. 9. It is shown that when $M_{try}$ is 2, the OOB error reaches a peak value of 0.000181. When $M_{try}$ value is 7, a minimum of OOB error is 0.000071. With $M_{try}$ increments, OOB error has ups and downs but overall tends to increase.

![Figure 9. Relationship between the number of split variables and OOB error](image)

Discussion

Driving factors of energy carbon footprint

The LMDI (Logarithmic Mean Divisia Index) decomposition method is an empirical research method proposed by Ang et al. in 1998 to decompose and measure energy carbon footprint factors. This method extends the Kaya identity and uses the log-average formula to solve the problem. The elements are decomposed and can effectively eliminate the residuals (Branger and Quirion, 2015). Introducing it into this study, it is conducive to the decomposition of energy carbon footprint driving factors. Considering the complexity of energy carbon footprint systems, the openness and availability of economic data, the influencing factors of energy carbon footprint (Y) is divided into energy structure (X1), energy efficiency (X2), energy intensity (X3) and economic development (X4). The energy carbon footprint factors are decomposed through LMDI, as shown in Fig. 10.

The energy structure effect is further exceeded than that of energy efficiency, which are the major contributor to the increase in energy carbon footprint. Energy intensity and economic development can curtail the increase in energy carbon footprints. In particular, the energy intensity effect was reduced from -0.1157 to -1.3189. The economic development effect was reduced from -0.0064 to -0.1329. It means that adjusting energy structure can reduce energy carbon footprints and particularly alleviate dependence on primary energy. And the use of new energy and the development of green energy can also alleviate the greenhouse effect.

Based on the accumulative effects of energy carbon footprint factors, the energy structure and energy intensity effects contribute the most to the increase in energy carbon footprint, which is shown in Fig. 11.
Duro (2006) provide a methodology for decomposing international inequalities in per capita CO₂ emissions into Kaya (multiplicative) factors which indicated that energy intensity have made a less significant contribution to energy carbon footprint. This paper demonstrate energy intensity effects contribute the most to the increase in energy carbon footprint. In fact, energy intensity is truly affected the energy carbon footprint. This paper is the continuation of energy carbon footprint driving factors.

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{Figure10}
\caption{Factor decomposition of carbon footprint in Liaoning}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{Figure11}
\caption{Cumulative effect of energy carbon footprint}
\end{figure}

Besides, Liu (2007) analyzes the change of industrial carbon emissions from 36 industrial sectors in China over the period 1998–2005 and the industrial carbon emissions are decomposed into carbon emissions coefficients of heat and electricity, energy intensity, industrial structural shift, industrial activity and final fuel shift. The overwhelming contributors to the change of China's industrial sectors’ carbon emissions in the period 1998–2005 were the industrial activity and energy intensity. This paper also verified that energy intensity was contribute the most to the increase in energy carbon footprint, which indicated the accuracy of LMDI.
**Strategies to decrease carbon footprint**

The random forest algorithm used for calculating energy carbon footprint in Liaoning Province reveals the dynamic mechanism and trance of energy carbon footprint. The result of random forest algorithm shows that carbon footprint of energy carbon footprint grow rapidly. This dynamic mechanism also provides a theoretical support and data basis for the growth of carbon footprint and the transformation of economic development model. The conclusion is that the economic development of the rapid increase in carbon footprint played a crucial role. Based on empirical results, combined with the national conditions and development of advanced foreign experience, it proposes relevant measures to mitigate the carbon footprint growth and development of low-carbon economy recommendations. The following recommendations should be taken into consideration so as to decrease the carbon footprint and promote the national economy for sustainable development.

**Increase energy efficiency and develop low-carbon economy**

Much effort should be made to optimize the energy structures, reduce environmental pollution, develop the green energy, accelerate the modernization of enterprises, increase the quality of economic growth, and finally create a new pattern to develop economy. At the same time, the low-carbon concepts such as green lifestyles, green manufacturers should be advocated and the public awareness of environmental protection should be heightened.

**A collaborative reverse logistics network should be established**

The reverse logistics can be outsourced to those third-party reverse logistics enterprises for expertise. Thus it can be exploited to increase the management efficiency, decrease resource consumption and reduce the management risks. A prompt product recalling system should be built and the renewable materials should be recommended to achieve the environmental protection.

**Design products in an eco-friendly way**

The effect on environment and resources should be completely considered in the life cycle of products. And the waste rate of products should be decreased in order to reduce the adverse effect of products on the environment and maximize the resource utilization rates. Meanwhile, renewable resources and secondary energy should be used to slow down exhaustion of resources and ensure sustainable supply of resources.

**Conclusions**

The goal of this study has been to gain an understanding of the calculation of energy carbon footprint in Liaoning Province. We characterized the energy categories for the calculation of carbon footprints: raw coal, coke, crude oil, petrol, diesel oil, black mineral oil and natural gas. The empirical model subsequently used to calculate the energy carbon footprint in Liaoning Province. In order to evaluate the performance of random forest algorithm, the PLS algorithm is compared with the models proposed in this paper. The results showed that random forest algorithm not only can control parameter deviations, but also it can enhance the precision and accuracy of algorithm.
The utility of random forest algorithm and results proposed in this paper can be extended by examining the energy carbon footprint of the other regions and countries.

As guidance for the future research, the random forest algorithm addressed in this paper can be enhanced to identify the main environmental and social driving factors of energy carbon footprint in Liaoning Province. It could be highly helpful for public sector to formulate energy policies. Besides, random forest algorithm can be used to deal with the other environmental and social numerical problems.

REFERENCES


