

Tentative Outline

Special Thematic Issue for the journal Recent Advances in Electrical & Electronic Engineering

Reliable and Efficient Interpretable Machine Learning Methods Analysis and Design

Guest Editor: Shiping Wen, Guogang Wen, Ke Li

- **Scope of the Thematic Issue:**

In the context of machine learning, interpretability refers to the degree to which a model can be explained and presented using human-perceivable terms, and it can also generally refer to the degree to which people can consistently predict the results of a model. The incompleteness of problem formulation, human curiosity and learning ability drive the need for interpretability of machine learning models, which means that when people use machine learning to solve some specific problem, they are no longer limited to just getting the prediction, but asking the model to explain how to get the prediction. Three important aspects of model interpretation are: (1) what drives the model's prediction; (2) why the model makes the prediction; (3) whether people can trust the model's prediction. The higher the interpretability of a machine learning model, the easier it is for people to understand why the model made certain decisions/predictions. The more interpretable a machine learning model is, the easier it is for people to understand why the model made certain predictions, thus establishing a certain level of trust in the machine learning model.

Nowadays, mainstream machine learning models are still performance-oriented. In order to have excellent accuracy, the scale and complexity of the model will be greatly improved, but the cost of doing so is to sacrifice the interpretability of the model, resulting in these high-performance models. It is difficult to explain why the prediction was made. If people cannot fully trust machine learning models, these models with excellent performance will not be able to be used in fields with low fault tolerance such as finance and healthcare. In-depth research on interpretable machine learning models in different application domains is significant and forward-looking. This special issue will bring together academic and industrial researchers for an in-depth discussion and detailed analysis of the cutting-edge technological achievements and major challenges in interpretable machine learning.

Keywords: interpretable learning; machine learning; reliable model; interpretable model; self-interpretation learning; efficient model.

Sub-topics:

The sub-topics to be covered within the issue should be provided:

- Interpretation methods based on feature summary statistics
- Interpretation methods based on data points
- Interpretation methods based on weights and model structure (model internals)
- Inherently interpretable models
- Application of visualization in interpretable machine learning
- Human-friendly interpretable machine learning
- Model-agnostic interpretation methods/tools
- Robots and programs with self-interpretation
- Self-driving based on interpretable machine learning
- Precision medicine based on interpretable machine learning
- Intelligent transportation based on interpretable machine learning
- The application of interpretable machine learning in high-risk engineering projects
- Research on the robustness of interpretable machine learning models

Tentative titles of the articles:

1. **Paper names:** Evaluating the performances of several interpretable artificial intelligence methods in

simulating nonlinear daily streamflow time series of hydropower reservoir

Novelty:

- Interpretable artificial intelligence methods provide new insights for daily runoff prediction.
- Several AI-based methods in daily runoff simulations are firstly compared.
- Various input variables and structures are used to choose the best models.
- AI-based methods obtain better performances than conventional methods.

Relationship: The emerging interpretable artificial intelligence methods produce satisfying forecasting performance as used to simulate the nonlinear daily streamflow time series.

2. **Paper names:** Deriving operation rule of power systems by interpretable machine learning optimized by swarm intelligence

Novelty:

- Input vectors are divided into disjointed clusters by the SOM method
- The input-output relationship per cluster is identified by machine learning.
- The model's hyper-parameters are optimized by swarm intelligence.
- The developed method obtains satisfying performance in real-world cases.

Relationship: Machine learning and swarm intelligence are dynamically integrated to improve the operation rule of hydropower reservoir.

3. **Paper names:** Interpretable generational adversarial network (GAN) model for long-term runoff prediction

Novelty:

- An efficient LN-GAN-SA model is presented by integrating layer normalization (LN), generational adversarial network (GAN), and stimulated annealing (SA) algorithm
- The hidden layer of the GAN model is construct by LN to accelerate the convergence speed, and the optimal parameters is determined by SA algorithm to balance the speed and stability of training.
- The hybrid LN-GAN-SA model has higher accuracy and stability than other models.

Relationship: The potential of GAN in long-term runoff prediction is proved while the hybrid LN-GAN-SA model has higher accuracy and stability than other models.

4. **Paper names:** Interpretable short-term operation optimization of cascade systems based on AI-based optimization method

Novelty:

- Short-term peak operation optimization of cascade systems is developed.
- AI-based optimization method is presented to dynamically search the scheduling scheme from a large amount of historical operation data
- The inaccurate simulation problem caused by basic data errors is avoided.

Relationship: AI-based method in guaranteeing the operational safety and reliability of multi-agent system is successfully proved.

5. **Paper names:** A novel interpretable model using VMD-based long short-term memory networks considering relevant factors with Bayesian optimization algorithm

Novelty:

- Nonlinear mapping is introduced to map the relevant factors.
- Bayesian Optimization Algorithm (BOA) is used in hyperparameter optimization.
- Proposed model periodically moves the data window which has high practicability.

Relationship: A hybrid interpretable model based on machine learning and parameters optimization is proposed.

6. **Paper names:** Dynamic prediction model via secondary decomposition model and interpretable machine learning tool in parallel computing

Novelty:

- Proposed an adaptive hybrid model to mitigate practical difficulties of traditional DEFMs.
- Integrating EEMD, AVMD and sample entropy for data preprocessing.
- The developed model with parameters optimization in parallel computing is built
- The hybrid decomposition method ASD made more contributions than BSA.

Relationship: This paper develops adaptive dynamic prediction model by linking data preprocessing into machine learning and parallel computing.

7. **Paper names:** High-performance parallel computing to determine the best operation schemes of adaptive water resources allocation under the changing environments

Novelty:

- Adaptive operation model for water resources allocation in the changing environments is proposed
- High-performance parallel computing is used to improve execution efficiency and objective values.
- The developed method produces better performances than several control methods.

Relationship: This paper provides a novel high-performance parallel interpretable computing method for smart resources management.

8. **Paper names:** Hybrid modeling in semi-humid region based on interpretable artificial intelligence coupled with metaheuristic optimization

Novelty:

- Hybrid modeling in semi-humid region is proposed.
- Interpretable artificial intelligence coupled with metaheuristic optimization can yield high performance.
- The developed method outperforms several benchmark models in the real-world experiments.

Relationship: Interpretable artificial intelligence approach and metaheuristic optimization are used to develop a novel hybrid flood susceptibility modeling in semi-humid region.

Schedule:

- ✧ Thematic issue submission deadline:
- Paper submission deadline: April. 1, 2023
- Notice of the first-round review results: August. 1, 2023
- Revision Due: October 1, 2023
- Final Notice of Acceptance/Reject: December 1, 2023

Contacts:

Guest Editor Name: Shiping Wen

Affiliation: University of Technology Sydney

Email: shiping.wen@uts.edu.au

Guest Editor Name: Guoguang Wen

Affiliation: Beijing Jiaotong University

Email: guoguang.wen@bjtu.edu.cn

Guest Editor Name: Ke Li

Affiliation: University of Exeter

Email: k.li@exeter.ac.uk