Main Ablation Study

We vary the main ingredients of BANANAS [2] in a factorial manner and benchmark the different configurations on NAS-Bench-301 [1].

Results are given in Figure 2. The choice of acquisition function optimizer is by far the strongest determinant of performance as also indicated by a four-way ANOVA on the final performance (Table 1).

Further Investigations

- Based on the Tabular + RF + EI + Mut configuration, the BO loop was run for 50 iterations (architecture evaluations).
- For edit distances ranging from 1 to 8, 100 test architectures were constructed each by mutating a fixed number of parameters (operations or edges) of the incumbent.
- For these test architectures, Kendall’s τ with respect to the predicted and true validation accuracy (Figure 3A), their true validation accuracy (Figure 3B) and their expected improvement and actual improvement (Figure 3C) was calculated.

TL;DR

- We vary the main ingredients of BANANAS, architecture encoding, surrogate candidate, acquisition function and acquisition function optimizer in a factorial manner and benchmark the different configurations on NAS-Bench-301.
- The choice of acquisition function optimizer (by default, BANANAS minimally mutates the incumbent architecture) is by far the strongest determinant of performance.
- Results hint that the surrogate model is not able to differentiate high-performing architectures well.
- Therefore, exploration of architectures or thorough optimization of the acquisition function may not be needed but minimally mutating the incumbent architecture is all you need.

Table 1: Results of a four-way ANOVA on the factors architecture encoding, surrogate candidate, acquisition function, and acquisition function optimizer. Type II sums of squares. Results are based on 20 replications.

References