Monte Carlo Tree Search based Space Transfer for Black-box Optimization

Shukuan Wang[1,2]*, **Ke Xue**[1,2]*, **Lei Song**[1,2], **Xiaobin Huang**[1,2], **Chao Qian**[1,2]†
[1]National Key Laboratory for Novel Software Technology, Nanjing University
[2]School of Artificial Intelligence, Nanjing University
{wangsk, xuek, huangxb, songl, qianc}@lamda.nju.edu.cn

Background

Black-box optimization

- The objective function lacks a mathematical analytical form, can only be evaluated by inputs

$$argmax_{x \in \mathbb{X}} f(x)$$

Expensive black-optimization problems

- Problem evaluations comes with high computational or economic costs.
- Only a limited number of evaluations available

MCTS-transfer

Challenges of expensive BBO: How to efficiently utilize data and knowledge of similar tasks?

Transfer optimization

MCTS–transfer

Challenges of expensive BBO: How to efficiently utilize data and knowledge of similar tasks?
Transfer optimization

Assumption: Similar tasks share similar characteristics

- Thus, learning from similar source tasks can accelerate the optimization of target tasks.

MCTS-transfer

Challenges of expensive BBO: How to efficiently utilize data and knowledge of similar tasks?
Transfer optimization

Assumption: Similar tasks share similar characteristics

- Thus, learning from similar source tasks can accelerate the optimization of target tasks.

What is being transferred in Transfer optimization?

- Search space: extract promising subspaces
- Initialization: better initializations for warm-start
- Surrogate model: multi-task GP, deep kernel learning
- Acquisition function: balance between source and target task

MCTS–transfer

Challenges of expensive BBO: How to efficiently utilize data and knowledge of similar tasks?
Transfer optimization

Assumption: Similar tasks share similar characteristics
- Thus, learning from similar source tasks can accelerate the optimization of target tasks.

What is being transferred in Transfer optimization?
- Search space: extract promising subspaces
- Initialization: better initializations for warm–start
- Surrogate model: multi–task GP, deep kernel learning
- Acquisition function: balance between source and target task

common component can be applied to different optimizers

Motivation

Transfer optimization on search space

- Space pruning [Wistuba et al., PKDD'15]
- Hypercube/ellipsoid space extraction [Perrone et al., NeurIPS'19]
- High-quality subspace integration [Li et al. KDD'22]

More applicable when source tasks are similar to target tasks;

Motivation

Transfer optimization on search space

- Space pruning [Wistuba et al., PKDD’15]
- Hypercube/ellipsoid space extraction [Perrone et al., NeurIPS’19]
- High-quality subspace integration [Li et al. KDD’22]

More applicable when source tasks are similar to target tasks;

However, it’s difficult to identify task similarities in advance in BBO

Key questions:

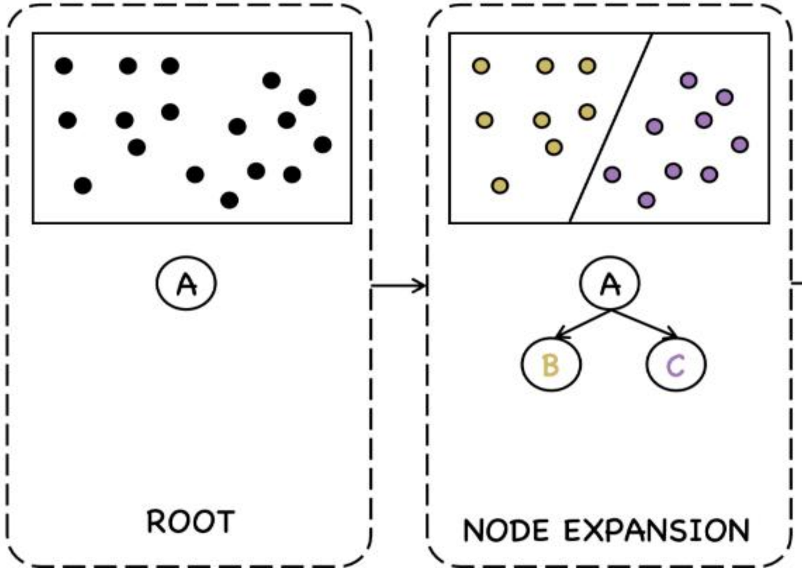
How to automatically identify the most relevant source tasks and leverage their information for search space transfer?

MCTS–transfer

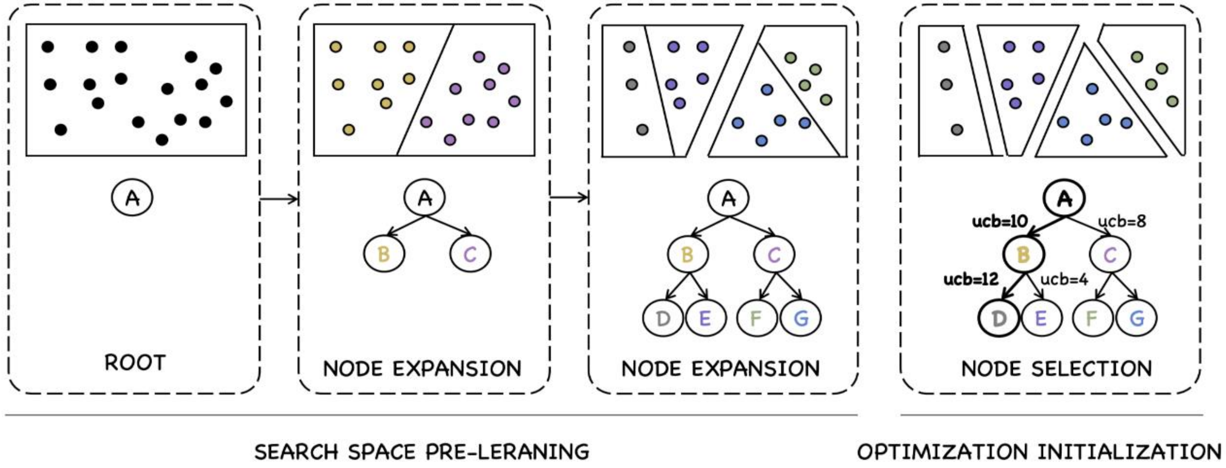
Main Idea: leverage MCTS to divide search spaces considering the source task

Modeling Assumption: the division of the search space corresponds to the splitting of tree nodes (A=B+C)

Tree Building Rule: for any node, the potential value of the left child node is higher than that of the right child node (B>C)



MCTS–transfer

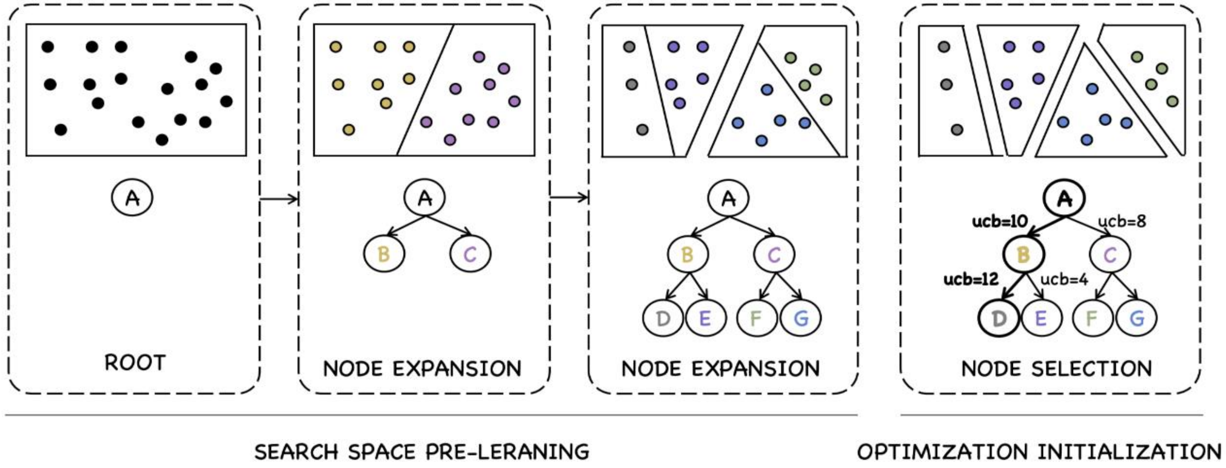


Pre–learning stage: divide the search space based on source task data

- Start from the root node, for node m, if it can be split:
  - Use K–Means to divide the samples in node into two clusters
  - Use a binary classifier to separate the two clusters and divide the space
  - Calculate node potential
    - The cluster with higher potential be the left child node

$$p_m = \frac{\sum_{i \le K} \sum_{(\boldsymbol{x}_{i,j}, y_{i,j}) \in D_i \cap \Omega_m} y_{i,j}}{\sum_{i \le K} |D_i \cap \Omega_m|}.$$

MCTS–transfer



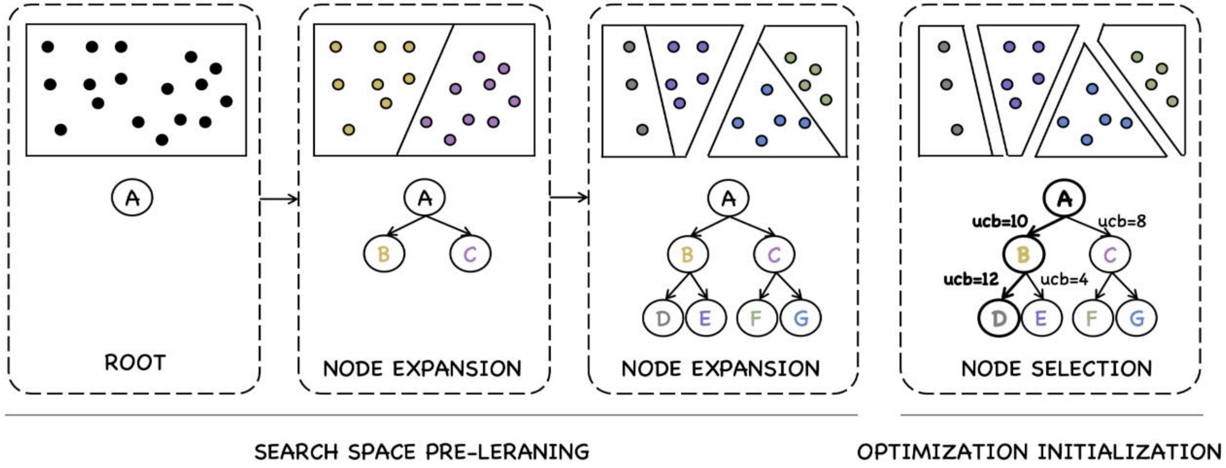
Optimization stage

- **Node Selection**: select the node *m* with higher UCB from ROOT

$$\mathrm{ucb}_m = \frac{\boxed{p_m}}{n_m} + 2C_p\sqrt{2\log\left(n_p\right)/n_m} \qquad p_m = \boxed{\gamma^{t-1}}\frac{\sum_{i\le K}\boxed{w_i}\,\bar{y}_{i,m}}{\sum_{i\le K} w_i} + \bar{y}_{T,m}$$

- **Simulation**: sample in the node *m*, select the query point *x* by acquisition function and evaluate *x*

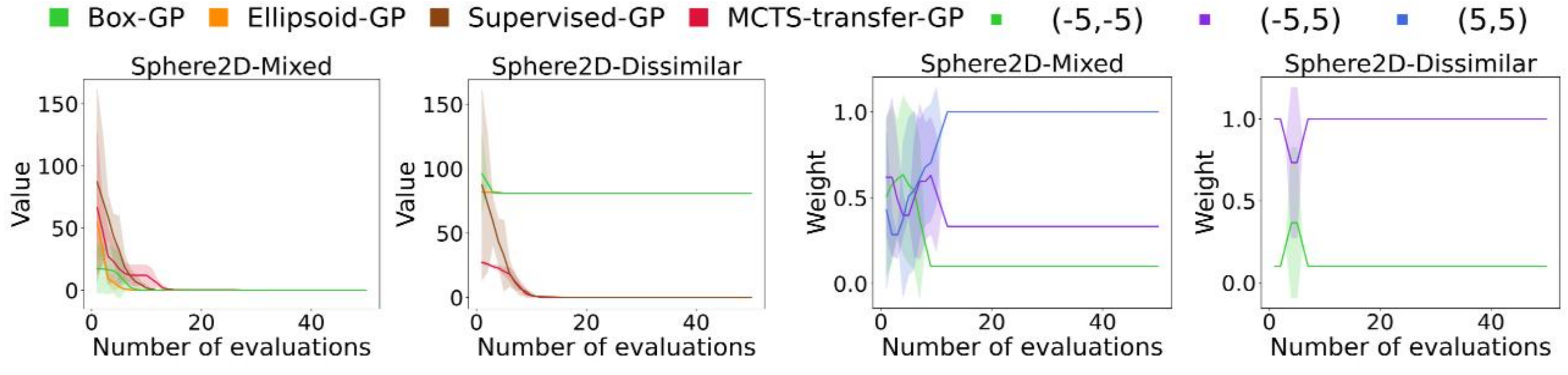
MCTS–transfer



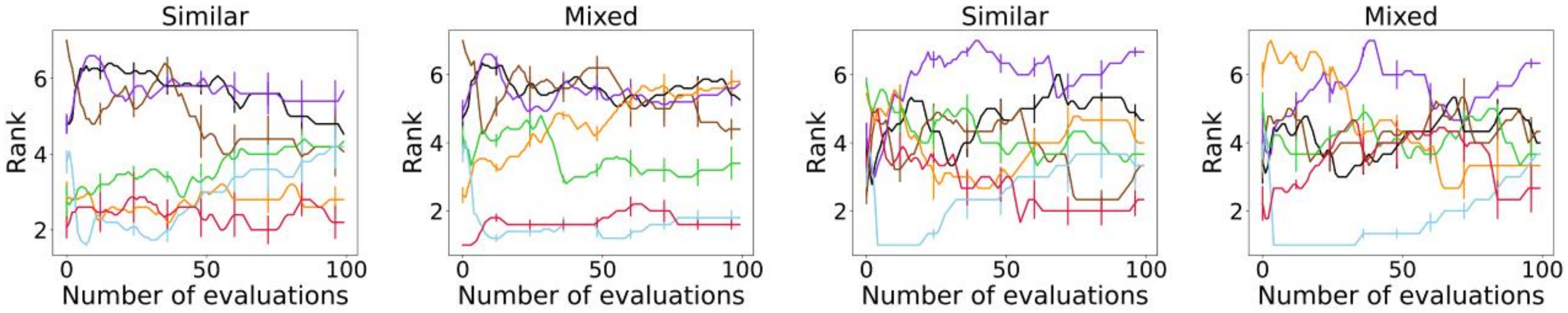
Optimization stage

- **Node Expansion**: expand all splitable nodes
- **Backward Update and Tree Reconstruction**:
  - Update similarity, ranking, and task weights w_i based on the evaluation
  - Update node potential based on w_i
  - check for any subtrees that violate the tree-building rules; if found, reconstruct the subtree.

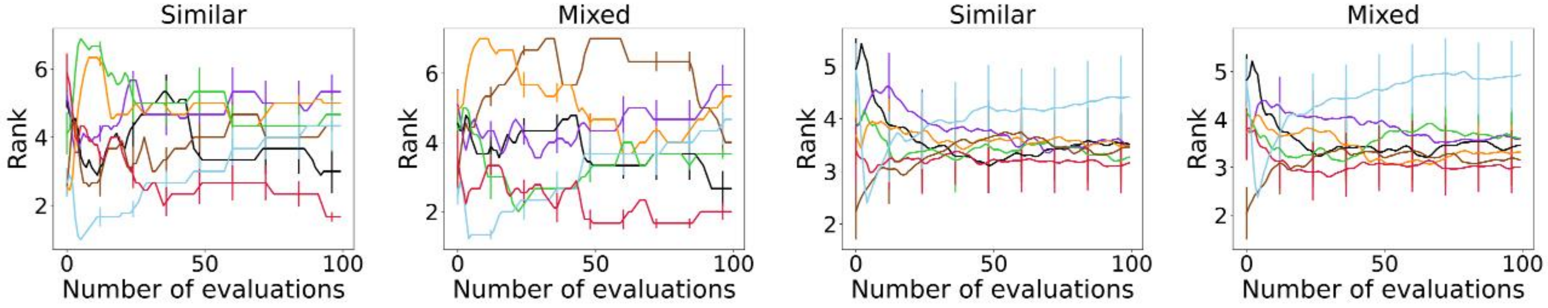
Comparasion Experiments



(a) Best value (b) Weight change curves



(a) BBOB (b) Real-world problems



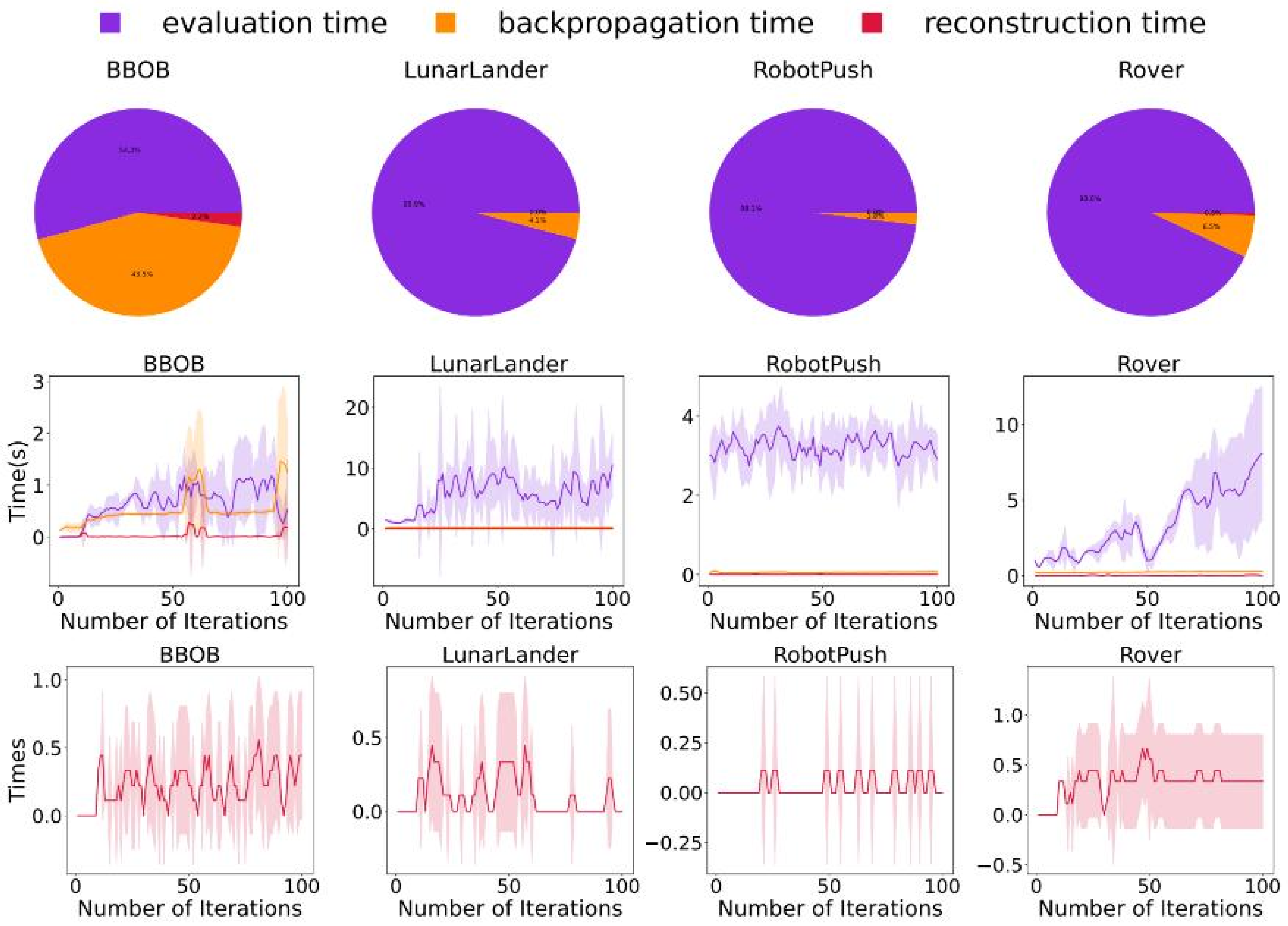
(a) Design-Bench (b) HPOB

validity: superiority especially in mixed and dissimilar setting

reasonableness： weights of similar tasks are higher

robustness: stable performance in complex or high-dimensional problems

Runtime Analysis



Computational cost:
the additional computation overhead is minor

Reconstruction frequency：
average frequency is low

Conclusion

Advantages

- Automatically identify similar source tasks and assign greater weight to fully utilize source task, accelerating the optimization process
- Dynamically adjust tree structure to increase the probability of the optimal solution locating in the best leaf node
- MCTS-transfer has surpassed the baselines in numerous experiments with introducing minor computational overhead

Future Work

- Extend MCTS-transfer to heterogeneous search space transfer
- Explore more node potential measurement methods
- Build theoretical guarantees

Thank you!