Analysis of influencing factors of carbon sequestration based on BP neural network and clustering decision tree model

Xinlin Yang*, Ying Maa

Shenyang Aerospace University School of Energy and Environment Shenyang, Liaoning, 110136

* Corresponding Author Email: 2904294947@qq.com, a2725613450@qq.com

Abstract. Global warming poses a serious threat to life on Earth, carbon sequestration technologies can mitigate its negative effects. In order to optimize the effect of carbon sequestration, firstly, this paper develops a model using BP neural network to determine what forest management plan is the most effective one at sequestering carbon dioxide. On the basis, the best way of using forest is obtained by using the decision tree model, and then we used K-means cluster analysis algorithm combined with confidence intervals andorigin, determined the transition points of different management areas.

Keywords: BP neural network, sparrow multi-objective optimization algorithm, hierarchical analysis, K-means, meta-cellular automata model.

1. Introduction

In an effort to alleviate the negative effect of Gglobal warming, effective measures should be taken In addition to reducing greenhouse gas emissions, we are also expected to find new solutions to increase the stocks of carbon dioxide sequestrated from the atmosphere, especially by biosphere or by mechanical means.

2. Model Building

2.1. Forest Mangement Plan

A carbon sequestration model is developed to discuss the expected effects of forests and their associated products on CO_2 sequestration, and to discuss the most effective forest management plans for CO_2 sequestration.

We first collected the carbon dioxide storage data of a regional forest and analyzed the carbon dioxide storage at different days of the week and different times of the day through MATLAB. The BP neural network prediction algorithm was then used to predict the forest carbon dioxide sequestration for different months of the year and different dates of the month, the BP neural networks includes an input layer,a hidden layer,a activation function and an output layer,which is shown as follows:

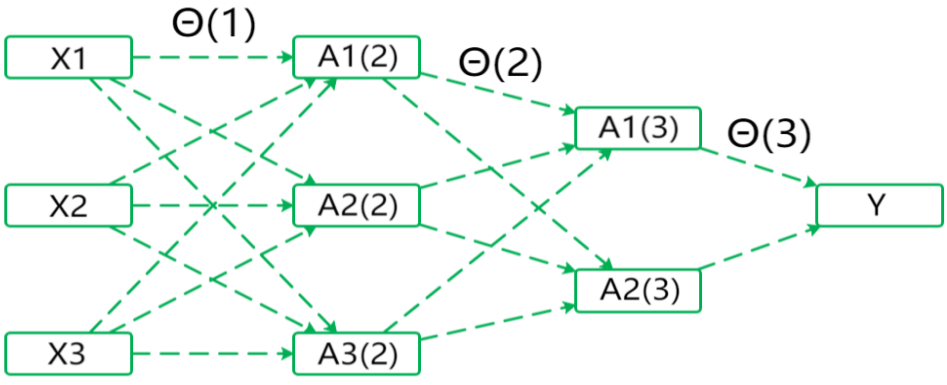

Figure 1. The structure and flow chart of BP neural network

The equation of the Sigmoid is:

$$\mathbf{F(z) = sigmoid(z)= } \left(\mathbf{1 + e^{-z}}\right)^{\mathbf{-1}} \quad (1)$$

The final data screening process produces period prediction images, as shown in the following images.

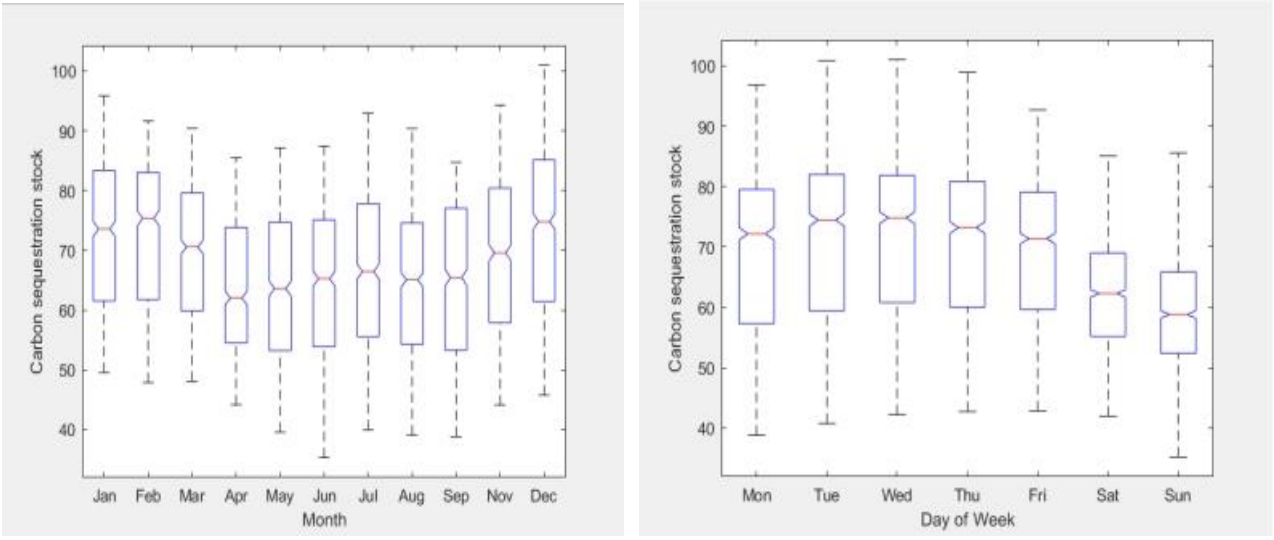

Figure 2. BP neural network prediction diagram

From the above image, we can see that the daily carbon sequestration in this forest area is increasing and then decreasing, and it is at a high and stable level from 8:00 to 18:00 every day, which is consistent with the pattern of plants absorbing carbon dioxide for photosynthesis.

The main plant types and groundwater systems in the forest have different effects on carbon dioxide sequestration in different geographical locations and climate types, and the plant types and groundwater systems further affect the carbon sequestration effect by influencing the plant root system. The better the climatic conditions match the geographical location and the plants growing in the area, the more significant the carbon sequestration effect.

Based on the above analysis, the following forest management plan was derived.

For different tree species in different regions, artificial intervention should be minimized, and carbon sequestration should mainly rely on trees' own photosynthesis, the regulating effect of tree roots on water environment and the self-regulating ability of forest ecosystem. However, some dead trees and trees of a single species due to unreasonable vegetation structure should be cleaned up appropriately and timely to ensure that the environmentally improved trees can better sequester carbon and also to promote wood products to complete carbon sequestration during their life cycle, thus improving the overall carbon sequestration efficiency.

2.2. Decision Model

Forest resource management plans are optimal for carbon sequestration and forest conservation under ideal conditions, but may not meet the needs of human society. Therefore, we try to construct a decision model to evaluate the plan and discuss different types of forest plans. First, the forest tree types can be divided into coniferous, mixed coniferous and deciduous broadleaf, and broadleaf forests.Secondly, we collected the factors affecting carbon sequestration to organize and categorize them, using the main influencing factors discussed above, starting from both biosphere and wood products and analyzing them hierarchically, the results are as figure 3.

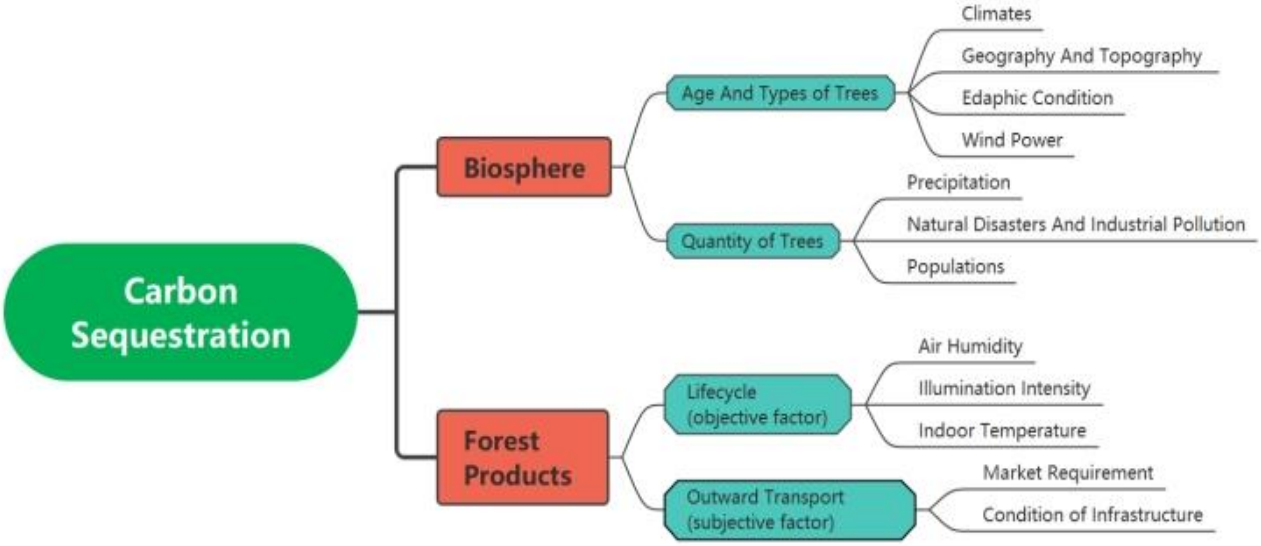

Figure 3. Decision-tree model analysis

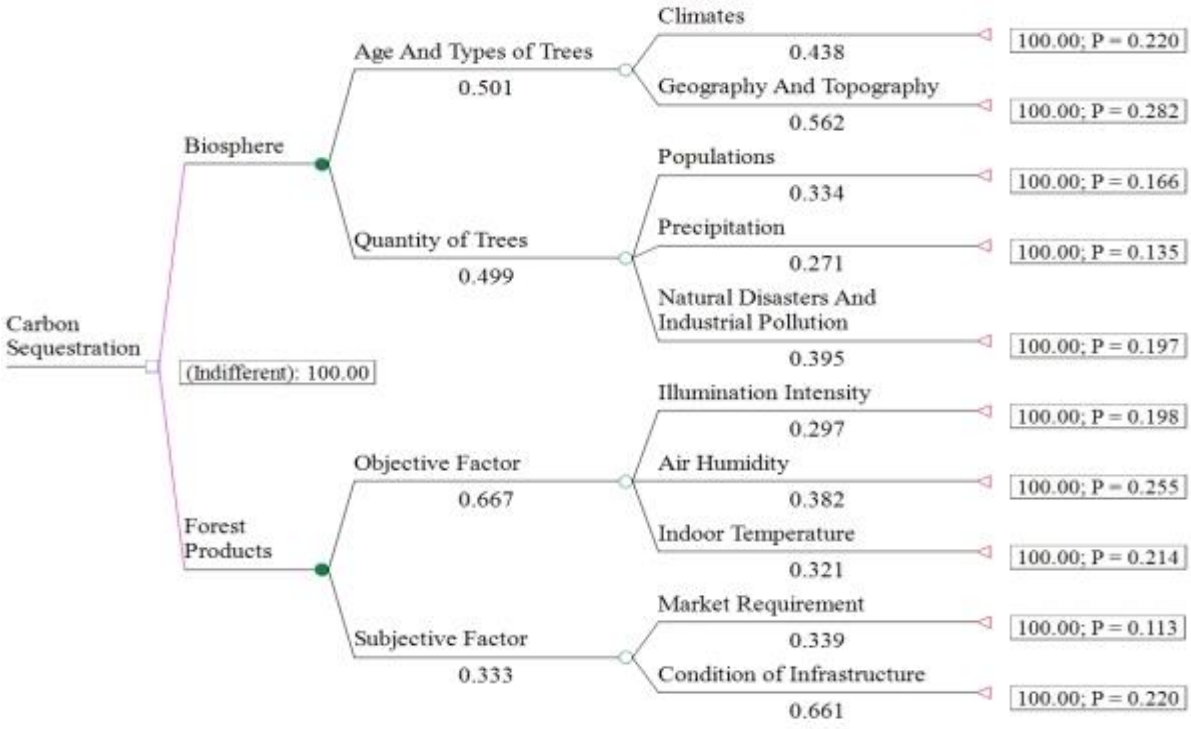

Figure 4. Hierarchical analysis of the related factors

From the above factors based on the analysis of the above principles, we can out we reasonably derive to establish the judgment matrix.

We used MATLAB and used the entropy value method for weight probing, and the weight analysis of the above judgment matrix was performed as follows.

The weights were obtained for indoor temperature as a percentage 9.81%; market requirement 5.17%; illumination intensity 9.09%; air humidity 11.69%; condition of infrastructure 10.10%; geography and topography 15.24%; populations 9.01%;

natural disasters and industrial pollution 10.69%; climates 11.90%; precipitation 7.29% and tested for consistency is follows:

$$CI=(t-n)/(n-1) \quad (2)$$

RI = [0 0 0.52 0.89 1.12 1.26 1.36 1.41 1.46 1.49 1.52 1.54 1.56 1.58 1.59];

CR = CI/RI(n); CI = 0.1259; CR= 0.0845

Because CR < 0.1, the consistency of this matrix turns out to be acceptable with a high degree of feasibility.

Then, we further combine the CART classification algorithm with the obtained weights to recursively divide the independent variable space for the training sample data influencing factors and build a decision tree model in turn, and then prune the tree using the method of validation data to obtain a decision tree classification model that meets the requirements, and select the test attributes according to the Gini coefficient,

If the Gini index of the sample set of influencing factors D: assuming k categories, the value of the gini coefficient of D can be calculated by the following equation.

$$\mathrm{Gini(D)} = 1 - \sum_{k=1}^{k} \left(\frac{|P_k|}{|D|}\right)^2 \tag{3}$$

Here P_k denotes the probability that the selected sample belongs to category k. The probability that this sample is misclassified is (1- P_k)

After the above CART analysis we went through Gini index comparison, combined with its reasonable influencing factors using TreeAge for decision tree analysis and idealization of values.

From the figure we see that geography and topography play an important role in influencing carbon sequestration through biosphere action. Therefore, we selected the cities of Shanghai, Beijing, Ulaanbaatar, Bandar Seri Begawan, Singapore, Manila, Bangkok , New Delhi, Rome, Athens, London, Paris, Berlin, Vladivostok, and Moscow, a total of 15 city, collected information on their altitude, population size, forest cover, and tree species.

We can conclude that the species of trees vary from city to city due to their different geographical locations and topographic conditions. And the species are also closely related to the temperature zones in which each place is located. Secondly, air humidity should also be part of the consideration, it will affect the life cycle of wood products, climate also affects the type and longevity of the number of trees. However, the decision tree model also has certain shortcomings, it ignores the correlation between data, and for data with inconsistent number of samples in each category, the information gain will be biased towards those features with more values, and in order to reduce these objective errors, we collected data on the length and width of different tree leaves and their carbon sequestration effects and used the K-means cluster analysis algorithm and Origin's 3D Confidence-Ellipsoid function to classify the different tree species and create a 3D classification confidence ellipsoid image, as shown in the following figure.

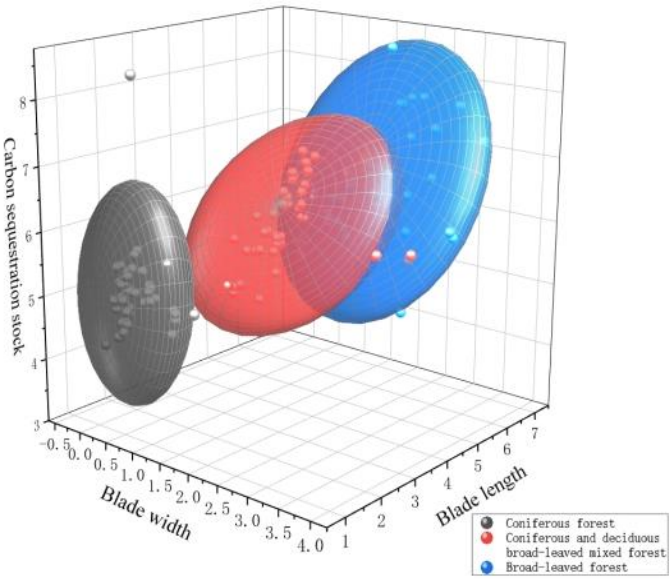

Figure 5. Cluster confidence ellipsoid interval Fig

From the above figure, we can see that different types of trees have different carbon sequestration effects, we should choose a differentiated forest resource management plan for their management.

Due to geographic location and climatic conditions, there may be overlap between some of the differentiated forest management plans, and the forest resources in the overlapping areas are suitable for multiple forest resource management plans. Therefore, in these areas there are transition points for multiple forest resource management plans. We first exclude the impact of heavy metal pollution from industrial production and wind and sand erosion on forest area due to environmental degradation under such conditions. After that, through in-depth research and analysis of the geographical location and climatic characteristics of different areas, we determined that the transition points between different forest resource management plans are basically located in the border area between broad-leaved forest zone and coniferous forest zone, i.e.,mixed coniferous and deciduous broad-leaved forest area.

3. Experiments

Based on the model constructed by the BP neural network prediction algorithm, we estimate the amount of carbon dioxide that will be sequestered by trees and their related products in the forest area selected for our study over a 100-year period. 36525000 kg/ha.

Based on the consideration of maximizing economic returns, we need to analyze the extraction sites in each forest area, and we use a meta-cellular automata model to provide the construction of a raster network map to simulate the process of extracting trees in forest areas and to simulate the extent of extraction by workers.

We then set up specific scenarios for our selected forest zones about differentiated forest resource management plans for different geographical locations, climatic conditions and number of types. In the selected area, divided the area into five functional zones, and developed a differentiated exploitation plan according to the different functions of the forest in the area.

Protected forests have the function of wind and sand control and soil and water conservation, so this part of the forest should be protected, and cutting and destruction of protected forests are strictly prohibited, and regular sampling and investigation of trees in protected forests and replanting measures for dead saplings should be carried out; timber forests as the main exploitation area should be moderately nurtured and intercut to accelerate the growth rate of trees, and thus shorten the tree establishment period, in order to obtain higher Economic forests are mainly used for producing fruits, oil, medicinal herbs and industrial raw materials, and intensive management methods should be adopted for trees in this area. The charcoal forests are mainly used to provide fuel for daily life and industrial production, and the trees in this area are fast-growing, high-yielding and have a large calorific value, so a reasonable exploitation plan should be formulated according to the growth cycle of the trees and crop rotation should be carried out in a timely manner The special-purpose forest is mainly for scientific experiments, and should be protected with the same or even higher protection measures as the protective forest. To sum up, we should divide the trees according to their species and differentiate the mining solution according to their specific use, so to obtain the best environmental protection effect and economic benefit.

Assuming that there are 10 years longer than the previous plan between harvesting time in the best management plan. Among the above five types of forests, timber forests, economic forests and fuelwood forests are used as the subjects of study due to the special nature of protection forests and special-purpose forests themselves that cannot be cut down. The growth cycle of timber forests is 5 to 8 years, while the growth cycle of fuelwood forests is 3 to 5 years. If the harvesting cycle is extended by 10 years, it is obvious that not only the carbon sequestration efficiency but also its economic benefits will plummet. Through general knowledge of geography and the collection of relevant information, it is obtained that the areas where timber forests and charcoal forests are planted are mostly poor mountainous areas where there is a shortage of fossil energy. Farmers in mountainous areas mainly rely on fuelwood forests to obtain fuel energy to meet their daily lives, and rely on the

export of timber forests to obtain economic income. Therefore, in this situation, this group is bound to bear the brunt, so it is especially crucial to change the development mode. In line with the current trend of the booming tertiary industry, during the logging ban, forest managers can negotiate with local residents to improve road transport and other infrastructures, so as to provide a fuller guarantee for timber transportation on the one hand, and lay the foundation for the development of tourism and agriculture on the other hand, as well as develop side businesses such as fungus planting and forest frog breeding to broaden income channels. As for the economic forest, during the ban period, we can speed up the promotion of good seed breeding and cold storage and preservation technology.

4. Test the Models and Sensitivity Analysis

We used the GRNN-PNN joint neural network algorithm to test our model and perform sensitivity analysis. We selected different types of forests (broadleaf, mixed coniferous and deciduous broadleaf, and broadleaf) as the test set for validation, and compared the correct rates under the GRNN and PNN neural network algorithms, and later verified them with the real values. From the results, the accuracy and sensitivity of the model is high as the correct rate of the model is 100% in the GRNN neural network environment and about 93.3333% in the PNN environment.

5. Conclusion

First, a carbon sequestration model was developed using a multi-objective planning algorithm to determine the amount of carbon dioxide that can be sequestered by the forest and its products over time and by temperature, and to determine a forest management plan that is most effective in terms of sequestering carbon dioxide. Next, a decision model was developed using hierarchical analysis, considering alternative ways of valuing the forest, with the aim of allowing forest managers to understand the best use of the forest and to determine a forest management plan that balances the various ways of assessing the forest. The amount of CO_2 sequestered by trees and their related products in this forest area is predicted to be about 36525000kg/ha by BP neural network. We differentiated the specific plan for different geographical and climatic conditions for protection forest, timber forest, economic forest, charcoal forest and special purpose forest zones. If the best management plan requires a longer10 interval between harvesting and harvesting than the original one, then the local area can be promoted by developing tourism, forest frog breeding, fungus cultivation and other industries to broaden the sales channels for local residents.

References

[1] РУТКОВСКАЯД., ПИЛИНЬСКИЙМ., РУТКОВСКИЙ Л.НЕЙРОННЫЕ СЕТИ, ГЕНЕТИЧЕСКИЕ АЛГОРИТ МЫ И НЕЧЕТКИЕ СИСТЕМЫ, 2013.

[2] L. Ingber.Simulated annealing:Practice versus theory.Mathematical and Computer Modelling, December 1993.

[3] Feng YJ, Chen M, Gong Y, Cheng TS, Wang YANYAN. Research progress on dynamic multi-objective optimization evolutionary algorithm. Journal of Automation, Vol. 11, 2020.

[4] Tang H-L, Lin Y-Song, Zhang D-B. Image classification model with feature selection and cluster analysis. Modern Electronics Technology, Vol. 8, 2021.

[5] Zhang Y, Kong XW, Wang ZF, Fu HY, Li M. Cluster analysis based on multi-view matrix decomposition. Journal of Automation, Vol. 12, 2018.

[6] Yao Yuesong, Zhang Xiangyong, Chen Shuai, Deng Che. Decision tree induction algorithm based on attribute purity. Journal of Computer Engineering and Design, Vol. 1, 2021.

[7] Liu Z. Y., Chu N.. A weighted clustering division decision tree algorithm. Telecommunications Technology, Vol. 11, 2020.

[8] Ye R D, Wang Z C, Luo K K, Lin Y. Bootstrap confidence intervals for multiple partial-normal overall common location parameters. Journal of Mathematical Physics, Vol. 1, 2021.

[9] Liu LJ, Hu JW, Sun HC. Adjustment method of judgment matrix in hierarchical analysis method. Journal of Weaponry Equipment Engineering, Vol. 2, 2020.

[10] Ying, Yu-Ti, Xiang, Ting-Ting, Lin, Wei-Lei, Wu, Jia-Sen, Yang, J., Jiang, Pei-Kun. Sequestration characteristics of phytosilica carbon in the apoplastic layer of five forest stands in subtropical China. Forestry Science, Vol. 3, 2015.

[11] Zhang X, Li Y, Ma Q, Liu LN. Research on the Development of Carbon Capture and Utilization and Storage Technology in China. Chinese Engineering Science, Vol. 6, 2021.