Research on the Interpretability Analysis Method of Transient Stability Assessment in Power Systems Based on Deep Learning

Qinfeng Ma[1], Qingqing Zhang[1], Mingshun Liu[1], Jie Zhang[2,*], Yihua Zhu[3], Zhuohang Liang[4], Su An[1], Qingxin Pu[1], Jiang Dai[1]

[1]Power Dispatching and Control Center of Guizhou Power Grid Co., Ltd.Guizhou, China
[2]State Key Laboratory of HVDC, Electric Power Research Institute, China Southern Power Grid Guangzhou 510663, China
[3]National Energy Power Grid Technology R&D Centre Guangzhou 510663, China
[4]Guangdong Provincial Key Laboratory of Intelligent Operation and Control for New Energy Power System, Guangzhou 510663, China

*zhangjie5@csg.cn

Abstract. In this paper, application of deep learning techniques and their interpretability analysis are explored in transient stability assessment of power systems. With the continuous expansion and increasing complexity of power system scales, traditional stability assessment methods are facing new challenges. Due to their outstanding data processing and learning capabilities, deep learning techniques are able to provide new insights for improving the accuracy and efficiency of transient stability assessment. By elaborating on the application process of deep learning models in power system stability assessment, which includes model selection, training and optimization strategies, this study demonstrates the advantages of deep learning in handling complex system data. Furthermore, this work emphasizes the importance of model interpretability, analyzes several mainstream interpretability methods, and explores their potential applications in power system stability assessment, highlighting the crucial role of enhancing model transparency in understanding prediction results, boosting decision-makers' confidence, and optimizing system design. Finally, a summary of the research findings on deep learning-based transient stability assessment methods for power systems is presented, and future research directions are outlined, indicating that integrating deep learning and interpretability analysis is able to support reasonable decision making for the safe operation of power systems.

Keywords: Power System Transient Stability, Deep Learning, Interpretability

1 INTRODUCTION

The "14th Five-Year" energy system plan aims to modernize power systems for large-scale renewable energy integration, emphasizing safe operations, digital advancement, and innovative technologies. It promotes grid reforms, enhances intelligence, and

A. Haldorai et al. (eds.), *Proceedings of the 2024 3rd International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2024)*, Atlantis Highlights in Intelligent Systems 11,
https://doi.org/10.2991/978-94-6463-490-7_44

adapts to centralized and distributed renewables. Key focuses include better power source coordination, expanded energy storage, and increased grid flexibility. As the economy grows and electricity demand rises, the integration of large-scale renewable energy and the expansion of electronic devices have significantly increased the complexity of power grids. This complexity, combined with advances in communication technology, has made power systems into highly complex systems with multifaceted information interactions. Traditional stability prediction methods, based on mechanistic models, struggle to accommodate the increasing randomness and intricacy of modern power systems, leading to a need for more effective predictive methods. These methods are crucial for assisting operators in making informed control decisions and enhancing grid stability [1].

Deep learning has revolutionized various industries by addressing complex, multidimensional data challenges, particularly in power system informatization. Traditional methods fall short in managing the data-intensive demands of today's power systems, making deep learning combined with big data a pivotal area of research for assessing transient stability. Despite its advantages, deep learning's "black box" nature poses challenges like model opacity and sample imbalance, complicating its application in power system stability.

This paper explores transient stability in power systems, starting with an introduction to key deep learning architectures—Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), and Deep Reinforcement Learning (DRL)—and their roles in stability assessment. It then discusses interpretability analysis, differentiating between inherently interpretable models and model-agnostic post-hoc methods, highlighting their applications and potential in transient stability assessments. The paper concludes by summarizing the benefits of these interpretability methods, emphasizing enhanced model transparency and improved decision-making, essential for meeting the challenges of new energy systems and complex power grids.

2 RESEARCH ON POWER SYSTEM TRANSIENT STABILITY ASSESSMENT METHODS BASED ON DEEP LEARNING

2.1 Research on Transient Stability Assessment Methods Based on Convolutional Neural Networks

The basic concept of Convolutional Neural Networks.

CNNs [2] are deep neural networks characterized by their convolutional structure, reducing memory usage. Key operations include local receptive fields, weight sharing, and pooling layers, effectively mitigating overfitting. In CNNs, convolutional and pooling layers alternate in hidden layers, facilitating feature extraction. Weight parameters are adjusted layer by layer through gradient descent, enhancing accuracy via iterative training. The final output layer employs the Softmax activation function for classification. CNNs optimize structure by exploiting features like local receptive fields and shared weights, fully utilizing data locality. Feature extraction layer parameters are learned from training data, eliminating manual extraction and establishing input-feature classification correlations.

The application of Convolutional Neural Networks in transient stability assessment

Literature [3] demonstrates the use of voltage trajectories as CNN input for assessing transient stability. Literature [4] explores feature selection for stability using trajectory clusters, Relief, and mRMR methods, establishing a dynamic CNN-based model with time windows and credibility metrics.

2.2 Research on Transient Stability Assessment Methods Based on Deep Belief Networks

The basic concept of Deep Belief Networks

Deep Belief Networks (DBN) [5], deep learning models built on probabilistic graphical models with layers of Restricted Boltzmann Machines (RBMs), are widely used in computer vision, natural language processing, and bioinformatics for their robust feature learning and generation capabilities. A Deep Belief Network typically consists of an input layer, several hidden layers, and an output layer, with each hidden layer made up of an RBM. An RBM includes a visible layer and a hidden layer, forming a two-layer neural network that models the relationship between the visible and hidden layers through a probabilistic distribution. The structure of an RBM, as shown in figure 1, includes n hidden units h and n visible units v, where both h and v are binary variables (taking values 0 or 1). There are direct weight connections between visible and hidden units, but no connections within the same layer.

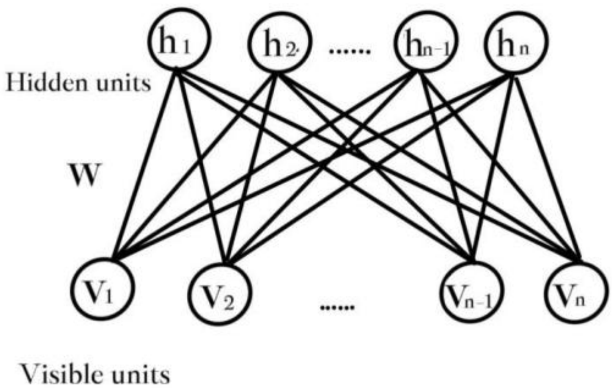

Fig. 1. Restricted Boltzmann Machine (RBM)

The application of Deep Belief Networks in transient stability assessment

Due to the powerful feature extraction capabilities and generalization ability of Deep Belief Networks, they have been widely applied in the assessment of transient stability in power systems [6]. Literature [7] proposes the integration of the NSGA-II algorithm with intelligent optimization and Deep Belief Networks, significantly enhancing the speed of transient stability assessment and the accuracy of preventive control strategies.

2.3 The challenges and solutions of Deep Learning in the application of transient stability in power systems

Deep learning is highly effective in pattern recognition for transient stability assessments in power systems but faces challenges such as needing large volumes of high-quality data, which are scarce, and the models' "black box" nature that reduces transparency and interpretability. Additionally, issues with generalization, overfitting, and real-time computational demands complicate its use. To address these, we suggest augmenting training datasets with synthetic data and simulations, developing interpretable AI for better transparency, and improving model robustness with techniques like regularization and cross-validation. Optimizing computational infrastructure and implementing hardware acceleration are recommended for efficient real-time processing. Continuous learning and adaptive strategies will keep the models updated with system changes, enhancing deep learning's reliability and supporting improved decision-making in power systems.

2.4 Conclusion

Additionally, deep learning models leveraging Graph Convolutional Networks (GCN) and Graph Attention Networks (GAN) process electrical grid topology data and enhance stability factor recognition, respectively. These models boost assessment speed, accuracy, and handle complex power system models, including those with renewable energy sources, by autonomously learning features from data. Consequently, they reduce reliance on expert knowledge and improve generalization capabilities. Deep learning's potential and advantages in power system transient stability assessment suggest broad future implementation, offering robust technical support for safe and stable power system operation. As technology advances, deep learning's role in this field is expected to grow significantly.

3 INTERPRETABILITY METHOD ANALYSIS AND ITS POTENTIAL APPLICATIONS IN POWER SYSTEM TRANSIENT STABILITY ASSESSMENT

3.1 Inherently Interpretable Model Analysis and Its Application in Power System Transient Stability

Decision Trees.

The basic concepts of Decision Trees

Decision tree models [8] are structured for classification and regression tasks, organizing decision rules in a tree format. Nodes, branches, and leaves test features, show outcomes, and represent classes or values, respectively. Building these trees involves selecting optimal features and values for splitting, while managing growth and pruning to improve generalizability and prevent overfitting. Metrics like information gain, gain ratio, and Gini impurity guide splitting decisions, with a focus on entropy or dataset purity. Although interpretable and computationally efficient, decision trees may overfit and struggle with continuous variables. Enhanced by ensemble methods like random forests and gradient boosting, decision trees remain popular due to their transparent decision-making process.

The application of Decision Trees in power system transient stability

In literature [9], Decision trees find application in fault diagnosis within power systems, analyzing historical fault data to identify characteristic parameter combinations during faults. They aid operations and maintenance by pinpointing specific circuits or components implicated in transformer failures. Decision trees' interpretability lies in their clear decision paths, enabling personnel to comprehend the reasoning behind each prediction.

Generalized Additive Models (GAMs)

The basic concepts of Generalized Additive Models

Generalized Additive Models (GAMs) [10] are flexible statistical tools that explore non-linear relationships in data by using smooth functions instead of linear influences, uncovering hidden patterns without predefined variable relationships. Suitable for diverse fields like environmental science, finance, and healthcare, GAMs handle both continuous and categorical variables. While they offer interpretability since each model component is treated separately, challenges remain in choosing suitable smoothing functions and managing computational complexity. Despite these hurdles, GAMs are popular in various research areas for their ability to handle complex relationships and have demonstrated significant value in numerous studies.

The application of Generalized Additive Models in power system transient stability

In power systems, GAMs can be used to capture the non-linear relationships between load demand and factors such as time and weather conditions [11]. For example, they can reveal how to most effectively allocate electrical resources under specific weather conditions. The interpretability of GAMs is particularly beneficial for analysis and planning, as the model can provide intuitive understanding of which factors have significant impacts on the power system under specific conditions. The results of the model help improve the performance of the electrical grid and provide decision support for future infrastructure upgrades.

3.2 Post-hoc Interpretability Method Analysis and Its Application in Power System Transient Stability

Local Interpretable Model-Agnostic Explanations (LIME)

The basic concepts of Local Interpretable Model-agnostic Explanations

LIME, introduced by Ribeiro et al. in 2016 [12], enhances interpretability of complex machine learning models by generating a simpler, local model that approximates a complex model's behavior for individual predictions. It identifies key influencing features, useful in areas like medical diagnosis and financial risk assessment. However, LIME's local explanations may not fully capture the global behavior of the model, and explanations can vary with perturbations. Selecting effective local models and visualizing explanations are crucial for LIME's implementation.

The application of Local Interpretable Model-agnostic Explanations in power system transient stability

In literature [13], researchers used XGBoost to develop a transient stability prediction model by analyzing generator data during power system faults, applying the LIME method to explain contributions of individual features to predictions.This interpretability aids operational personnel in understanding and addressing issues based on the model's insights.

SHapley Additive exPlanations(SHAP)

The basic concepts of SHapley Additive exPlanations

SHAP, developed by Lundberg and Lee in 2017 [14], applies Shapley values from cooperative game theory to interpret model predictions by quantifying each feature's contribution. Originally for fair payoff allocation, SHAP treats each feature as a cooperative participant, calculating its impact on the prediction. By comparing predictions with and without each feature across all combinations, SHAP derives the feature's average impact. Widely used in healthcare, finance, and marketing, SHAP explains model decisions, like disease diagnosis or loan approval. However, it grapples with computational complexity and clarity for non-experts.

The application of SHapley Additive exPlanations in power system transient stability

SHAP has excelled in power systems as a supervised learning method. For instance, in literature [15], it's applied for attribution analysis in transient voltage stability assessment. By averaging Shapley values, this approach ranks feature importance, clarifying each feature's contribution to predictions. This quantification demonstrates SHAP's effectiveness in improving accuracy and interpretability in power system tasks.

3.3 Conclusion

The application of deep learning in power systems has enhanced the ability to handle complex data and make precise predictions, but it has also created a need for model interpretability. By incorporating interpretability methods such as LIME, Grad-CAM, and SHAP, model transparency can be enhanced, providing visualization and understanding of the decision-making process, thereby increasing user trust in model predictions. Although current interpretability methods still have limitations, they have made significant progress in understanding and explaining deep learning models, especially in critical areas such as the identification of power system states.

4 CONCLUSION AND OUTLOOK

This paper provides a review of current transient stability assessment methods in power systems using deep learning, emphasizing the importance of interpretability. It covers various deep learning networks and their applications in this domain. Interpretability methods, categorized into inherently interpretable models and post-training methods, are discussed, highlighting their roles in enhancing model credibility and supporting grid control decisions. Future research in explainable AI is anticipated to advance intelligence in power systems.

ACKNOWLEDGMENTS

This work was financially supported by the Science and Technology Project of China Southern Power Grid (No.GZKJXM20232014).

REFERENCES

1. Shuang Wu;Le Zheng;Wei Hu;Rui Yu;Baisi Liu,"Improved Deep Belief Network and Model Interpretation Method for Power System Transient Stability Assessment" [J]. Journal of Modern Power Systems and Clean Energy,2020(01).
2. J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.

3. Ren C, Xu Y, Zhang R. An interpretable deep learning method for power system transient stability assessment via tree regularization[J]. IEEE Transactions on Power Systems, 2021, 37(5): 3359-3369.
4. Chunyu Chen;Mingjian Cui;Fangxing Fran Li;Shengfei Yin;Xinan Wang. " Model-Free Emergency Frequency Control Based on Reinforcement Learning," in IEEE Transactions on Industrial Informatics., vol. PP, no. 99, pp. 1-1,.2020, doi: 10.1109/TPWRS.2021.3084469.
5. G. E. Hinton, S. Osindero and Y. Teh, "A Fast Learning Algorithm for Deep Belief Nets," in Neural Computation, vol. 18, no. 7, pp. 1527-1554, July 2006, doi: 10.1162/neco.2006.18.7.1527.
6. Jürgen Schmidhuber, " Deep learning in neural networks: An overview," in Neural Networks, vol. 61, pp. 85-117, 2015, doi: 10.1016/j.neunet.2014.09.003.
7. Xia Shiwei, Luo Xiao, Chan K W, et al. "Probabilistic transient stability constrained optimal power flow for power systems with multiple correlated uncertain wind generations. "IEEE Transactions on Sustainable Energy, pp. 1113-1144, 2016.
8. Dhyan Chandra Yadav;Saurabh Pal, " Prediction of thyroid disease using decision tree ensemble method," in Human-Intelligent Systems Integration, vol. 2, pp. 1-7, 2015, doi: 10.1007/s42454-020-00006-y.
9. Yan Ding;Shan Zhong;Lei Hua, " Automatic Recognition of Student Emotions Based on Deep Neural Network and Its Application in Depression Detection," in Journal of Medical Imaging and Health Informatics, vol. 10, pp. 2634-2641, 2020, doi: 10.1166/jmihi.2020.3265.
10. HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [J]. arXiv preprint arXiv:1704.04861, 2017
11. Hambali, Moshood & Akinyemi, & Oladunjoye, M. & N., Yusuf. (2017). Electric Power Load Forecast Using Decision Tree Algorithms. 7. 29-42..
12. Pavlo Sokolov;Muhammad Shakeel Virk, " Analytical parametrizations of droplet collision efficiency on cylinders – A review study," in Cold Regions Science and Technology, vol. 155, pp. 314-326, 2015, doi: 10.1016/j.coldregions.2018.08.015.
13. RIBEIRO M T, SINGH S, GUESTRIN C."Why should I trust you?": explaining the predictions of any classifier[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2016, San Francisco, USA: 1135-1144.
14. Yufei Xia;Chuanzhe Liu;YuYing Li;Nana Liu, " A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," in Expert Systems With Applications. vol. 78, pp. 225-241, 2017, doi: 10.1016/j.eswa.2017.02.017.
15. LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, USA: 47684777.