Micro-Genetic Algorithm Embedded Multi-Population Differential Evolution based Neural Network for Short-Term Load Forecasting

Colin Paul Joy
School of computing, Engineering and Digital technologies
Teesside University
Middlesbrough, UK
c.joy@tees.ac.uk

Dr Gobind Pillai
School of computing, Engineering and Digital technologies
Teesside University
Middlesbrough, UK
g.g.pillai@tees.ac.uk

Dr Yingke chen
School of computing, Engineering and Digital technologies
Teesside University
Middlesbrough, UK
y.chen@tees.ac.uk

Dr Kamlesh Mistry
Dept. of computer and information sciences
Northumbria University
Newcastle Upon Tyne, UK
k.mistry@northumbria.ac.uk

***Abstract*—The load of a power system usually presents a certain range of nonlinear fluctuation with time. Even then, the load characteristics still follow certain rules which can be exploited to optimise and improve the accuracy of computer-based Short-Term Load Forecasting (STLF) models. Therefore, this paper presents a mGA (micro–Genetic Algorithm) embedded multi-population DE (Differential Evolution) to optimise an Artificial Neural Network (ANN) STLF model. Firstly, the mGA embedded multi-population DE is proposed, to improve and balance the global and local search. Then the proposed DE is applied to optimise the weights during the training of the ANN. The overall model's performance is evaluated using publicly available Panama electricity load dataset against four state-of-the-art machine learning algorithms. The evaluation results show that the proposed DE based NN STLF model has higher prediction accuracy compared to the other selected machine learning algorithms.**

Keywords—Machine Learning, Differential Evolution, Neural Networks, Electricity Load Forecasting

I. Introduction

The process of predicting the demand for electricity consumption during a specific time frame in a specific area is referred to as Load forecasting. Load forecasting is an effective and important technique that assists in the management and operation of power systems which can contribute to significant reductions in costs when conducted precisely [1]. Currently, electric storage such as batteries and pumped hydro storages have limited capacities. Because of the lack of mechanisms to store electricity in large quantities, at the national scale, the amount of electricity that is produced at a given time has to cover the demand and compensate for the losses. Hence, there is always a requirement to track the load by system generation for effective operation of a power system.

The forecasting carried out for a single day to several weeks ahead is usually referred to as short-term load forecasting (STLF). This process supervises the production of electricity in power stations [2] by tackling electricity wastage, manpower wastage and assist in mapping power infrastructure. Faultless and precise load prediction has forever been a continuous challenge in research since if the predicted load is lesser than required load there will arise a shortage of electricity. Meanwhile, if the predicted load is surplus to the required load this will lead to wastage of energy, machine resources and drive up the costs. Load forecasting helps to establish balance in regional power supply and demand while simultaneously making sure the whole system is run safely.

In order to build the load prediction models in this paper, the city of Panama is chosen as it shares its electricity load data to the public providing us with much needed research data considering most collective and multi-site electricity demand statistics are difficult to obtain. Based on methodologies proposed by the National Dispatch Centre (CND) [3], the organization that is in charge of power system planning and operations in Panama, the final aim of forecasting with minimum deviation is to predict and supply the demand while incurring minimum losses. The CND performs the forecast planning weekly and ensures short term forecasting for the following week is efficient enough to tackle security issues in the electrical system. To conduct short term scheduling, the CND utilizes an optimization software that handles input hourly [5] by scrutinizing data related to load forecast, power plants and the power grid. The Nostradamus Artificial Neural Network (ANN) by Hitachi ABB [6] is employed by the National Dispatch Centre for optimization helping in hourly and weekly predictions [7].

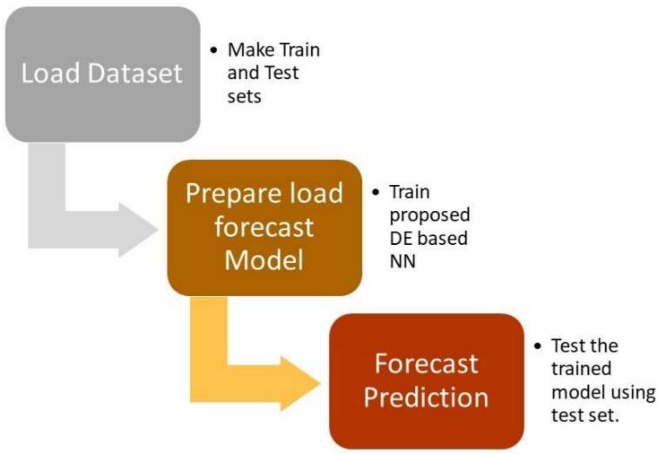

Fig. 1. System Architecture diagram

978-1-6654-4389-0/21/$31.00 ©2021 IEEE

This paper proposes a novel variant of DE to improve ANN performance. On this basis, we have established an ANN STLF model based on the proposed DE to optimise the neural network, and the model is evaluated by comparing with other state-of-the-art machine learning algorithms. The overall system architecture is illustrated in Fig 1.

II. Literature Review

Methodological advances and the rapid development of the internet has seen an improvement in accuracy of STLF. Of all the approaches to enhance the accuracy, it is predominantly divided into two categories: classical statistics methods and artificial intelligence techniques.

The statistics approaches include linear regression, Kalman filtering method, and state space model etc., which excel in load prediction but falls short in enabling the system with insight into the reason for more structural changes. The parameter selection of Support Vector Machines (SVM) is vital and directly influences the efficacy of forecasting. Machine based schemes of forecasting load encompass expert systems, fuzzy logic, ANN etc. which attain competitive advantages for nonlinear load mapping and generalisation, although offering hinderances to criteria making and parameter setting. With the advancements in technology and computer aided predictions, the process of power load forecasting has seen improvements via ANN, wavelet transforms, fuzzy algorithms, SVM etc [8-11].

Owing to improved capabilities of nonlinear relationship modelling between the inputs and outputs, ANNs are quite often employed to forecast load [8-11]. The ANN method for short term load forecasting, operates by making computers mimic the mathematical model of a human brain [4]. ANN networks account for significant part of forecasting models where a plethora of research is dedicated to its optimisation [5-6]. By using training data to learn, these networks are effective in extracting the innate connections between input variables [12]. Load forecasting with the help of neural networks was first researched upon in publications made in the late 1980's, the number of subsequent research increasing substantially from then on [13].

Recently, many intelligent systems like Particle Swarm Optimisation algorithms have been introduced to handle the training of the ANN networks in STLF [14,15]. An STLF forecasting system founded on Adaptive Cauchy variation Particle Swarm Optimisation (ACMPSO) and Long Short-Term Memory (LSTM) neural network was proposed by Wei et al in [7]. The ACMPSO was effective in using the nonlinear inertia weight to enhance the global optimization capability and improve the convergence speed of the PSO. In order to minimize the risk of local stagnation or convergence into local optima, a mutation operator is further embedded into this algorithm. Zeng [8] presented a PSO-BP STLF model that consists of data mining techniques along with an ant colony algorithm to optimise the PSO- Back propagation system.

The above literature shows that ANN is very popular for predicting STELF. Moreover, involvement of evolutionary algorithms is also increasing in this field.

III. Methodology

In this section, we introduce the proposed model for STLF. Section A gives information about conventional DE and section B talks about the proposed DE and its application.

A. Conventional DE

Storn and Price [16] proposed DE to deal with global optimisation in continuous spaces. DE algorithm employs the scaling factor between two individuals, which is also called as mutation factor. DE algorithm starts its search with the random initialisation of vectors and tries to improve them to further obtain optimal solution. In DE, the population with *Np* vectors in g generations is denoted as *P={X1, X2,…,XNp}* where, *Xi=(xi,a,…,xi,D)*. Conventional DE consists of three important steps:

1) Mutation

In this step, three vectors from the population are selected randomly and the following mutation equation is applied:

$$V_i = X_{i_1} + F(X_{i_2} - X_{i_3}) \quad (1)$$

Where, *F∈[0,2]* it further controls the augmentation of the differential vector of *(Xi2−Xi3)*. In DE, the *F* value plays very important role in controlling the exploration ability. Higher the *F* value higher the exploration ability and vise-versa.

2) Crossover

This step is applied to improve the diversity of the population by crossing the mutant and parent vector as follows:

$$U_{i,d} = \begin{cases} V_{i,d}, & if\ rand_d(0,1) \leq C_r or\ d_{rand} = d \\ x_{i,d}, & otherwise \end{cases} \quad (2)$$

Where d is the dimension, C_r is the cross rate parameter, and the trial vector U_i can be generated as given below:

$$U_i = (U_{i,1}, \dots, U_{i,D}) \quad (3)$$

3) Selection

This is the final step in each iteration, where U_i or X_i vector is selected based on their fitness value. The best fitness value vector is selected and sent for next iteration until the end condition is met.

B. Proposed DE

The conventional Differential Evolution only employs one set of population to explore the search space [16]. The single set of population restricts the search in only one direction or one set of features. The application of optimising the weights for ANN can lead to multiple clusters of features, which makes conventional DE less useful in this scenario.

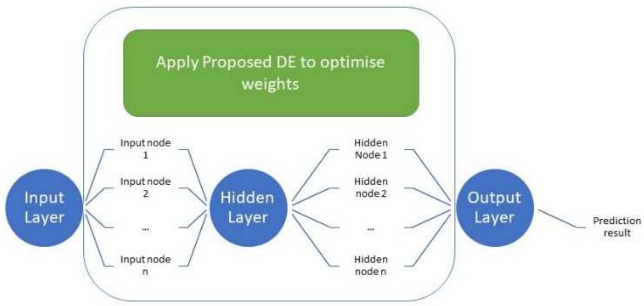

Fig. 2. Flow diagram of proposed DE based NN model

In this paper, we employ a multi-population DE to optimise the weights for ANN classifier. Firstly, initialising number of populations with the same size of 30 members. Each population has its own fitness criteria i.e. the populations will look for different output criteria. However, it simply means that multiple DE's will run in parallel to generate multiple sets of features.

In order to further diversify the populations and reduce the risk of pre-mature convergence, we employ micro–Genetic Algorithm's (mGAs) diversity maintenance strategy of non-replaceable memory [17]. During the first iteration, we store all members of each population in one set of non-replaceable memory. These stored members are further used to aid the micro-population generation. After the first iteration, we conduct the DE on all populations and when the system reaches 30 iterations (selected based on experiments) a set of combined micro-population will be created. This micro-population will consist of 5 members including the global best member from each population and three most correlating members from the non-replaceable memory. A GA with three mutation strategies will be applied to the micro-population. If the newly generated members are better than the previous global bests then they will be replaced. In this version of GA we have employed gaussian mutation [17], Cauchy mutation [17] and Levys flight mutations [17]. This will improve the diversity and balance the global and local exploration.

The overall structure of the proposed DE is illustrated in Algorithm 1. The flow diagram of the proposed model is given in Fig 2. In this paper, the STLF ANN is set with 9 input layers, 11 hidden layers, and 1 output layer. The target error was set as 0.0001.

Algorithm 1: The Pseudo Code of the Proposed DE Variant

```
Step-1 Initialise multiple sets of populations.
Step-2 Evaluate the fitness value for each population using separate
criteria.
Step-3 Store the current version of population in Non-replaceable
memory.
Step-4 While (satisfying termination criteria)
        Perform standard DE steps as follows:
        Mutation using Equation 1.
        Crossover using Equation 2 and 3.
        Perform selection.
     Re-evaluate the fitness value for everyone in the   current
     population.
      If (iterations >=30)
         Select best individual from each population and compare it with
     members from Non-replaceable memory.
         Select the three most co-relating members and add them to
     Micro population along with the multiple global bests.
          Run GA for 100 iterations using Micro-population
                 Evaluate fitness for everyone in micro population using
     fitness criteria.
           Swap global bests if the newly generated members have higher
     fitness value.
       End If
      Re-evaluate the fitness value for everyone in all populations.
End While

Step-4 End
```

IV. Evaluations

This research work was conducted using a standard hardware computer with Intel i7-10700 processor with 16 Gigabyte RAM. All work is implemented using Python as the base language. The publicly available Panama electricity load dataset was used [3]. This dataset provides hourly records from January 2015 until June 2020. This dataset is further transformed into train and test sets for the ANN model while maintaining the chronological order.

Literature shows that decision-makers find it hard to trust machine generated predictions and demand to see interpretability in machine learning (ML) algorithms [18]. To meet these demands, a set of ML algorithms with proven track record of forecast prediction are selected. The proposed model is evaluated against the selected set of ML algorithms, which includes K-Nearest Neighbour regressor (KNNR), Support vector regression (SVR), Random forest regressor (RFR), and extreme gradient boosting regressor (XGBR). All the models including the proposed model are trained and tested using the same dataset. To evaluate the performance, the results are converted into 5 different metrices: Mean Absolute Percent error (MAPE), Square Root of Mean Square Error (RMSE), Peak load absolute precent error (Peak), Valley load absolute precent error (Valley) and Energy absolute precent error (Energy).

TABLE I. Error distribution by model

Model	Metric	Mean	SD
KNN	MAPE	4.56	3.41
	RMSE	54.63	41.68
	Peak	5.19	2.48
	Valley	2.91	2.37
	Energy	2.66	1.79
SVR	MAPE	4.08	3.41
	RMSE	49.82	42.25
	Peak	3.43	1.92
	Valley	4.38	2.73
	Energy	2.18	1.78
RF	MAPE	4.11	3.17
	RMSE	49.97	39.21
	Peak	3.94	2.35
	Valley	3.68	2.82
	Energy	1.71	1.54
XGBR	MAPE	3.66	2.95
	RMSE	44.52	36.09
	Peak	2.93	1.99
	Valley	3.04	3.13
	Energy	1.75	1.30
Proposed model	MAPE	3.23	2.50
	RMSE	44.01	33.10
	Peak	2.75	1.90
	Valley	2.91	3.11
	Energy	1.66	1.22

TABLE II. Evaluation metrices by model forf avergage of 14 testing weeks.

	MAPE	RMSE	Peak	Valley	Energy
KNNR	4.43	66.7	5.10	2.88	2.61
SVR	4.22	62.3	3.42	4.36	2.19
RFR	4.16	61.6	3.92	3.62	1.69
XGBR	3.77	55.1	2.90	3.10	1.77
Proposed model	3.33	54.25	2.77	2.97	1.65

The results illustrated in Table 1, shows that KNNR has a larger standard deviation, followed by SVR, RFR and XGBR. Moreover, the proposed model shows the best performance when compared to the selected ML models. The proposed model requires higher computation time due to the involvement of multi-population DE. While KNNR requires significantly lower computation time but shows the worst performance.

To further evaluate the models' performance, all models are evaluated in terms of testing week, where each week has a different load profile. Table 2 shows the average performance for each ML model. The weekly evaluation demonstrates that proposed model improved on all 5 metrics, while XGBR shows very close performance.

V. CONCLUSION

This paper proposes a mGA embedded multi-population DE to optimise STLF Neural Network Models. In the aspect of combining the optimisation of the Intelligent Algorithm with the fitting of the ANN, the superior optimisation ability of the proposed DE is used to guide the updating of weights. The model was developed and benchmarked with publicly available data and previous forecasts for Panama's power system. Results along 14 testing weeks confirmed the suitability of the proposed model.

REFERENCES

[1] J. Wang, S. Zhu, W. Zhang, and H. Lu, "Combined modeling for electric load forecasting with adaptive particle swarm optimization," Energy, vol. 35, no. 4, pp. 1671–1678, 2010.

[2] J. Runhai, Z. Tianming, J. Yizhi and H. Hui, "Short-Term Non-Residential Load Forecasting Based on Multiple Sequences LSTM Recurrent Neural Network", *IEEE Access*, vol. 1, no. 6, pp. 59438-59448, 2018.

[3] CND—ETESA, Metodologías de Detalle. Available online: https://www.cnd.com.pa/images/doc/norm_metodologiasdetalle_nov2020.pdf.

[4] M. Cao, W. Li and X. Qian, "Power load forecasting algorithm based on classification recognition deep confidence network", Power demand side management, no. 2, pp. 44-50, 2020.

[5] J. Ping and N. Ying, "A Hybrid Double Forecasting System of Short Term Power Load Based on Swarm Intelligence and Nonlinear Integration Mechanism", APPLIED SCIENCES-BASEL, vol. 10, no. 4, 2020.

[6] J. T. Lalis, B. D. Gerardo and Y. Byun, "A Mechanism for Online and Dynamic Forecasting of Monthly Electric Load Consumption Using Parallel Adaptive Multilayer Perceptron (PAMLP)", 17th IEEE/ACIS International Conference on Computer and Information Science (ICIS), vol. 791, pp. 177-188, June 2018.

[7] T. Wei and T. Pan, "Short-term Power Load Forecasting Based on LSTM Neural Network Optimized by Improved PSO", Journal of System Simulation.

[8] Q. Zhang, Y. Ma, G. Li, J. Ma, and J. Ding, "Short-term load forecasting based on frequency domain decomposition and deep learning," Mathematical Problems in Engineering, vol. 2020, Article ID 7240320, 9 pages, 2020

[9] J. Huang, Y. Tang, and S. Chen, "Energy demand forecasting: combining cointegration analysis and artificial intelligence algorithm," Mathematical Problems in Engineering, vol. 2018, Article ID 5194810, 13 pages, 2018.

[10] S. Zheng, Q. Zhong, L. Peng, and X. Chai, "A simple method of residential electricity load forecasting by improved Bayesian neural networks," Mathematical Problems in Engineering, vol. 2018, Article ID 4276176, 16 pages, 2018.

[11] L. Wu, C. Kong, X. Hao, and W. Chen, "A short-term load forecasting method based on GRU-CNN hybrid neural network model," Mathematical Problems in Engineering, vol. 2020, Article ID 1428104, 9 pages, 2020.

[12] H. Shayeghi, H. Shayanfar, and G. Azimi, "A hybrid particle swarm optimization back propagation algorithm for short term load forecasting," International Journal on, Technical and Physical Problems (IJTPE), vol. 4, no. 2, pp. 12–22, 2010.

[13] 23.H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," IEEE Transactions on Power Systems, vol. 16, no. 1, pp. 44–55, 2001.

[14] 24. X. Zheng, X. Ran, and M. Cai, "Short-term load forecasting of power system based on neural network intelligent algorithm," IEEE Access, , 1 page, 2020.

[15] 25.Y. K. Semero, J. Zhang, and D. Zheng, "EMD–PSO–ANFIS-based hybrid approach for short-term load forecasting in microgrids," IET Generation, Transmission and Distribution, vol. 14, no. 3, pp. 470–475, 2019.

[16] R. Storn and K. Price, "Journal of Global Optimization" 11: 341, 1997. [29] N. Strisciuglio, G. Azzopardi, M. Vento and N. Petkov, "Multiscale blood vessel delineation using B-cosfire filters," In: International Con- ference on Computer Analysis of Images and Patterns. Springer, pp.300–312, 2015.

[17] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim, and B. Fielding, "A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition," IEEE Transactions on Cybernetics, 47(6), pp.1496 - 1509, 2016.

[18] Bertsimas, D.; Delarue, A.; Jaillet, P.; Martin, S. The Price of Interpretability. 2019.