Optimizing Bidirectional LSTM for Energy Consumption Prediction Using Chaotic Particle Swarm Optimization and Hyperparameter Tuning

Candra Juni Cahyo Kusuma[1*], Khairunnisa[2]
[1]Universitas PGRI Yogyakarta, Indonesia
[2]National Taiwan University of Science and Technology, Taiwan
Email: candrajunick02@gmail.com

Article Info

Article history:

Received: January 20, 2024
Revised: March 12, 2024
Accepted: June 30, 2024
Available Online: July 30, 2024

Keywords:

Bi-LSTM
CPSO
Hyperparameter Tuning
Energy Consumption Prediction

ABSTRACT

This study aims to improve the accuracy of energy consumption prediction using the Bidirectional Long Short-Term Memory (BLSTM) model which is known to be able to handle temporal dependencies in time series data. However, the performance of BLSTM is greatly affected by the hyperparameter configuration, which often requires manual tuning which is inefficient. To address this, this study proposes an optimization framework that combines BLSTM with Chaotic Particle Swarm Optimization (CPSO) to automatically adjust hyperparameters such as the number of hidden units and learning rate. Experiments show that BLSTM optimized with CPSO produces higher prediction accuracy compared to traditional methods such as grid search and random search. By utilizing the chaos map, CPSO improves exploration and exploitation capabilities, accelerates convergence, and finds more optimal solutions. The integration of CPSO and BLSTM shows promising results for improving the performance of time series prediction models, especially in energy consumption forecasting.

1. INTRODUCTION

As global energy demand continues to rise, it is imperative to optimize energy distribution, reduce operational costs, and minimize energy consumption [1]. This is because numerous countries, including industrialized nations, are unable to accurately identify and forecast energy consumption patterns due to various factors, including the complex dynamics of the global energy market, fluctuations in demand, and the impact of climate change[2].

In recent years, deep learning-based methods such as Long Short-Term Memory (LSTM) have demonstrated superior performance in predicting time series data, including energy consumption[3]–[5]. BiDirectional LSTM (BiLSTM) enhances the efficacy of LSTM by leveraging information from both temporal directions, thereby improving the model's capacity to comprehend intricate data patterns [6]–[10]. However, the hyperparameter configurations in BiLSTM influence the model's performance, necessitating an effective optimization strategy [11]–[16].

Particle Swarm Optimization (PSO) represents a potential avenue for system hyperparameter optimization[17], [18]. It has been demonstrated that the conventional PSO method encounters local static obstacles when confronted with intricate or complex optimization problems[19], [20]. To address this limitation, the Chaotic Particle Swarm Optimization (CPSO) algorithm was developed. CPSO incorporates chaos dynamics to facilitate the expansion of the search space and prevent premature convergence[21]–[23].

This research entails the creation of a BiLSTM power consumption prediction model that has been optimized using CPSO to enhance prediction accuracy. Additionally, the hyperparameters of the BiLSTM model are tuned to identify the optimal parameter set for the characteristics of energy consumption data. The objective is to develop a more efficient model for various consumption scenarios.

2. METHOD

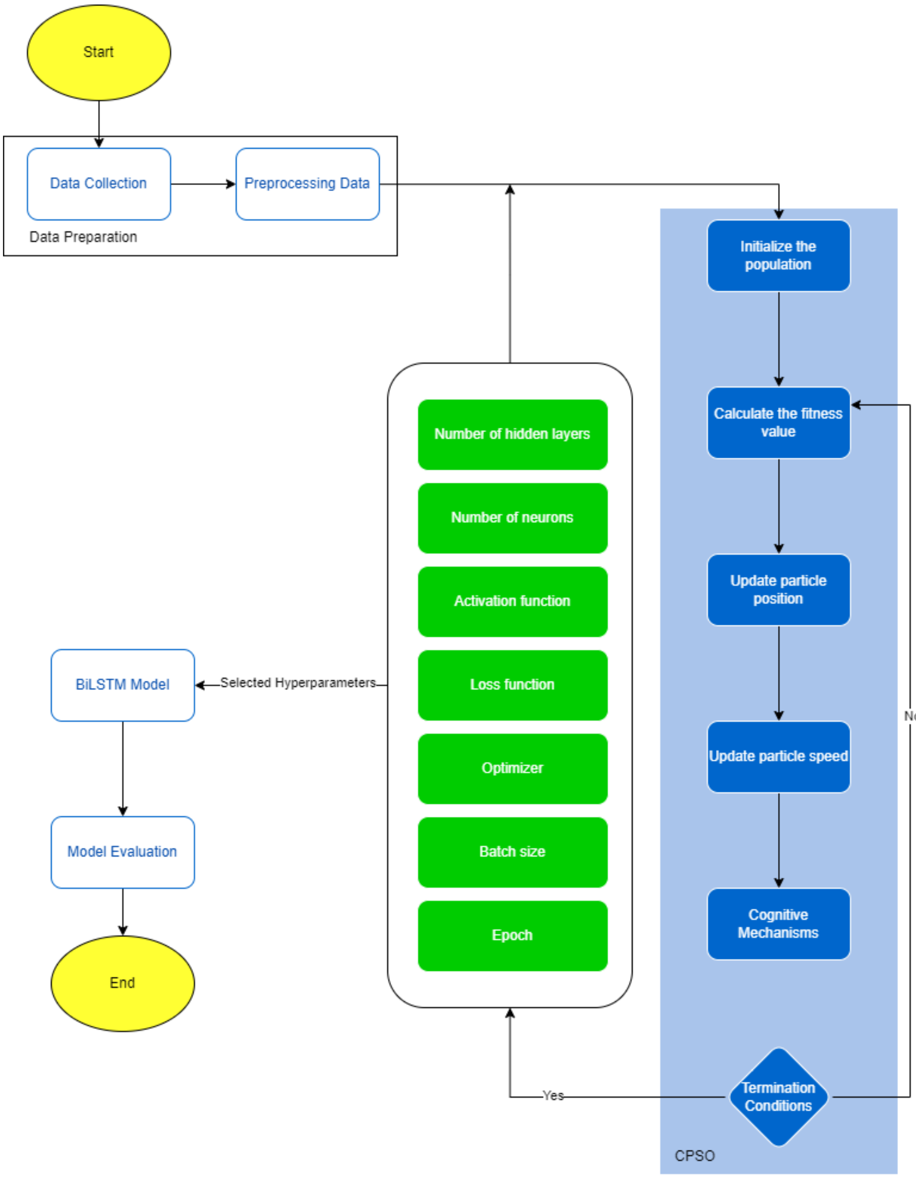

Figure 1. Research methodology

The present study adopts the approach of Bidirectional Long Short Term Memory (BiLSTM) deep learning technique to predict the energy consumption. The phases of this research work consist of data collection, building the BiLSTM model, and tuning the hyperparameters by applying Chaotic Particle Swarm Optimizer (CPSO). In detail, the research methodology is outlined as follows.

2.1. Data Preparation

The energy consumption data used in this study was obtained from the UC Machine Learning data, which provides time series information for energy consumption appliances[24]. The data collection spans roughly 4.5 months and has a length of 10 minutes. A ZigBee WSN is used to gather the variables related to interior temperature and humidity. For around 3.3 minutes, each wireless node sends out temperature and humidity data. Next, during the course of ten minutes, the data that was wirelessly retrieved within the house was averaged. Every ten minutes, an m-bus energy meter records energy data. The experimental dataset and the public dataset at rp5.ru, which includes records from Chievres Airport in Belgium, were combined using the date and time fields. To test the regression and exclude characteristics (parameters) that were not predictively relevant, eight two random variables were inserted into the dataset.

2.2. BiLSTM Model Development

The BiLSTM model is constructed in a manner that allows for the examination of both past and future energy consumption characteristics. The input to the model is a collection of energy consumption characteristics that incorporate bi-directional LSTM optimized for the appropriate number of neurons. The model also includes a dropout layer to decrease the risk of overfitting and a dense layer as the output that performs the task of energy consumption prediction.

The model architecture employed in this research paper is distinct from that utilized in other documents, namely the Bidirectional Long Short-Term Memory (Bi-LSTM) approach. In other words, BiLSTM represents an advancement of LSTM, whereby the model is capable of processing time series data in two directions: forward and backward. Therefore, this model is also capable of utilising contextual information from both the initial and final points of the sequence, which is of particular importance when working with sequential data, such as in the prediction of stock prices. In this architectural configuration, the number of hidden units utilized is 121, which was determined through the process of hyperparameter optimization. The final layer of the model comprises a single neuron, whose function is to output a continuous variable that is well suited to the regression task at hand. The model was constructed using the optimizer Adam with the optimal learning rate of 0.0061 and the mean squared error (MSE) loss function, a standard regression loss function. The MSE function measures the discrepancy between the predicted and actual values. The model was trained over 20 epochs with a batch size of 32, and the validation data were employed to monitor the model's performance in real time during training.

2.3. Hyperparameter Optimization with CPSO

Hyperparameters are optimized for BiLSTM networks through hyperparameter regularization via chaotic particle swarm optimization (CPSO). The CPSO method was selected due to its capacity to alter the focus of the search, thus circumventing local solution traps through the introduction of chaos.

The specified hyperparameters exert an influence on the architecture of genes, including factors such as the learning rate, the number of neurons in each layer, the batch size, and others. The approach employs CPSO, which maps the solution with changing particle velocity and position according to a fitness function, namely RMSE on validation data.

The CPSO algorithm commences with a number of particles representing the hyperparameters to be combined and progresses iteratively until it reaches a state of convergence or meets the pre-defined stopping criteria.

2.4. Model Evaluation

Once the training and optimization processes have been completed, the model is evaluated on the test dataset using root mean square error (RMSE), and mean absolute error (MAE) as evaluation metrics. The effectiveness of the proposed approach is analyzed using a comparison with regular models, such as a regular BiDirectional long short-term memory (BiLSTM) model without CPSO optimization.

3. RESULTS AND DISCUSSION

The objective of this investigation is to utilize the Bidirectional LSTM (BiLSTM) model to predict the consumption of energy resources and enhance the model's performance with the incorporation of Chaotic Particle Swarm Optimization (CPSO). The predictions generated by the Duo models, namely the BiLSTM and the BiLSTM-CPSO, have been evaluated in terms of their Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2). The ensuing section presents the assessment outcomes of the two models.

Table 1. Comparison of evaluation results

Model	MAE	RMSE
BiLSTM	47.7279	89.1674
BiLSTM-CPSO	51.7223	88.8565

The results for the BiLSTM model without CPSO yielded a lower mean absolute error (MAE) value (47.7279) than those obtained with the BiLSTM model that was optimized using CPSO (51.7223). The root mean square error (RMSE) is a measure of the discrepancy between the predicted and actual values. In this case, the differences between the two models are not statistically significant, as evidenced by the slight reduction in the RMSE for the BiLSTM-CPSO model. The RMSE for the BiLSTM model is 89.1674, while that for the BiLSTM-CPSO model is 88.8565. This discrepancy demonstrates that although the hyperparameter is transformed through the use of CPSO, it does not result in a reduction in MSE. However, in certain instances, it does facilitate an intuitive decline in MSE.

The overall coefficient of determination, indicated by the R^2 value of 0.2110, demonstrates that the BiLSTM-CPSO model accounts for approximately 21% of the total variation in the energy consumption data. Although this R^2 value is not particularly high, it is sufficient to demonstrate an enhancement in the model's generalization ability relative to the BiLSTM model, which was not optimized and did not result in more impressive coefficients of determination.

Chaotic Particle Swarm Optimization (CPSO) was originally designed to circumvent the common local optima encountered in traditional PSO algorithms. Despite a reduction in RMSE and other metrics, the expansion of CPSO does not necessarily guarantee improvement in all evaluation metrics. This may be attributed to factors such as the complexity of the model and the nature of the dataset, which may not necessitate further optimization using CPSO.

Overall, the energy consumption prediction accuracy of the BiLSTM model utilized in this study was satisfactory, as evidenced by the reasonable MAE and RMSE values. The incorporation of a bi-directional architecture within an LSTM architecture enables the model to gain temporal features more effectively than an LSTM architecture that is uni-directional and only processes data in one orientation.

It is important to note that the BiLSTM-CPSO model exhibits a higher MAE value, which indicates that while the CPSO is effective in identifying superior hyperparameters, this optimization technique does not necessarily yield optimal results for all metrics. Further investigation with diverse experimental settings is essential, and the potential benefit of integrating multiple optimization strategies to enhance prediction accuracy should also be considered.

4. CONCLUSION

The findings of the study's evaluation indicate that while the BiLSTM optimization through the use of CPSO can reduce the RMSE value, it does not result in a significant enhancement in the MAE. For an R2 value of 0.2110, the BiLSTM-CPSO model demonstrates superior generalization capabilities compared to the BiLSTM model without any optimization. This study presents avenues for future research, particularly the combination of CPSO with other optimization techniques to further enhance the model's performance. For further research, it will be continued with the addition of data, for example the variable household energy consumption data.

REFERENCES

[1] S. K. Hora, R. Poongodan, R. P. de Prado, M. Wozniak, and P. B. Divakarachari, “Long Short-Term Memory Network-Based Metaheuristic for Effective Electric Energy Consumption Prediction,” *Appl. Sci.*, vol. 11, no. 23, 2021, doi: 10.3390/APP112311263.

[2] B. J. Van Ruijven, E. De Cian, and I. Sue Wing, “Amplification of future energy demand growth due to climate change,” *Nat. Commun.*, vol. 10, no. 1, p. 2762, 2019.

[3] G. N. Kouziokas, “Long Short-Term Memory (LSTM) Deep Neural Networks in Energy Appliances Prediction,” 2019, vol. 2019, pp. 1–5. doi: 10.1109/PACET48583.2019.8956252.

[4] S. H. Lee, T. Lee, S. Kim, and S. Park, “Energy Consumption Prediction System Based on Deep Learning with Edge Computing,” 2019, pp. 473–477. doi: 10.1109/ELTECH.2019.8839589.

[5] E. Balraj, T. Pugalendhi, M. R. Sureshkumar, and K. Vijayasarathi, “Optimized LSTM Model for Electric Load Forecasting using Deep Learning with Genetic Algorithm,” 2023, pp. 301–304. doi: 10.1109/ICICCS56967.2023.10142580.

[6] A. S. Varghese, S. Sarang, V. Yadav, B. Karotra, and N. Gandhi, “Bidirectional LSTM Joint Model for Intent Classification and Named Entity Recognition in Natural Language Understanding,” *Int. J. Hybrid Intell. Syst.*, vol. 16, no. 1, pp. 13–23, 2020, doi: 10.3233/HIS-190275.

[7] A. Chao, H. Zhijie, Z. Yan, C. Dunliang, and Q. Yinchuan, “Bidirectional LSTM neural network-based statement identification method and apparatus.” 2016.

[8] E. Kiperwasser and Y. Goldberg, “Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations,” *arXiv: Computation and Language*. 2016. [Online]. Available: https://www.arxiv.org/pdf/1603.04351

[9] J. Cross and L. Huang, “Incremental Parsing with Minimal Features Using Bi-Directional LSTM,” 2016, vol. 2, pp. 32–37. doi: 10.18653/V1/P16-2006.

[10] M. M. Rahman, Y. Watanobe, and K. Nakamura, “A Bidirectional LSTM Language Model for Code Evaluation and Repair,” *Symmetry (Basel).*, vol. 13, no. 2, 2021, doi: 10.3390/SYM13020247.

[11] A. B. Gumelar, E. M. Yuniarno, D. P. Adi, A. G. Sooai, I. Sugiarto, and M. H. Purnomo, “BiLSTM-CNN Hyperparameter Optimization for Speech Emotion and Stress Recognition,” 2021. doi: 10.1109/IES53407.2021.9594024.

[12] A. Sinha, T. Khandait, and R. Mohanty, “A Gradient-based Bilevel Optimization Approach for Tuning Hyperparameters in Machine Learning.,” *arXiv: Learning*. 2020. [Online]. Available: https://arxiv.org/pdf/2007.11022.pdf

[13] M. Munem, T. M. Rubaith Bashar, M. H. Roni, M. Shahriar, T. B. Shawkat, and H. Rahaman, “Electric power load forecasting based on multivariate LSTM neural network using bayesian optimization,” *2020 IEEE Electr. Power Energy Conf. EPEC 2020*, vol. 3, 2020, doi: 10.1109/EPEC48502.2020.9320123.

[14] J. J. Cherian *et al.*, “Efficient hyperparameter optimization by way of PAC-Bayes bound minimization.,” *arXiv: Machine Learning*. 2020. [Online]. Available: https://arxiv.org/pdf/2008.06431

[15] K. C. Ukaoha and E. C. Igodan, “Architecture Optimization Model for the Deep Neural Network For Binary Classification Problems,” p. 0, 2020, doi: 10.21608/IJICIS.2020.18509.1008.

[16] Q. Lyu and J. Zhu, “Revisit Long Short-Term Memory: An Optimization Perspective.” 2015. [Online]. Available: https://ml.cs.tsinghua.edu.cn/~jun/pub/lstm-parallel.pdf

[17] F. Ye, “Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data.,” *PLoS One*, vol. 12, no. 12, 2017, doi: 10.1371/JOURNAL.PONE.0188746.

[18] H. Xing and X. Pan, “Application of improved particle swarm optimization in system identification,” 2018. doi: 10.1109/CCDC.2018.8407336.

[19] K. E. Parsopoulos, “Particle Swarm Methods,” 2018, pp. 639–685. doi: 10.1007/978-3-319-07124-4_22.

[20] M. Kaminski, “Neural Network Training Using Particle Swarm Optimization - a Case Study,” 2019, pp. 115–120. doi: 10.1109/MMAR.2019.8864679.

[21] G. Shang and Y. Jing, “Research on Chaos Particle Swarm Optimization Algorithm,” *Pattern Recognit. Artif. Intell.*, 2006.

[22] D. Tian and Z. Shi, “MPSO: Modified particle swarm optimization and its applications,” *Swarm Evol. Comput.*, vol. 41, pp. 49–68, 2018, doi: 10.1016/j.swevo.2018.01.011.

[23] H. Zhao, Z. Ru, X. Chang, and S. Li, “Reliability Analysis Using Chaotic Particle Swarm Optimization,” *Qual. Reliab. Eng. Int.*, vol. 31, no. 8, pp. 1537–1552, 2015, doi: 10.1002/QRE.1689.

[24] L. Candanedo, “Appliances Energy Prediction.” UCI Machine Learning Repository, 2017. [Online]. Available: https://doi.org/10.24432/C5VC8G