Short-term photovoltaic output prediction based on advanced prediction error NGA-ELMAN cascade neural network

Weihui Xu[1] and Zhaoke Wang[1]

[1]North China University of Water Resources and Electric Power - Huayuan Campus

March 24, 2024

Abstract

Improving the accuracy of photovoltaic power prediction is crucial for grid scheduling planning and is essential for the safe, stable, and economic operation of power systems. Based on the statistical characterization of the data, a two-stage PV power prediction model with error correction is developed. First, an Elman neural network model optimized by a small habitat genetic algorithm is introduced; subsequently, a more accurate model for the preliminary prediction error probability distribution is established, based on its distribution characteristics. This model aims to achieve error correction of the preliminary prediction results. The empirical results, derived from actual PV power curves and meteorological data, demonstrate the effectiveness of the proposed method.

Short-term photovoltaic output prediction based on advanced prediction error NGA-ELMAN cascade neural network

Weihui Xu[1],Zhaoke Wang[1]

North China University of Water Resources and Electric Power,zhengzhou , 450003,China

Email: keelectrical@yeah.net

Abstract: Improving the accuracy of photovoltaic power prediction is crucial for grid scheduling planning and is essential for the safe, stable, and economic operation of power systems. Based on the statistical characterization of the data, a two-stage PV power prediction model with error correction is developed. First, an Elman neural network model optimized by a small habitat genetic algorithm is introduced; subsequently, a more accurate model for the preliminary prediction error probability distribution is established, based on its distribution characteristics. This model aims to achieve error correction of the preliminary prediction results. The empirical results, derived from actual PV power curves and meteorological data, demonstrate the effectiveness of the proposed method.

Introduction: With the advancement of technology and reduction in costs, photovoltaic (PV) power generation has emerged as an important component of the renewable energy sector [1]. However, the power output from PV systems is significantly influenced by environmental factors, rendering its prediction highly uncertain and thereby impacting the stability and dispatch efficiency of the power grid. Current prediction methods comprise physical models [2], statistical analysis [3], and machine learning, with the latter gaining prominence in this field due to its ability to handle complex data and resist environmental disturbances [4]. The advantage of machine learning methods lies in their capability to achieve more reliable grid operation and efficient energy utilization through refined prediction methods. Strategies to improve the predictive performance of machine learning models include data preprocessing such as data cleaning and standardization [5], feature engineering [6], selecting appropriate models and optimizing hyperparameters [7, 8], and applying ensemble learning methods. For deep learning, this also includes adjusting network architecture, applying regularization techniques, and tuning the learning rate [9]. These strategies work together to enhance the model's prediction accuracy and generalization ability. Despite significant research achievements in the field of PV power prediction, exploring new methods to further enhance predictive capability remains of great research value. This paper proposes a two-stage PV power prediction model aimed at further improving the accuracy of PV power generation predictions by combining advanced machine learning techniques with error correction mechanisms. Empirical analysis shows that this model brings an innovative research method to the field of PV power prediction, demonstrating its potential in improving prediction accuracy.

Adaptive Genetic Algorithm: Genetic algorithms derive their foundation from the core concepts of Darwin's theory of evolution. They simulate real-world problems on a computer, mimicking the process of chromosome selection, crossover, and mutation in biological evolution, as illustrated in Figure 1.

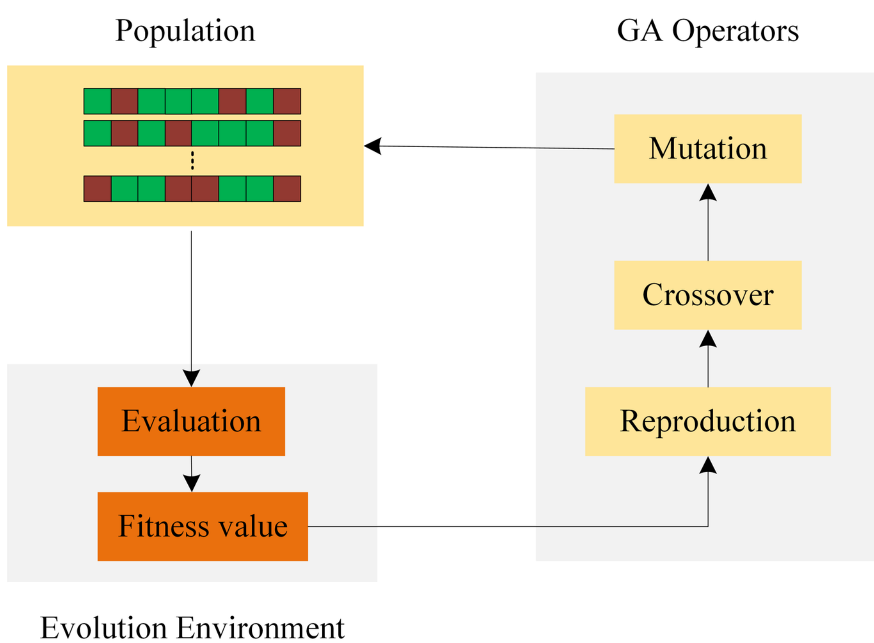

Fig 1 *Genetic algorithm flow chart*

This paper incorporates adaptive crossover and mutation operators proposed by Srinivas, and uses the Sigmoid growth curve as the adaptive adjustment curve for crossover rate Pc and mutation ratePm . This approach is used to improve these rates, with the mathematical expressions as follows:

In the formula: P_{c_max} is the maximum crossover probability; P_{c_min} is the minimum crossover probability; F_{max} is the maximum fitness value of each generation; F_{avg} is the average fitness value of each generation; F_{fit} is the individual fitness value; P_{m_max} is the maximum mutation probability; P_{m_min} is the minimum mutation probability.

Adaptive Dynamic Niche Radius Technique: Niche technology is a concept derived from the natural phenomenon of like attracting like. In dealing with multimodal problems, niche technology helps maintain the diversity of solutions, thereby reducing the probability of falling into local optima. It primarily comprises two components: population division and individual fitness updating. Initially, niche populations are divided based on genetic similarity between individuals, with each niche undergoing its genetic evolution operations independently, and the optimal individual functioning solely within its own group. Subsequently, each individual in the niche updates their fitness based on the sharing function, with the updated fitness determining the optimal individual.

Hamming distance $\mathrm{d_{ij}}$ is a standard reflecting the genetic similarity between individuals. The division rule is as follows:

In the formula: x_{ik} andx_{jk} represent the k-th variable of individuals i and j, respectively; m is the number of variables for each individual; N is the population size; $\sigma_{\mathrm{rad}}$is the niche radius. When$d_{ij} < \sigma_{\mathrm{rad}}$, the individuals are classified into the same niche.

This paper adopts the niche technology based on fitness sharing proposed by Goldberg. This mechanism adjusts individual fitness through the sharing function $f_s(d_{ij})$, which reflects the degree of closeness between individuals within a niche. The mathematical expression for this is as follows:

In the formula: fi is the fitness of individual i;f_i ' is the shared fitness of individual i;$f_s(d_{ij})$ is the fitness sharing function; N is the number of individuals within the niche.

The niche radius serves as a crucial criterion for dividing niches. In conventional algorithms, the use of a fixed niche radius is common. However, with the Hamming distance between individuals decreasing in the later stages of evolution, a consequent reduction in the number of niches occurs, adversely affecting population diversity. Therefore, this study introduces an adaptive niche radius, delineated by the following formula:

In the formula: t represents the iteration number; $\Delta\delta$ is an adaptive adjustment factor. When the number of niches becomes too small in the later stages, $\Delta\delta$ will increase appropriately, thereby reducing the niche radius, which in turn increases the number of niches.

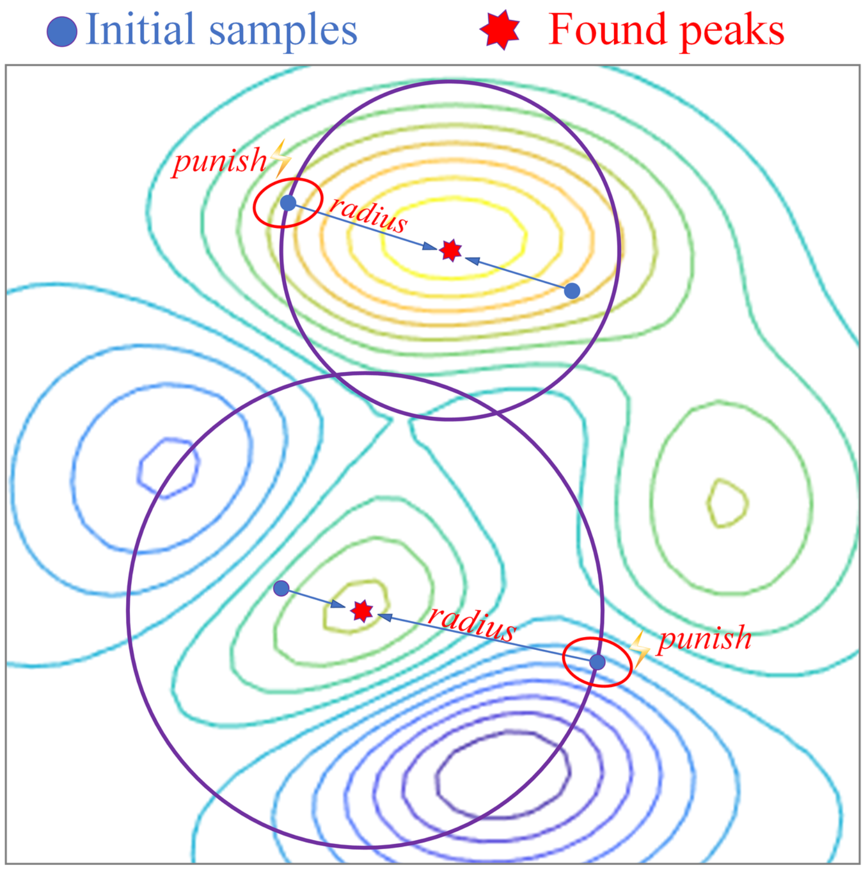

Fig 2 *The principle diagram of the niche*

ELMAN neural network based string level over prediction error modeling: This study introduces a cascaded lead forecasting error network, anchored in the ELMAN neural network architecture, and consisting of three distinct ELMAN neural networks, each assigned a unique function: the historical data error generation network (Network 1), the lead forecasting error network (Network 2), and the prediction output network (Network 3). The core idea behind this model's design involves using the error between Network 1's predicted values and the actual values as the target output for training Network 2. Network 2's testing set inputs mirror those of Network 3, rendering Network 2's output the lead forecasting error of Network 3's output, namely, predicting the error between Network 3's predicted values and their actual values in advance. Consequently, the subtraction of Network 2's forecasted error from Network 3's predicted output is aimed at reducing the actual error. The model's principle is depicted in Figure 3.

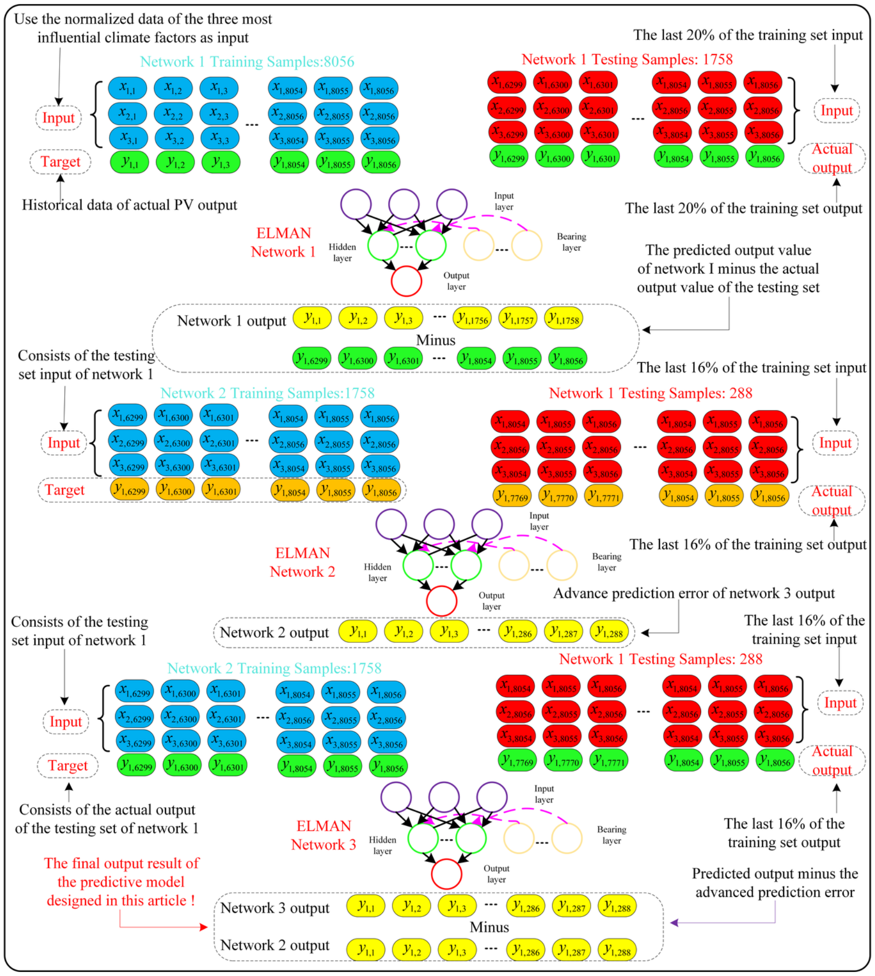

Fig 3 *Model schematic diagram*

Taking January in winter as an example, the training set for Network 1 is composed of the first 8056 samples representing the three most significant historical climatic factors and historical PV output data; the testing set includes the last 20% of the training set. The training set inputs for Network 2 mirror the testing set inputs of Network 1; the target outputs for the training set consist of the predicted PV outputs from Network 1 minus the actual PV output values; the testing set constitutes the last 16% of the training set; the output values represent the advance prediction errors of PV output predicted by Network 3. The training set for Network 3 is the same as the testing set for Network 1; the testing set is the last 16% of the training set. The final model output is the PV output predicted by Network 3 for a certain day minus the advance prediction error from Network 2, thereby achieving improved prediction accuracy of the model.

The principle of using NGA-ELMAN for predicting photovoltaic (PV) output involves representing weights and biases with individuals in a genetic algorithm and optimizing the ELMAN neural network by continuously searching for and updating the optimal individual through evolutionary operations and niche technology, aiming to minimize the probability of falling into local optima. The implementation steps are as follows:

Step 1: Identify the main climatic factors affecting PV output in each season based on Pearson correlation analysis.

Step 2: Normalize the data and divide it into training and testing sets.

Step 3: Determine the structure of the ELMAN neural network and the parameters of the algorithm.

Step 4: Initialize the population using floating-point encoding, where the length of each individual is the sum of the total number of weights and thresholds in the network, given by the formula $C_{\text{len}}=I_{\text{in}}H_{\text{hid}}+H_{\text{hid}}+H_{\text{hid}}H_{\text{out}}+H_{\text{out}}$. Here, in I_{in} represents the number of neurons in the input layer, H_{hid} the number of neurons in the hidden layer, and H_{out} the number of neurons in the output layer.

Step 5: Set the fitness function as the mean absolute error between the network's predicted values and the actual values. The mathematical expression is: , where m is the number of nodes in the network's output layer; y_i is the predicted output value of the network; y_t is the actual value; φrepresents a threshold function.

Step 6: Calculate the fitness of all individuals using training data, and employ an elitist retention strategy to select the top M individuals with the highest fitness. Perform selection and elimination on all individuals in the population using the roulette wheel algorithm, where the probability of an individual being retained for the next generation is .Conduct crossover and mutation operations on individuals to generate the offspring population. Calculate the fitness of the offspring individuals. Add the retained M individuals to the offspring population and adjust the fitness of individuals using the niche technique. Employ a tournament selection mechanism to select the top P individuals to form a new population. Repeat the above steps until the convergence accuracy is satisfied or the maximum number of iterations is reached, to obtain the optimal individual.

Case Study Analysis: This paper selects data from the Australian Solar Photovoltaic Research and Development Center, covering the period from 0:00 on January 1, 2018, to 0:00 on December 30, 2018. The dataset consists of photovoltaic output and various climatic factors, with a sample point every 5 minutes throughout the year.

The number of neurons in the hidden layer exerts a significant influence on network performance, making the selection of an appropriate number of nodes imperative. This study identifies the optimal number of nodes using an empirical formula and conducts an evaluation based on the test error of the ELMAN network. The empirical formula is specified as follows:

where m is the number of nodes in the input layer; h is the number of nodes in the output layer; a is an integer between [1,10]. The prediction error of Elman neural networks with different numbers of neurons is shown in Figure 4, with the optimal number of neurons for the hidden layer determined to be $n_{\text{hid}}=11$.

Hosted file

`image13.emf` available at https://authorea.com/users/758834/articles/733023-short-term-photovoltaic-output-prediction-based-on-advanced-prediction-error-nga-elman-cascade-neural-network

Fig 4 *Optimization of hidden unit number*

This paper sets the learning rate of the ELMAN neural network to 0.01, the training times to 1000, the target minimum error to 0.00001, the momentum factor to 0.01, and the minimum performance gradient to 0.00001.

Given the training duration of the algorithm model, this paper specifies the number of evolutionary iterations at 100 generations, the population size at 40, the maximum crossover rate at 0.85, the minimum crossover rate at 0.55, the minimum mutation rate at 0.35, and the number of individuals retained through the elitist strategy at 20.

To confirm the effectiveness of the proposed NGA-EMANCNN neural network prediction model, this study utilizes various statistical error measurement methods to evaluate the performance of the prediction model, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

NGA-LMANNCNN Photovoltaic Output Prediction Model Results: The paper divides the dataset into training and testing sets at an 8:2 ratio, using the training set to develop the prediction model and the testing set to assess the model's performance and practicality subsequent to training. Given the runtime of the program, only the third ELMAN sub-neural network is optimized using NGA.

Taking January as an example, the output results of Network 1 on the test set are shown in Figure 5, the error figure is shown in Figure 6, and the calculation results of the model measurement index are shown in Table 1.

Hosted file

`image14.emf` available at https://authorea.com/users/758834/articles/733023-short-term-photovoltaic-output-prediction-based-on-advanced-prediction-error-nga-elman-cascade-neural-network

Fig 5 *Model Prediction Comparison Chart*

Hosted file

`image15.emf` available at https://authorea.com/users/758834/articles/733023-short-term-photovoltaic-output-prediction-based-on-advanced-prediction-error-nga-elman-cascade-neural-network

Fig 6 *Model error comparison diagram*

Table 1 Model error results

Index	error correction	No correction
RMSE/kW	4.16	4.95
MAE/kW	2.27	2.92
MAPE/%	0.13	0.17

The comparison indicates that the MGA-ELMAN model without error correction exhibits a certain model bias, with overall predictions being lower than the actual values. The cascaded error correction model proposed in this paper can mitigate the model bias inherent in the preliminary prediction model, significantly enhancing the accuracy of the predictions.

Conclusion: Accurate prediction of PV output enables the power dispatch department to timely adjust the scheduling plan, significantly enhancing the acceptance of PV grid connection and the stability and security of the power system. In this paper, an error-corrected PV output prediction model using a serial MGA-Elman neural network is established, based on the PV output data from actual PV power stations. The effectiveness of the serial error-corrected prediction model is verified through comparison with the MGA-Elman error-corrected prediction model, thereby providing a reference for predicting PV output.

Reference

1. Ding, H., Zhou, D., Liu, G., and Zhou, P., 'Cost Reduction or Electricity Penetration: Government R&D-Induced Pv Development and Future Policy Schemes', Renewable and Sustainable Energy Reviews, 2020, 124, p. 109752.

2. Mayer, M.J. and Gróf, G., 'Extensive Comparison of Physical Models for Photovoltaic Power Forecasting', Applied Energy, 2021, 283, p. 116239.

3. Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., and Lughi, V., 'Advanced Methods for Photovoltaic Output Power Forecasting: A Review', Applied Sciences, 2020, 10, (2), p. 487.

4. Mohamad Radzi, P.N.L., Akhter, M.N., Mekhilef, S., and Mohamed Shah, N., 'Review on the Application of

Photovoltaic Forecasting Using Machine Learning for Very Short-to Long-Term Forecasting', Sustainability, 2023, 15, (4), p. 2942.

5. Ge, L., Du, T., Li, C., Li, Y., Yan, J., and Rafiq, M.U., 'Virtual Collection for Distributed Photovoltaic Data: Challenges, Methodologies, and Applications', Energies, 2022, 15, (23), p. 8783.

6. Li, D.H., Cheung, G.H., and Lam, J.C., 'Analysis of the Operational Performance and Efficiency Characteristic for Photovoltaic System in Hong Kong', Energy Conversion and Management, 2005, 46, (7-8), pp. 1107-1118.

7. Sun, W., Zhang, T., Tao, R., and Wang, A., 'Short-Term Photovoltaic Power Prediction Modeling Based on Adaboost Algorithm and Elman', in, 2020 10th International Conference on Power and Energy Systems (ICPES), (IEEE, 2020)

8. Ma, X. and Zhang, X., 'A Short-Term Prediction Model to Forecast Power of Photovoltaic Based on Mfa-Elman', Energy Reports, 2022, 8, pp. 495-507.

9. Li, G., Xie, S., Wang, B., Xin, J., Li, Y., and Du, S., 'Photovoltaic Power Forecasting with a Hybrid Deep Learning Approach', IEEE Access, 2020, 8, pp. 175871-175880.