Journal of Physics: Conference Series

Optimized Transfer Learning Based Short-term Electrical Load Interval Prediction

To cite this article: Qinyi Bai *et al* 2023 *J. Phys.: Conf. Ser.* **2661** 012033

Optimized Transfer Learning Based Short-term Electrical Load Interval Prediction

Qinyi Bai[1, a], Dongbin Huang[1, b], Ling Xiao[2, c*]

[1]Department of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[2]Department of Mathematics and Statistics, The Xuzhou University of Technology, Xuzhou, 221018, China

[a]Email: baiqinyi9911@163.com; [b] Email: huangdb@cqupt.edu.cn; [*c] Email: xiaoling5231@163.com

Abstract: Electrical load forecasting is an essential foundation for power reliable and economical operation of the power grid. Most forecasting models regard the prediction results as deterministic variables, which ignores the randomness and volatility of the power load. At the same time, insufficient historical load data often lead to undertrained models, which affects the accuracy of capturing uncertain information. Therefore, we proposed an optimized transfer learning-based method for short-term load-interval prediction. A deep learning quantile regression model would be constructed by source domain data in the method, and the weights of the source model would be optimized to avoid negative transfer. Then, the target model is constructed by parameter transfer based on key layers and is tuned with hyperparameters by target domain data. From the experimental discussion, it is known that the model with an optimized transfer learning strategy can accurately quantify the fluctuation range of future power load.

1. Introduction

The entry of renewable energy with intermittent nature into the grid system will bring uncertainty. This can also make the supply and deployment of electricity increasingly complex, which brings significant new challenges to power load forecasting. These challenges also make uncertain information essential in decision-making in power systems.

Short-term load forecasting (STLF) focuses on addressing daily operations, system safety analysis, and scheduling maintenance [1]. The research on STLF approaches can be classified mainly into traditional technologies and emerging approaches [2]. The majority of the existing statistical methods use linear models [3], and the electric load is susceptible to external factors such as weather and holidays [4], so the accuracy can be limited in a fluctuating environment.

With the advancement of artificial intelligence, machine learning models including support vector machines [5], radial basis function neural networks [6], random forests [7], and deep learning models [8] have better nonlinear fitting capability than traditional technologies due to their sensitivity to data features and more complex internal structure [9].

Probabilistic forecasting can effectively capture the uncertainty information of load forecasting. Traditional point forecasting models can only obtain the exact forecast value at every step. In contrast, the interval forecasting method in probabilistic forecasting can predict the future fluctuation range of actual values [10], which is a useful tool for quantifying uncertainty information. Due to its more adaptable and effective features, quantile regression is a significant extension of conventional mean regression in interval prediction. It can directly calculate the distribution function and quantile and can be successfully combined with neural networks to improve the accuracy of interval prediction [11].

However, existing STLF models generally rely on a great deal of historical load data to train. Transfer learning becomes a solution to the problem of difficult access to realistic data [12]. On the one hand, the time cost is reduced by pre-training the models [13]. On the other hand, correlated data patterns are shared by constructing feature spaces [14]. The accuracy of mapping data for load prediction can be improved, however, this has not been addressed in studies on probabilistic prediction.

A short-term load-interval prediction approach based on optimized transfer learning is proposed in this study and is motivated by the notions mentioned above. The remainder of the essay is structured as follows. Quantile regression and transfer learning methodology are covered in Section II. The interval prediction approach based on optimized transfer learning is described in depth in Section III. Section IV analyzes and discusses the experimental results. Section V provides a summary of this work.

2. Related work

The related work in this paper is divided into two subsections. The benefits of the approach utilized in this work are explained in Section A, which also presents the ideas behind transfer learning. Section B introduces the basic definition of quantile regression.

2.1 Transfer learning

Transfer learning (TL) improves the learning task in the target domain by sufficiently learning historical data of the domain or task in question to absorb knowledge [15]. This approach relaxes the condition that traditional machine learning must have sufficient training samples and extends machine learning for practical applications. It is currently widely used in artificial intelligence, such as computer vision [16], natural language processing [17], and time series forecasting [18].

$$[\omega_S, \omega_T] \rightarrow \Omega_T, \omega_S \in \Omega_S \tag{1}$$

where ω_S and Ω_S are the knowledge and task of the source domain respectively, ω_T and Ω_T are the knowledge and task of the target domain respectively.

Due to the difficulties in getting them, there aren't enough genuine data available for the job or topic in question. To address the issue of data scarcity and because the majority of tasks or data are relevant, transfer learning can be used to share instances, parameters, features, and relationships to new models in a certain way to accelerate and optimize the learning of models. By identifying common or a priori parameters between different models, parameter transfer, sometimes referred to as model transfer, can assist reduce the target model's training time and increase its prediction accuracy.

2.2 Quantile regression

It is frequently challenging to map the more dispersed distribution using traditional mean regression. To correctly portray the mapping information of the explanatory factors to the response variables at different quantile points, we used quantile regression (QR) [19]. Assuming a linear relationship between the independent variable $X = [X_1, X_2, \dots, X_k]'$ and the dependent variable Y, quantile regression can characterize the mapping information of each particular explanatory variable X_i to the responsive variable Y_i at different quantile points $\mu(0 < \mu < 1)$with the following expressions:

$$Q_{Y_i}(\mu|X) = X'\alpha(\mu) + \varepsilon(\mu) \tag{2}$$

where $Q_{Y_i}(\mu|X)$ denotes the μ-the conditional quantile of the responsive variable Y_i under the explanatory variable X_i; $\alpha(\mu)$ denotes the vector of regression coefficients based on the conditional quantile point μ, $\alpha(\mu) = [\alpha_1(\mu), \alpha_2(\mu), \dots, \alpha_k(\mu)]'$, k is the number of quantile points; $\varepsilon(\mu)$ denotes the error of different quantile points μ. $\alpha(\mu)$ can be obtained according to the following equation:

$$\hat{\alpha}(\mu) = \arg min \sum_{i=1}^{S} \delta_\mu(Y_i - X_i'\alpha) \tag{3}$$

where s is the sample size; $X_i = X_{1i}, X_{2i}, \dots, X_{ki}$; $\delta_\mu(Y_i - X_i{}'\alpha)$ is the asymmetric loss function at the conditional quantile point μ, which is expressed as a segmented linear function as:

$$\delta_\mu(Y_i - X_i{}'\alpha) = (Y_i - X_i{}'\alpha)(\tau - I(Y_i - X_i{}'\alpha)) \tag{4}$$

where $I(Y_i - X_i{}'\alpha)$ is the indicator function as follows:

$$I(Y_i - X_i{}'\alpha) = \begin{cases} 0, Y_i - X_i{}'\alpha \geq 0 \\ 1, Y_i - X_i{}'\alpha < 0 \end{cases} \tag{5}$$

3. Methodology

3.1 LSTMQR model

The explanatory and responsive variables of real problems mostly show nonlinear relationships. Long and short-term memory neural networks (LSTM) [20] can well estimate nonlinearity and its gate structure can better handle long-term dependence information. Therefore, quantile regression combined with LSTM neural network can well quantify the uncertainty information of nonlinear relationships. The quantile loss function is as follows:

$$Loss = min_{z,d} \sum_{i=1}^{s} \delta_\mu \left(Y_i - g\left(X_i, \hat{Z}, \hat{d}\right)\right) \tag{6}$$

where $\delta_\mu(\sigma)$ is the loss function shown in Equation (4), and $\hat{Z}$ and $\hat{d}$ are the weights and biases of the LSTM, respectively. The function $g\left(X_i, \hat{Z}, \hat{d}\right)$ is optimized by continuously adjusting the weights and biases, and the expression of the optimal quantile at quantile point μ is as follows:

$$\hat{Q}_Y(\mu|X) = g\left(X, \hat{Z}_{(\mu)}, \hat{d}_{(\mu)}\right) \tag{7}$$

3.2 The optimized transfer learning model

The LSTMQR model has initial weights that are typically created at random, and interval prediction performance is directly correlated with these weights. The genetic algorithm can help the model to obtain reliable initial parameters with its powerful global search capability [21]. The selection of the fitness function directly affects the genetic algorithm's convergence speed and global search ability. If the fitness evaluation function is not properly designed, a series of deception problems that prevent the generation of individuals with high fitness can occur, such as premature convergence and evolutionary stagnation [22].

Therefore, this work adopts GASA to refine the initial weights matrix of LSTMQR to avoid the deception problem [23]. The minimization interval prediction evaluation is used as the fitness evaluation function, and the optimization search process is as Algorithm 1.

The optimization process of the LSTMQR model based on GASA:
Input: population size $\boldsymbol{M}$, source domain dataset $\boldsymbol{X}$, termination of evolution $\boldsymbol{G}$ **Output:** optimal individual $\boldsymbol{o}^*$
1. Chromosomes encoded as a vector of 800×1; 2. Initialized population $\boldsymbol{p(0) = P_{800\times M}}$; 3. $\boldsymbol{F \leftarrow CWC_{min}}$; 4. $\boldsymbol{g = 1}$; 5. $\boldsymbol{While\ g < G}$ 6. $\quad \boldsymbol{for\ m = 1, \dots, M}$ 7. $\quad\quad \boldsymbol{f(0) \leftarrow F(p(0), X)}$; 8. $\quad \boldsymbol{End\ for}$

$\boldsymbol{o} \leftarrow \boldsymbol{Select}(\boldsymbol{p}(\boldsymbol{0}), \boldsymbol{f}(\boldsymbol{0}), \boldsymbol{G_{gap}})$;
$\boldsymbol{o} \leftarrow \boldsymbol{Recombin}(\boldsymbol{o}, \boldsymbol{P_c})$;
$\boldsymbol{o} \leftarrow \boldsymbol{Mutate}(\boldsymbol{o}, \boldsymbol{P_m})$;
$\boldsymbol{f'}(\boldsymbol{0}) \leftarrow \boldsymbol{F}(\boldsymbol{o}, \boldsymbol{X})$;
$\boldsymbol{o} \leftarrow \boldsymbol{Neighbor}(\boldsymbol{o}, \boldsymbol{p_S}, \boldsymbol{p_R}, \boldsymbol{p_I})$;
$\boldsymbol{f''}(\boldsymbol{0}) \leftarrow \boldsymbol{F}(\boldsymbol{o}, \boldsymbol{X})$;
According to $\boldsymbol{p_D} = \begin{cases} \boldsymbol{1} & , \boldsymbol{\Delta E} \geq \boldsymbol{0} \\ \boldsymbol{exp}(-\frac{\boldsymbol{\Delta E}}{\boldsymbol{T}}), & \boldsymbol{\Delta E} < \boldsymbol{0} \end{cases}$ to determine whether to update; /*where $\boldsymbol{\Delta E} = \boldsymbol{f''}(\boldsymbol{0}) - \boldsymbol{f'}(\boldsymbol{0})$; $\boldsymbol{T}$ is the temperature */
$\boldsymbol{o} \leftarrow \boldsymbol{Reins}(\boldsymbol{o}, \boldsymbol{p}(\boldsymbol{0}))$;
$\boldsymbol{f'''}(\boldsymbol{0}) \leftarrow \boldsymbol{F}(\boldsymbol{o}, \boldsymbol{X})$;
$\boldsymbol{if}\ \boldsymbol{f'''} > \boldsymbol{f^*}$
  $\boldsymbol{f^*} = \boldsymbol{f'''}$;
  $\boldsymbol{o^*} \Leftarrow \boldsymbol{o}$;
$\boldsymbol{else}$
  $\boldsymbol{o^*} \Leftarrow \boldsymbol{p}(\boldsymbol{0})$;
$\boldsymbol{End\ if}$
$\boldsymbol{End\ While}$

The proposed method in this paper combines deep learning and quantile regression to quantify the uncertainty information of the electric load and uses GASA to optimize the performance of transfer learning. The prediction framework mainly consists of an optimized pre-training stage and a fine-tuning stage.

(1) Optimized pre-training stage

In this phase, we will construct a deep learning quantile regression model, train it using data from the source domain, and then implement the GASA algorithm to optimize its initial weight matrix (considering that the learned knowledge in transfer learning will be reflected in the weight matrix, so this paper chooses to optimize the model in the pre-training rather than the fine-tuning phase). The parameters are output when the global optimal solution is obtained and the model is retrained with the optimal initial weights to fully learn the knowledge of the distribution features of the source data.

Meanwhile, considering the complex structure of the LSTMQR model, the key layers including the input and hidden layers of the source model are frozen. In this way, the integrity of the parameters is ensured.

(2) Fine-tuning stage

A new LSTMQR model will be constructed as the target model in this phase. The corresponding network layers in the target model are replaced by unfreezing the transfer layers obtained in the previous stage. Then, the weights of the fully connected layer in the target model would be initialized. Train the target model with the training and validation sets from the target domain data. The important hyperparameters are also adjusted to fit the data distribution characteristics of the target domain. Finally, the prediction intervals for different quartiles of future load are obtained using the test set according to interval construction theory. The interval prediction results under different models and different strategies are evaluated with the above four metrics.

4. Case studies

4.1 Data Sources

This paper is validated using actual load data from New York State, USA, sourced from the NYISO. Data from the source domain and data from the target domain make up the loaded dataset, where the first has 8, 670 items and have a data scale from Jan. 1 to Dec. 31, 2019, for Western New York State's 1-hour electrical load values (MW).

The latter is the NYC 1-hour electrical load values (MW) data granularity, which includes 2, 160 entries with data scales from Jan. 1 to Mar. 31, 2019. Jan. 1 to Mar. 17 data are used to train the models. Mar. 18 to Mar. 24 data are used to adjust the parameters of the models. The effectiveness of interval forecasting is evaluated using data from Mar. 25 to Mar. 31.

4.2 Parameter setting

The algorithm parameters are configured as follows to ensure that the GASA algorithm quickly converges to the global standards of excellence: the per-generation gap G_{gap} in GA is 0.95; P_c is 0.7 for the crossover probability; P_m is 0.01 for the variation probability. The probability of selecting the exchange structure p_S in the simulated annealing algorithm is 0.2; p_R is 0.5; p_I is 0.3; 5 cycles for the inner layer and 10 cycles for the outer layer; the initial temperature T is 0.025; and the cooling factor is 0.98.

4.3 Metrics

Two measures, FICP and FINAW, are chosen for evaluation to confirm the prediction performance of the abovementioned models. They are defined as follows:

$$FICP = \frac{1}{k}\sum_{t=1}^{k} \varepsilon_t \times 100\% \tag{8}$$

$$FINAW = \frac{1}{kR}\sum_{t=1}^{k}(U_t - L_t) \times 100\% \tag{9}$$

where ε_t denotes the coverage state of the prediction interval, if the true value y_t is covered by the higher limit U_t and the bottom level L_t of the prediction interval, then $\varepsilon_t = 1$; otherwise $\varepsilon_t = 0$. R is the variation in the actual value y_t between its highest and lowest values.

However, both of these metrics evaluate only one aspect of the prediction interval individually. The CWC, which is frequently used in practical issues to assess overall quality, is expressed as follows:

$$CWC = FINAW\left(1 + \xi \mathrm{e}^{-\eta(FICP-\tau)}\right) \tag{10}$$

$$\xi = \begin{cases} 0, FICP \geq \tau \\ 1, FICP < \tau \end{cases} \tag{11}$$

where ξ and η are the determining parameters, $\eta = 10$, ξ depends on whether the FICP reaches a given confidence level τ. In addition, quantile scoring (QS) was used to evaluate the sharpness and reliability of the quantile regression, with the following expression:

$$L_i(\mu, y_i, \hat{y}_{i,\mu}) = max\left(\left(\frac{\mu}{10} - 1\right)(y_i - \hat{y}_{i,\mu}), \frac{\mu}{10}(y_i - \hat{y}_{i,\mu})\right) \tag{12}$$

$$QS = \frac{1}{M_\mu M}\sum_{\mu=1}^{M_\mu}\sum_{i=1}^{M} L_i(\mu, y_i, \hat{y}_{i,\mu}) \tag{13}$$

where L_i is the quantile score of the i-th load point, y_i denotes the i-th actual load, the quantile score of the i-th load point at quantile μ is represented by $\hat{y}_{i,\mu}$, the numbers of quantile sites and loads are represented by M_μ and M, respectively.

4.4 Experimental Results

4.4.1 Performance comparison of pre-trained models. The results of the pre-trained models' interval prediction are displayed in Table 1. Compared with ENNQR, GRUQR, and LSTMQR improve QS by 61.15% on average and PICP by 61.22% on average. This is attributed to their complex internal gate structure that can consistently mine the internal relationships of time series at different quantile points. LSTMQR, on the other hand, achieves better interval prediction results with its excellent ability to handle long-term dependent information. However, under the non-migration strategy, none of their PICPs reached the given confidence level.

Table 1. Interval Prediction Results of The Pre-trained Models.

Model	PICP (%)	PINAW (%)	QS
ENNQR	32.73	5.20	85.03
GRUQR	86.06	9.43	33.05
LSTMQR	87.88	8.60	33.01

4.4.2 Analysis of optimization effects. Figure 1 displays the outcomes of the search for optimization. GASA optimizes the weights in 30 generations with the lowest CWC as the goal. It can be seen that the CWC is 15% in the 1st generation and the global optimal solution is obtained after 5 evolutions. The CWC at the 16th generation is 13.4%. This shows that the GASA can efficiently improve the convergence speed.

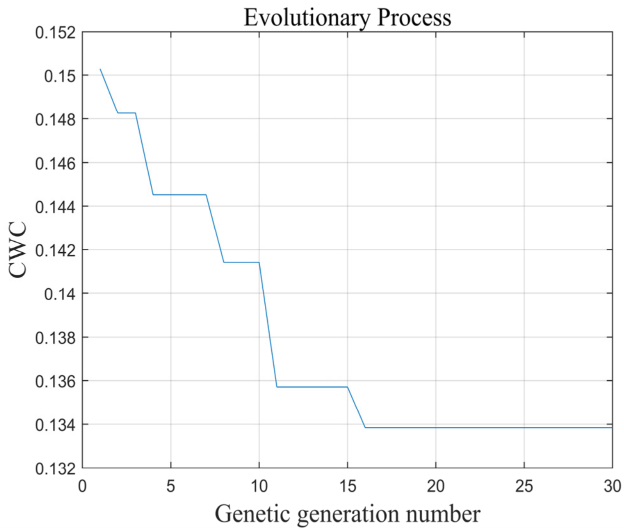

Figure1. GASA optimization process.

4.4.3 Ablation experiment. Figure 2 displays the interval prediction outcomes of the model used in this study. The findings show that OTL-LSTMQR covers the majority of the real values within a narrow FI width. Compared with the other strategies in Table 2, the performance of interval prediction can be enhanced by the suggested policy. It can be seen that the addition of the transfer strategy brings the PICP closer to the confidence level. However, this improvement is at the expense of interval width. The addition of the optimized transfer strategy effectively solves this problem. The model can reach the confidence level while the PINAW becomes 68.25% narrower than the previous strategy.

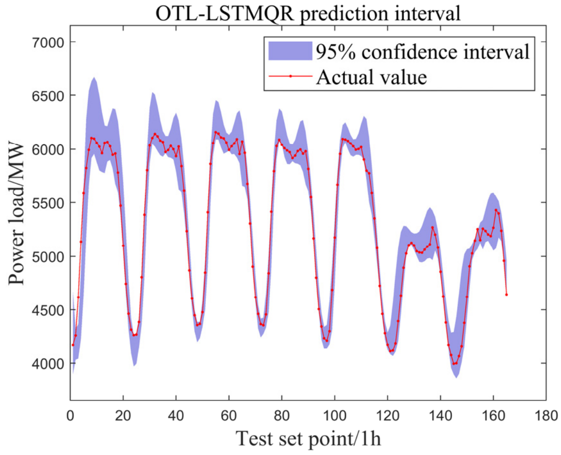

Figure 2. Prediction interval at 95% confidence level.

Table 2. Interval Prediction Results for Different Strategies.

Model	LSTMQR	TL-LSTMQR	OTL-LSTMQR
PICP (%)	87.88	94.55	95.15
PINAW (%)	8.60	57.00	18.10
CWC (%)	3527.05	182.52	18.10
QS	33.05	241.88	71.98

5. Conclusion

Decision-making about the grid depends on the accurate measurement of uncertainty information in electric demand. In this paper, an optimized transfer learning-based method for short-term load-interval prediction is proposed for the problem of poor forecasting performance due to insufficient historical load samples. After experimental analysis, the following conclusions can be obtained:

(1) The fitness evaluation function's design and the SA embedding effectively avoid the possible deception problem of GA. It enables the combinatorial optimization algorithm to find the global optimal solution in a short number of iterations.

(2) Comparing with ENNQR, GRUQR, LSTMQR, and TL-LSTMQR, the proposed strategy in this paper achieves the narrowest FINAW under the condition of satisfying the pre-defined interval coverage. this proves its superior interval prediction performance. The superior interval prediction performance is demonstrated.

Acknowledgement

Ling Xiao is the assigned author (xiaoling5231@163.com). The Science and Technology Innovation Project of The Chengdu-Chongqing Twin Cities Economic Zone (NO. KJCX2020027), the Chongqing Municipal Education Committee (NO. KJQN202100604), and the National Natural Science Foundation of China (NO. 71901045) all provided funding for this work.

References

[1] Boroojeni, K.G., Amini, M.H., Bahrami, S., Iyengar, S.S., Sarwat, A.I., Karabasoglu, O. (2017) A novel multi-time-scale modeling for electric power demand forecasting: from short-term to medium-term horizon. Electric Power Systems Research, 142: 58-73.

[2] Jorgensen, K.L., Shaker, H.R. (2020) Wind Power Forecasting Using Machine Learning: State of the Art, Trends and Challenges. In: 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE). Oshawa. pp. 44-50.

[3] Radziukynas, V., Klementavicius, A. (2014) Short-term wind speed forecasting with ARIMA model. In: 2014 55th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON). Riga. pp. 145-149.
[4] Wang, Y., Zhang, N., Chen, X. (2021) A short-term residential load forecasting model based on LSTM recurrent neural network considering weather features. Energies, 14: 2737.
[5] Wu, J., Wang, Y. G., Tian, Y. C., Burrage, K., Cao, T. (2021) Support vector regression with asymmetric loss for optimal electric load forecasting. Energy, 223: 119969.
[6] Salkuti, Reddy, S. (2018) Short-term electrical load forecasting using radial basis function neural networks considering weather factors. Electrical Engineering, 100: 1985-1995.
[7] Li, Y., Jia, Y.J., Li, L., Hao, J.S., Zhang, X.Y. (2020) Short-term power load forecasting based on a stochastic forest algorithm. Power System Protection and Control, 48: 8.
[8] Kim, J., Moon, J., Hwang, E., Kang, P. (2019) Recurrent inception convolution neural network for multi-short-term load forecasting. Energy and Buildings, 194: 328-341.
[9] Wen, L.L., Zhou, K., Yang, S., Lu, X. (2019) Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting. Energy, 171: 1053-1065.
[10] Hu, Z.Y., Bao, Y.K., Chiong, R., Xiong, T. (2015) Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection. Energy, 84: 419-431.
[11] He, Y.Y., Zhang, W.Y. (2020) Probability density forecasting of wind power based on multi-core parallel quantile regression neural network. Knowledge-Based Systems, 209: 106431-106447.
[12] Ren, C., Xu, Y., Dai, B.J., Zhang, R. (2021) An integrated transfer learning method for power system dynamic security assessment of unlearned faults with missing data. IEEE Transactions on Power Systems, 36: 4856-4859.
[13] Ren, C., Xu, Y. (2019) Transfer Learning-based Power System Online Dynamic Security Assessment: Using One Model to Assess Many Unlearned Faults. IEEE Transactions on Power Systems, 35: 821-824.
[14] Bashath, S., Perera, N., Tripathi, S., Manjang, K., Dehmer, M., Streib, F.E. (2022) A data-centric review of deep transfer learning with applications to text data. Information Sciences, 585: 498-528.
[15] Xu, X., Meng, Z. (2020) A hybrid transfer learning model for short-term electric load forecasting. Electrical Engineering, 102: 1371-1381.
[16] Gatys, L.A., Ecker, A.S., Bethge, M. (2016) Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414-2423.
[17] Howard, J., Ruder, S. (2018) Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.
[18] Zhang, Y., Ying, S., Wen, Z. (2022) Multitask transfer learning with kernel representation. Neural Computing and Applications, 34: 12709-12721.
[19] He, Y.Y., Liu, R., Han, A.Y. (2017) Short-term power load probability density forecasting method based on real-time price and support vector quantile regression. Proceedings of the CSEE, 37:768-775.
[20] Gao, C., Sun, Y.Q., Zhao, H.F., Cao, P.F. (2022) Research on short-term load forecasting based on ICOA-LSTM. Electronic Measurement Technology, 45: 88-95.
[21] Jiang, Q., Huang, R., Huang, Y., Chen, S., He, Y., Lan, L., Liu, C. (2019) Application of BP neural network based on genetic algorithm optimization in the evaluation of power grid investment risk. IEEE Access, 7: 154827-154835.
[22] Chen, F., Wu, X.Y. (2009) Directed mutation operator in the research of solving GA deceptive problem. Systems Engineering and Electronics, 31:204-207.
[23] Oudani, M. (2021) A simulated annealing algorithm for intermodal transportation on incomplete networks. Applied Sciences, 11: 4467-4481.