

A Reinforcement Learning-Based Two-Modes Surrogate-Assisted EA for the Inversion of Soil Thermal Resistivity for Power Cables

Yuhan Jiang, Shiyong Yang

College of Electrical Engineering, Zhejiang University, Hangzhou, 310027, China, eesyyang@zju.edu.cn

It is essential to have an accurate soil thermal resistivity to determine the real time ampacity to make full use of a power cable. However, the determination of the soil thermal resistivity involves the solution of an expensive multi-physical field inverse problem including a high-fidelity model, and the existing optimization methods still encounter some difficulties. In this respect, a reinforcement learning-based two-modes surrogate-assisted evolutionary algorithm (EA) is proposed. In the proposed algorithm, Mode 1 focuses on balancing the convergence and diversity while Mode 2 aims to accelerate the convergence. The search will adaptively switch between the two modes. Using the reinforcement learning, two adaptive evolutionary operator selection strategies are designed to select an operator for each individual in the two modes. Also, a new ensemble approach of surrogate models is proposed to ensure the solution accuracy and robustness. The superiorities of the proposed method are confirmed by the optimization results of a test function and a prototype cable.

Index Terms—Ampacity, power cable, inverse problem, reinforcement learning, soil thermal resistivity.

I. INTRODUCTION

AMPACITY of a power cable is defined as the load current when the conductor temperature reaches a maximum allowable limit and is determined under the worst environmental conditions. In other word, the cable conductor temperature under this normal ampacity is lower than the allowable temperature limit in most conditions, resulting in an underuse of the cable. In this regard, it is desirable to monitor the conductor temperature to determine the real time ampacity to make full use of a cable in engineering application. Nevertheless, it is impossible to directly measure the conductor temperature for the constructed cable line. Moreover, due to the diversity in environmental parameters affecting conductor temperature in different regions and seasons, the constitutive parameters of the medium, especially the soil thermal resistivity (STR), will vary significantly and be difficult to measure online [1]. Thus, it is demanding to inverse the STR to calculate the conductor temperature based on the measured temperature of cable sheath. However, the inversion of the STR is essentially a computationally expensive multi-physical field inverse problem involving high-fidelity model such as finite element model (FEM). Although widely used to solve such problems, surrogate-assisted evolutionary algorithms (SAEAs) still face some challenges [2]. Firstly, most SAEAs either converge slowly or prematurely, as the same evolutionary operator are applied to all individuals without considering differences among individuals; Secondly, although multi-surrogate models can be used simultaneously to improve the robustness of the predicted results, an improper weight to weight the results of models generally leads to information loss and low accuracy.

To address the above inefficiencies, a reinforcement learning (RL) based two-modes surrogate-assisted evolutionary algorithm (RLTM-SAEA) is proposed. Mode 1 focuses on balancing the convergence and diversity while Mode 2 aims to accelerate the convergence. The search will adaptively switch between two modes. The quantum-inspired evolutionary algorithm (QEA) is utilized as the search engine [3]. Two adaptive evolutionary operator selection strategies (AEOSs)

based on RL are designed for two modes to select an operator. In the searching process, each individual will select an evolutionary operator based on the corresponding AEOS. Additionally, a new ensemble approach to finally compromise the results of different surrogate models is proposed to ensure the solution accuracy and robustness.

II. A REINFORCEMENT LEARNING-BASED TWO-MODES SURROGATE-ASSISTED EVOLUTIONARY ALGORITHM

A. *Q-Learning and Q-Table*

To equip the proposed RLTM-SAEA to have reinforcement learning ability, *Q*-learning [4], a widely used RL algorithm, is used to design the selection framework. In *Q*-learning, a *Q*-table is constructed to store *Q*-values, i.e., the expected cumulative rewards obtained by selected actions. Based on the current *Q*-table and its state, the agent tends to select the action with the highest *Q*-value. After the action, the agent obtains a reward and transits to a new state, prompting an update to the *Q*-table based on this feedback. Obviously, an appropriate reward strategy is essential for guiding the agent toward an optimal action.

B. *The Iterative Procedure of the Proposed RLTM-SAEA*

The proposed RLTM-SAEA adaptively switches between the two modes. It starts from Mode 1 until no improvement on the diversity of an *elite* group for some consecutive generations; the algorithm then switches to Mode 2 until no improvement on the current best fitness value over some consecutive generations. This procedure is repeated until a stop criterion is satisfied. Moreover, four rotation operators in the existing QEA are used as optional actions for each individual to increase the diversity of the operations, and will be detailed in the full paper.

C. *The AEOS for Mode 1*

Since Mode 1 focuses on balancing the convergence and diversity, *QI*-table is designed for operator selection based on both the fitness value and density of an individual. Consequently, the fitness ranking value of each individual and the density ranking value are used to encode the hierarchical state variables of the *QI*-table, and will be detailed in the full

paper for space limitations. To accelerate the convergence, a reward $r(x_i, u_i)$ is proposed based on fitness value from the i th current individual x_i and its offspring u_i , as shown in (1):

$$r(x_i, u_i) = \begin{cases} r_0 \{(f(u_i) - f(x_i))/f(x_i)\} & f(u_i) \geq f(x_i) \\ 0 & \text{else} \end{cases} \quad (1)$$

where r_0 is a user-defined positive base reward. Obviously, an individual is rewarded if it produces a better offspring.

To prevent offspring clustering, a penalty $c(u_i)$ is designed based on the density from the u_i , as shown in (2):

$$c(u_i) = \begin{cases} -r_0 \{(s(u_i) - \alpha)/\alpha\} & s(u_i) > \alpha \\ 0 & \text{else} \end{cases} \quad (2)$$

where $s(u_i)$ is the density of the i th offspring calculated according to [2], α is the density threshold. Obviously, an individual is punished if its offspring is too densely packed.

The total reward from the x_i and its offspring u_i is thus

$$R_1(x_i, u_i) = r(x_i, u_i) + c(u_i) \quad (3)$$

D. The AEOS for Mode 2

Since Mode 2 aims to accelerate the convergence of the algorithm, the fitness of an individual will be solely used to design $Q2$ -table for operator selection, other things being equal to those of Mode 1. Consequently, the fitness ranking value of each individual is used to encode the hierarchical state variables of the $Q2$ -table. The reward $R_2(x_i, u_i)$ is then designed as

$$R_2(x_i, u_i) = \begin{cases} r_0 \{(f(u_i) - f(x_i))/f(x_i)\} & f(u_i) \geq f(x_i) \\ -r_0 \{(f(x_i) - f(u_i))/f(x_i)\} & \text{else} \end{cases} \quad (4)$$

An individual is rewarded for producing a better offspring; otherwise, it is negatively awarded to ensure a fast convergence.

E. A New Ensemble Approach of Surrogate Models

To improve the robustness of the predicted results, a new ensemble approach of surrogate models without any weight is introduced. To start, Latin hypercube sampling is first implemented to generate an initial training database (DB). In each generation, firstly, one utilizes this DB to construct three different surrogate models: a polynomial regression, a radial basis function, and a Kriging model. Then, the three models are employed to predict the comparative relationship of the fitness values between all parents and their corresponding offspring in the current population. For a pair of an individual and its offspring $\langle x_i, u_i \rangle$, each model vote on the two comparative relationships (only "better" and "worse" cases are considered; if "equality" occurs, exact fitness evaluation is adapted). The relationship with the majority of votes is taken as the true result. The surrogate model earns one point for a correct prediction and zero for an incorrect prediction. Finally, the model with the highest score is employed to evaluate fitness in (1) or (4). The offspring with the highest fitness value is re-evaluated by a high-fidelity model and updated in DB. All scores are reset to zero for the next generation.

III. NUMERICAL EXAMPLES

To validate the proposed algorithm, a 10-dimensional Ackley function whose global optimum is 0, is first solved using the proposed algorithm and its performances are compared with a popular SAEA called EAPSO [2]. For a fair comparison, the two algorithms terminate after 400 exact fitness evaluations, and every algorithm runs randomly 20 times. The statistic performances of the two algorithms are tabulated in Table I. From these results, it is obviously that the proposed algorithm achieves a high accuracy and maintains a good stability in terms of all statistic performances.

TABLE I
OPTIMIZED RESULTS OF TEST FUNCTION

Algorithm	Best	Mean	Worst	SD
EAPSO	0.3313	1.0706	1.7029	0.6419
Proposed	0.0308	0.0843	0.1775	0.0554

To demonstrate the advantages of the proposed algorithm in inverting the STR of a power cable, a duct laying cable model is selected as the prototype. The soil consists of the backfill and native soils with different thermal resistivities. The proposed RLTM-SAEA and a standard QEA (SQEA) are employed to inverse the STR, respectively. The FEM is used as the high-fidelity model. The inversed results are summarized in Table II. Also, the conductor temperatures and real time ampacity under the inversed STRs by two algorithms are compared in Table II. It is evident that the STRs inversed by proposed algorithm are very close to those by SQEA while the former uses approximately 59.78% of the calls of FEM by the latter. The conductor temperatures and real time ampacity under the inversed STRs by proposed algorithm are very close to those by SQEA. Moreover, the computed real time ampacity achieves an approximate 27.50% increase compared to its normal ampacity (1023.86A). Consequently, the proposed approach significantly reduces computational burden without compromising the solution quality compared to the originally computationally intensive approach.

TABLE II
RESULTS AND PERFORMANCES OF DIFFERENT APPROACHES FOR THE
PROTOTYPE CABLE OPTIMIZATIONS

Algorithm	STR (K·m·W ⁻¹)		MAPE of conductor temp.	Real time ampacity (A)	No.FEM
	Backfill soil	Native soil			
Proposed	1.64	0.69	0.03	1305.41	550
SQEA	1.43	0.72	/	1314.83	920

MAPE is the mean absolute percentage error between conductor temperatures calculated by two algorithms, and No.FEM refer to calls of FEM.

REFERENCES

- [1] C. Bates, K. Malmedal, and D. Cain, "Cable ampacity calculations: A comparison of methods," *IEEE Trans. Ind. Appl.*, vol. 52, no. 1, pp. 112-118, Jan. 2016.
- [2] F. Li, X. Cai, and L. Gao, "Ensemble of surrogates assisted particle swarm optimization of medium scale expensive problems," *Appl. Soft Comput.*, vol. 74, pp. 291-305, Jan. 2019.
- [3] S. L. Ho, S. Yang, P. Ni, and J. Huang, "A quantum-inspired evolutionary algorithm for multi-objective design," *IEEE Trans. Magn.*, vol. 49, no. 5, pp. 1609-1612, May 2013.
- [4] P. Li, J. Hao, H. Tang, X. Fu, Y. Zhen, and K. Tang, "Bridging evolutionary algorithms and reinforcement learning: A comprehensive survey on hybrid algorithms," *IEEE Trans. Evol. Comput.*, Aug. 2024, to be published.