BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Electrical Science & Engineering
https://ojs.bilpublishing.com/index.php/ese

ARTICLE

An Approach to Carbon Emissions Prediction Using Generalized Regression Neural Network Improved by Genetic Algorithm

Zhida Guo[1] Jingyuan Fu[2*]

1. School of Economics and Management, Dalian Jiaotong University, Dalian, China
2. Academic Unit of Human Communication, Development, and Information Sciences, Faculty of education, the University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history
Received: 26 March 2020
Accepted: 3 April 2020
Published Online: 30 April 2020

Keywords:
Carbon emissions
Genetic Algorithm
Generalized Regression Neural Network
Smooth Factor
Prediction

ABSTRACT

The study on scientific analysis and prediction of China's future carbon emissions is conducive to balancing the relationship between economic development and carbon emissions in the new era, and actively responding to climate change policy. Through the analysis of the application of the generalized regression neural network (GRNN) in prediction, this paper improved the prediction method of GRNN. Genetic algorithm (GA) was adopted to search the optimal smooth factor as the only factor of GRNN, which was then used for prediction in GRNN. During the prediction of carbon dioxide emissions using the improved method, the increments of data were taken into account. The target values were obtained after the calculation of the predicted results. Finally, compared with the results of GRNN, the improved method realized higher prediction accuracy. It thus offers a new way of predicting total carbon dioxide emissions, and the prediction results can provide macroscopic guidance and decision-making reference for China's environmental protection and trading of carbon emissions.

1. Introduction

With the continuous development of economic globalization and the implementation of strategies of "Introduce In" and "Go out", China is fully integrated into the world in both depth and breadth. Nevertheless, any major domestic challenge is likely to evolve into an international one, such as carbon emissions and energy shortages [1]. Ecological environment and sustainable development are arousing more concerns in the party-government and social sectors, where the work arrangement of carbon emissions and carbon sequestration is much obvious. In the meantime, climate change and relevant environmental events have exerted serious impacts on the sustainable development of human society, thereby becoming a common challenge facing the international community today [2]. Under this context, China can neither do well alone nor undertake international responsibility. In the new era, regardless of the future challenges resulted from carbon emissions, to what extent China can realize accurate predictions require comprehensive research and timely response. The study on accurate prediction of China's future carbon emissions by scientific methods will lay a theoretic basis for relevant policy-making on China's carbon emissions and carbon sequestration.

**Corresponding Author:*
Jingyuan Fu,
Academic Unit of Human Communication, Development, and Information Sciences, Faculty of education, the University of Hong Kong, Hong Kong, China;
Email: jingyuanfu122@163.com

DOI: https://doi.org/10.30564/ese.v2i1.1772

So far, the prediction of carbon emissions has been researched by a large number of scholars at home and abroad. Pan Jinghu et al. [3] accurately identified the spatial characteristics of carbon emissions by building a spatial regression model to simulate the spatial distribution of regional carbon emissions. Wang Yong et al. [4] made scenario predictions of China's industrial carbon emission peaks and assessed their potential mitigation. Tulson Maimaiti et al. [5] established a carbon-emission prediction model based on the generalized neural network targeting agricultural production in Xinjiang; they also carried out a quantified analysis of the factors influencing carbon emissions using the average impact value method. China's carbon emissions in 2010-2050 were predicted by Du Qiang et al. [6] based on an improved Kaya identity model, and by Blanford et al. [7] through MERGE model. STIRPAT models were constructed by Chinese scholars such as Qu Shenning [8], Song Jiekun [9], Zhang Leqing [10], Wang Yanpeng [11], Wang Xianen [12] respectively to predict the carbon emissions of provinces and municipalities in the whole country. Many domestic scholars took the relationship between carbon emissions and energy consumption factors into account. Through the prediction and estimation of the energy consumption structure, a carbon emissions prediction method was constructed [13-16]. Zhao Aiwen et al. [17] performed the prediction of the national carbon emissions by the grey correlation method. Artificial neural network (ANN) and genetic algorithm (GA) have shown great advantages in solving complex and highly nonlinear problems, which hence get widely applied in the fields of control, prediction, and optimization. The generalized regression neural network (GRNN) is a form of neural network with good non-linear approximation performance and robustness. It can also achieve an excellent prediction effect even with unstable or insufficient data. GRNN provides a suitable solution to curve fitting [18-19]. The parameters that GRNN needs to adjust have only one smooth factor, leading to great computing advantages [20]. For GRNN, the network structure and the connection weights between the neurons are determined by the input learning samples. The training of the network is to determine the only parameter smooth factor σ . Specht proposed a method of multiple experiments to determine the range of the smooth factor ($\sigma_{\min}$, $\sigma_{\max}$), to change it progressively in the range. Then these values of equidifferent σ are used in GRNN prediction, and the mean square error between the predicted value and the actual value is taken as the evaluation index. The value of the minimum error is the optimal smooth factor. This is the most popular way of prediction by GRNN in practice, which however fails to obtain satisfactory results in the prediction of carbon emissions. The smooth factor can also be selected by an appropriate objective function and obtained by the optimization method. In this study, a genetic algorithm was used to optimize the smooth factor of GRNN. The shortcomings of traditional prediction theory and methods were overcome by the advantages of the above two methods. The improved algorithm was then applied to the process of carbon emissions prediction for accuracy improvement.

2. Methods and Data

2.1 Improvement of GRNN Model Based on GA Improved Method of Prediction Process

The improved method of the prediction process can be divided into the following two steps: (1) The optimal value of GRNN's unique parameter smooth factor is obtained by genetic algorithms, in which values of each parameter are input, and smooth factor of the population is randomly generated; finally, the optimum values of the smooth factor are output; (2) The optimal values of the smooth factor is transmitted to GRNN which is the main body. The index system data and the optimal smooth factor for carbon emissions prediction are input, and the prediction result of carbon emissions is output[21]. The first step is to achieve an improvement in the genetic algorithm on GRNN. It is designed based on the genetic progress generalized regression neural network model. The basic procedure of the improved prediction method is shown in Figure 1.

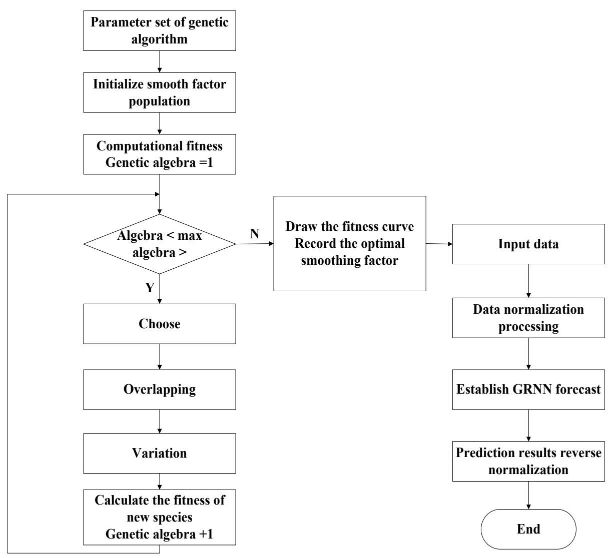

Figure 1. Flow chart of carbon emission prediction method improved based on general regression neural network

2.2 Genetic Improvement of Optimal Smooth Factor

The genetic algorithm of randomly generating individual smooth factor of the population is presented in this section. According to the size and genetic selection, individuals with crossover and mutation are screened. The individuals with the best values are kept, and those lacking adaptations are eliminated. The new group inherits the generation of information and performs better than the previous generation. This procedure is repeated until the condition gets satisfactory[22]. The genetic algorithm is composed of three modules, which are encoding and decoding, individual fitness evaluation and genetic operation, and genetic algorithm (including chromosome selection, crossover, and mutation and so on). Based on the basic genetic algorithm, the following techniques are used for optimization.

2.2.1 Encoding and Decoding

Encoding is the smooth factor of the feasible solution from the solution space to the algorithm that can deal with the search space method. Binary encoding is consistent with the principles of computer processing information, which facilitates the crossover, mutation and other operations of the chromosome[23]. The binary coding mode is carried out among the randomly generated individuals within the prescribed range; each individual corresponds to a value of the smooth sigma factor. In the crossover and mutation operations, the binary value of the operation and the individual's fitness value are directly calculated. The binary value should be decoded to decimal.

2.2.2 Assessment of Fitness

Fitness is used to measure the excellent degree of the optimal solution that each smooth factor individuals in the population may reach or close to. The function of calculating the individuals' fitness is the guided search option evaluation function based on the genetic algorithm. How to construct a fitness function is one of the key problems of the genetic algorithm. In the adaptation degree function, a GRNN is created based on five samples. The target values of the three samples in the trained network are predicted. The predicted results and the actual values of Euclidean distance are taken as the reciprocal of the fitness function. The smaller the Euclidean distance is, the greater the fitness value is. The fitness function is expressed below:

$$f(x) = \frac{1}{\sqrt{\sum_{i=1}^{3}(x_i - X_i)^2}} \tag{1}$$

where x is the predictive value; X is the practical value.

2.2.3 Genetic Operation

(1) Selection

Selection is used to determine the crossover or mutation of the individual operation. In this paper, the selection probability of each chromosome is selected according to the selection probability of each chromosome. The function for calculating the probability of selection is expressed as:

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^{N} f(x_j)} \tag{2}$$

The selection probability of random selected chromosome is obtained through the roulette selection method:

① Generate a uniform distribution of random numbers in the (0,1) interval r ;

② If $r \le q_1$, then the chromosome x_1 is selected;

③ If $q_{k-1} < r < q_k (2 \le k \le N)$, then the chromosome x_k is selected.

The following function is obtained:

$$q_i = \sum_{j=1}^{i} P(x_j) \quad (i = 1, 2, ..., N) \tag{3}$$

where q_i is the cumulative probability of chromosome $x_i = (i = 1, 2, ..., N)$.

(2) Crossover

Crossover refers to the exchange of two genes that are selected from some of the genes on the chromosome so that new individuals are generated by combining information from their parents. The crossover operation is determined by the setting of the crossover probability. In crossover operation, the crossover is randomly generated, and then part of the genes on two chromosomes exchange. Afterward, a new individual is generated. If it is not a crossover operation, then a new individual is the choice of two individuals.

(3) Mutation

Mutation means that the position of one gene on a chromosome gets changed, such as from 0 to 1 or 1 to 0; in fact, it is the change of the offspring gene according to the small probability. Similar to crossover, mutation operation is also determined by the setting of the mutation probability. If so, the mutation is randomly generated, and two new individuals corresponding mutation are formulated after crossover, following update for two individuals; if not, then do not do the operation.

DOI: https://doi.org/10.30564/ese.v2i1.1772

3. Application of Carbon emissions Prediction Algorithm

3.1 Basic Permits

To verify the effectiveness of the algorithm in the theoretical research, the algorithm was programmed under the MATLAB environment, and the MATLAB neural network toolbox was used to build generalized regression neural network prediction model with genetic improvement and the improved generalized regression neural network prediction algorithm. Then carbon emissions were predicted according to the improved prediction model and algorithm[24]. The predicted results were compared with the practical ones, thereby laying a basis for theoretical research in this field.

3.2 Index System

Based on the data collected from 2001 to 2011, the carbon emissions of China from 2019 to 2029 were predicted in this study. Many factors were influencing carbon emissions [25]. Through the gray correlation degree model, those influential factors with the correlation degree above 0.6 were selected to constitute the index system. A total of eight elements, including the proportion of the total respectively, the factors that affect these total population (10,000 people), the proportion of urban residents (%), GDP per capita (yuan), the ratio of GDP to industrial output (%), total energy consumption (million tons of standard coal), the consumption proportion of coal (%), the consumption proportion of oil (%), are taken into account, respectively representing by I_1, I_2, I_3, I_4, I_5, I_6, I_7 and I_8. China's carbon emissions data can be found from the Energy Information Agency (EIA), Carbon Dioxide Information Analysis Centre (CDIAC), International Energy Agency (IEA), World Resources Institute (WRI) and many other international energy agencies, which has no significant difference. In summary, the most authoritative, updated and timely EIA and IEA were selected as sources of carbon emissions data, expressed as V_1 and V_2 respectively. This will make the study more objective, universal and scientific. The arithmetic average value of EIA and IEA carbon emissions was taken as the target value of prediction. The average value of carbon emissions (10000 tons) is expressed as $\overline{V}$. Table 1 lists the detailed information.

3.3 Data Processing

During the prediction process, in most cases, the prediction of the data increment is better than that of the data directly. The incremental data were input and predicted. Those for 8 consecutive years were classified into as a group. The incremental data of each group seven years ago were taken for seeking smooth sigma factor and GRNN training. The 8-year incremental data were used for prediction. In this way, the incremental data were divided into 3 groups, respectively, to predict the carbon emissions during 2019~2029.

To prevent an increase in network training time caused by outlier sample data, normalization processing should be conducted on various data. It aims at converting data to numerical values between 0 and 1.

Table 1. Carbon emissions and the influencing factors during 2019~2029

Serial Number	Year	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	V_1	V_2	$\overline{V}$
1	2019	127627	37.66	8622	39.7	40.5	150406	68.3	21.8	3226.52225	3396.15	3311.336125
2	2020	128453	39.09	9398	39.4	41.5	159431	68	22.3	3422.08573	3605.39	3513.737865
3	2021	129227	40.53	10542	40.5	41.2	183792	69.8	21.2	3959.96585	4176.63	4068.297925
4	2022	129988	41.76	12336	40.8	40.4	213456	69.5	21.3	4596.97027	4837.28	4717.125135
5	2023	130756	42.99	14185	41.8	40.5	235997	70.8	19.8	5116.34858	5403.09	5259.71929
6	2024	131448	44.34	16500	42.2	40.9	258676	71.1	19.3	5575.198	5913.49	5744.344
7	2025	132129	45.89	20169	41.6	41.9	280508	71.1	18.8	5908.42776	6316.44	6112.43388
8	2026	132802	46.99	23708	41.5	41.8	291448	70.3	18.3	6166.56551	6489.98	6328.272755
9	2027	133450	48.34	25608	39.7	43.4	306647	70.4	17.9	6816.09505	6792.94	6804.517525
10	2028	134091	49.95	30015	40.0	43.2	324939	68	19	7446.51986	7252.77	7349.64493
11	2029	134735	51.27	35198	39.8	43.4	348002	68.4	18.6	8126.69441	7954.79	8040.742205

DOI: https://doi.org/10.30564/ese.v2i1.1772

$$x' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

$$y' = y_{\min} + y(y_{\max} - y_{\min}) \tag{5}$$

In equation (4), normalized value x' as the original value x_i , and affect the same element x_i in the minimum value $x_{\min}$, and affect the same element value x_i of maximum value $x_{\max}$. The predictive value obtained ranged from 0 to 1. Therefore, the predicted value y' should be reduced to the actual value by the formula (5). For the normalized value, the minimum value $y_{\min}$ of the network is the direct output value; the maximum value $y_{\max}$ of the network output value.

3.4 Prediction Process

In the use of GA optimizing smooth factor, operating parameters were set as below: smooth sigma factor ranging 0.05-1, precision arithmetic of 0.0001, the population size of 50, the biggest genetic algebra of 12, crossover probability of 0.9 and mutation probability of 0.09. In terms of the fitness function, the 8-year the incremental data of each group were divided into two parts: that of the first 5 years, and that of the last 3 years. The first part was used to train GRNN, and the second part was to predict increment values of the carbon emissions.

3.5 Result Analysis

It was observed that all the maximum values of genetic algorithms in the fitness curves and average fitness degree curves did not appear large-scale shocks, and the convergence of the algorithm was relatively smooth. The population in the evolutionary process of the two curves showed mutual convergence; successive generations of the individuals with the maximum adaptation did not evolve. It suggested that populations reached mature, as shown in Figure 2 and Figure 3.

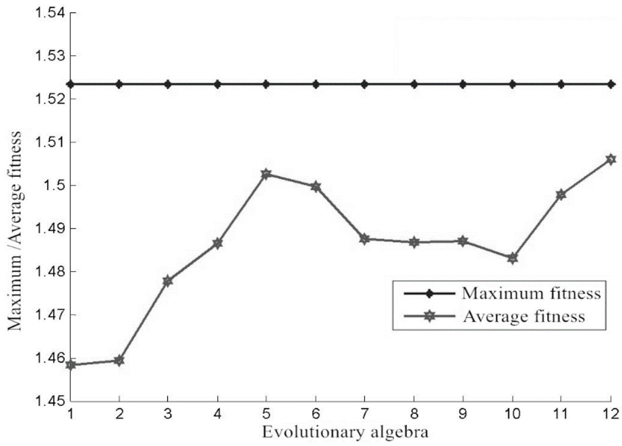

Figure 2. The smooth factor fitness curve of the carbon emission increment in 2019

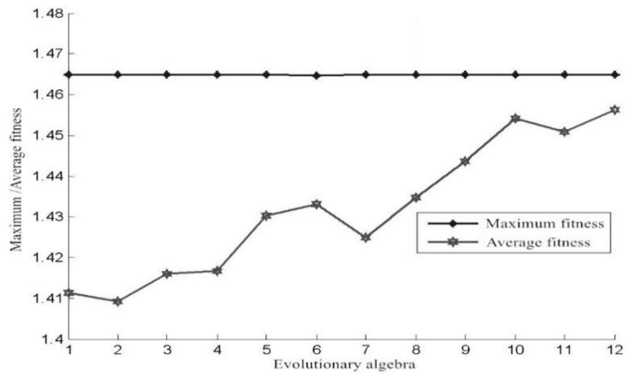

Figure 3 The smooth factor fitness curve of the carbon emission increment in 2029

According to the genetic algorithm, GRNN's smooth factor for predicting carbon emission increment in 2019 and 2029 was 0.084212 and 0.94416. The carbon emissions before the increment from 2019 to 2029 were predicted, as shown in table 2.

The general GRNN was used to fit training samples and determine smooth factor sigma. As a result, the test value of a general GRNN smooth factor sigma was obtained to be 0.1 in the range of 0.1 ~ 0.5. As a result, the predicted

Table 2. prediction results of GRNN and genetic improvement in GRNN

Year	Carbon emissions reference value	Smooth factor σ		Prediction results		Absolute relative error	
		General GRNN	Genetic improvement GRNN	General GRNN	Genetic improvement GRNN	General GRNN	Genetic improvement GRNN
2019	7349.64493	0.1	0.084212	6804.5	7172.60753	7.42%	2.41%
2029	8040.742205	0.1	0.94416	7349.6	7792.18493	8.60%	3.09%
Average						8.01%	2.75%

DOI: https://doi.org/10.30564/ese.v2i1.1772

and actual values of carbon emissions had no remarkable difference, indicating higher prediction accuracy. In the two years of carbon emissions prediction, the relative error of the genetic improved generalized regression neural network was less than 3%. This proved that and the accuracy of carbon emission prediction was greatly improved.

4. Discussion and Conclusions

Accurate prediction of carbon emissions is of great significance for environmental control and the policymaking related to carbon emissions. This study proposed a new method of genetic improvement based on the generalized regression neural network, so as to enhance the accuracy of carbon emission prediction. By directly compiling the MATLAB program code, the genetic improvement based on generalized regression neural network algorithm model was constructed to predict carbon emissions from 2019 to 2029. The comparison clearly demonstrated that the choice of the only parameter smooth factor exerted a significant impact on the prediction accuracy of carbon emissions when the generalized regression neural network was employed. Based on the prediction of data increment, the desired target value was obtained, making the prediction of carbon emissions more sensitive. Hence the prediction results closer to the actual ones were obtained. The algorithm was improved to adapt to the function setting and innovative “5 + 3” pattern was adopted to build a generalized regression neural network (GRNN). A row of five samples was trained, among which the three samples adjacent to the target value were predicted. The predicted result and the actual value were applied in Euclidean distance as the reciprocal of the fitness function. The smooth factor was optimized on the basis of the changing trend of the target data. The predicted results showed that the use of the sample was equal to the number of training. In terms of prediction for the same target value, the GRNN model with genetic improvement outperformed the unimproved ones. The improved method was able to effectively reduce the prediction error of carbon emissions.

References

[1] A. Hu, & W Zhang. Major Challenges and Active Responses in the New Era of China. Journal of Beijing Jiaotong University (Social Science Edition), 2018, 17: 1-8.

[2] J. Wu, J. Qu, H. Li, H. Zhang, L. Xu. Practical Research on Multidisciplinary Cooperative Mechanism of Climate Change, Ecological Economy. 2018, 34: 128-133.

[3] J. Pan, X. Zhao. Spatial Difference Modeling of Carbon emissions in China Based on Spatial Regression Model, Journal of Environmental Science. 2018, 27: 131-140.

[4] Y. Wang, Y. Bi, E. Wang. Scenario Prediction and Emission Reduction Potential Assessment of Industrial Carbon emissions Peak in China, China Population, Resources and Environment. 2017, 27: 131-140.

[5] Tulson M., W. Ding, J. Xie. Carbon emissions Prediction and Influencing Factors Analysis Based on Neural Network. Environmental Engineering, 2017, 35: 156-160.

[6] Du Q., Chen Q., & Lu N. (2012) China’s Future Carbon emissions Forecast Based on Improved IPAT Model. Journal of Environmental Science, 32, 2294-2302.

[7] Geoffrey J. B., Richard G. R. Thomas F. Rutherford. Revised Emissions Growth Projections for China: Why Post Kyoto Climate Policy Must Look East. Project on International Climate Agreements, 2008, 3: 1-26.

[8] S. Qu, Z. Guo. Prediction of China’s Peak Carbon emissions Based on STIRPAT Model[J]. China Population, Resources and Environment, 2010, 20: 10-14.

[9] J. Song. Carbon emissions Prediction Model Based on STIRPAT and Partial Least Square Regression[J]. Statistics and Decision-making, 2011, 24: 19-23.

[10] L. Zhang, S. Chen, W. Wang. Measurement and Trend Prediction of Carbon emissions Effect of Construction Land Change in Anhui Province in Recent 15 Years Based on STIRPAT Model[J]. Journal of Environmental Science, 2013, 33: 950-958.

[11] Y. Wang. Evolution and Forecast of Energy Consumption Carbon emissions in Henan Province. Enterprise Economy, 2013, 6: 26-32.

[12] X. Wang, Y. Wang, H. Duan. Prediction and Controllability of Peak Emissions of Regional Energy Consumption. China Population, Resources and Environment, 2014, 24: 9-16.

[13] B. Wei.Scenario Analysis of Energy Demand and Carbon Dioxide Emission in China. Beijing: China Environmental Science Press, 2007: 75-78.

[14] Yongbin Zhu, Zhen Wang, Li. Pang. Peak Forecast of Energy Consumption and Carbon emissions in China Based on Economic Simulation. Geography Journal, 2009, 8: 935-944.

[15] Sen Bao, Lixin Tian, Junshuai Wang. Prediction of Energy Production and Consumption Trends and Carbon emissions in China. Journal of Natural Resources, 2010, 25: 1248-1253.

[16] Xiaoming Li, Anjian Wang, Wenjia Ding. Global CO2 Emission Trend Analysis Based on Energy Demand Theory. Journal of Earth Science, 2010, 31:

741-748.
[17] Aiwen Zhao, Dong Li. Grey Prediction of Carbon emissions in China. Mathematics in Practice and Theory, 2012, 42: 61-69.
[18] Teo Lian Seng, M Khalid, R. Yusof. Adaptive GRNN for the Modeling of Dynamic Plants. Proceedings of the 2002 IEEE International Symposium on Intelligent Control, 2002, 11: 27-30.
[19] Donald F Specht. A General Regression Neural Network. IEEE Transactions on Neural Networks, 1991, 2: 568-576.
[20] Specht D F. The general regression neural network-Rediscovered. Sustainable Livestock Production & Human Welfare, 1993, 6: 1033-1034.
[21] Wiener, J.B. Climate change policy and policy change in China. UCLA Law Review, 2008, 55: 1805-1826.
[22] Baareh, A. K. Evolutionary design of a carbon dioxide emission prediction model using genetic programming. CARBON, 2018, 9: 298-303.
[23] Aziz, B., Arina, N., & Yee, C. J. Inventory Routing Problem with Carbon Emission Consideration. Matematika, 2019, 35: 39-49.
[24] Cariou, P., Cheaitou, A., Larbi, R., & Hamdan, S.. Liner shipping network design with emission control areas: A genetic algorithm-based approach, Transportation Research Part D: Transport and Environment, 2018, 63: 604-621.
[25] Nagapurkar, P., & Smith, J. D.. Techno-economic optimization and social costs assessment of microgrid-conventional grid integration using genetic algorithm and Artificial Neural Networks: A case study for two US cities. Journal of Cleaner Production, 2019, 229: 552-569.

  DOI: https://doi.org/10.30564/ese.v2i1.1772