HPOBench: A Collection of Reproducible Multi-Fidelity Benchmark Problems for HPO

/K/aut/oml/HPOBench
Why should you care?

Available data is growing.
Applications of ML are growing.
Model size and complexity is growing.

We need efficient hyperparameter optimization methods!

→ Multi-fidelity optimization

BUT

To develop, improve, understand and compare methods we need benchmark problems that are realistic, efficient and available for a long time.

→ HPOBench
Contributions

- The first collection of **containerized multi-fidelity** HPO benchmarks with 100+ benchmark problems

- The first set of HPO benchmarks that
  - are available as **raw and tabular** versions
  - which also support **multi-objective optimization** and **transfer-HPO across datasets**

- An exemplary **large-scale study** evaluating >10 optimization methods on all benchmarks
HPO Benchmarks

Benchmark ingredients:

Ideal features:

Efficiency

Reproducibility

Flexibility
Efficiency

- **raw benchmark**
  - algo
  - data
  - loss

- **tabular benchmark**
  - $f(\lambda, b_1)$
  - $\cdots$
  - $f(\lambda, b_m)$

- **surrogate benchmark**
  - ML model predicts

---

- **Optimizer**
  - query
  - observe
  - configuration space
  - fidelity space
Unify interface → Collect dependencies → Package everything in a container
## Flexibility

<table>
<thead>
<tr>
<th>Family</th>
<th>#benchs</th>
<th>#cont(log)</th>
<th>#int(log)</th>
<th>#cat</th>
<th>#ord</th>
<th>fidelity</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartpole</td>
<td>1</td>
<td>4(1)</td>
<td>3(3)</td>
<td>-</td>
<td>-</td>
<td>repetitions</td>
<td>✓</td>
</tr>
<tr>
<td>BNN</td>
<td>2</td>
<td>3(1)</td>
<td>2(2)</td>
<td>-</td>
<td>-</td>
<td>samples</td>
<td>✓</td>
</tr>
<tr>
<td>Net</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>time</td>
<td>✗</td>
</tr>
<tr>
<td>NBHPO</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>6</td>
<td>epochs</td>
<td>✓  ✓</td>
</tr>
<tr>
<td>NB101</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>26</td>
<td>-</td>
<td>epochs</td>
<td>✓  ✓</td>
</tr>
<tr>
<td>NB201</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>epochs</td>
<td>✓  ✓</td>
</tr>
<tr>
<td>NB1Shot1</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>9</td>
<td>-</td>
<td>epochs</td>
<td>✓  ✓</td>
</tr>
<tr>
<td>LogReg</td>
<td>20</td>
<td>2(2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>iter</td>
<td>✓, ✓</td>
</tr>
<tr>
<td>SVM</td>
<td>20</td>
<td>2(2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>data</td>
<td>✓, ✓</td>
</tr>
<tr>
<td>RandomForest</td>
<td>20</td>
<td>1</td>
<td>3(2)</td>
<td>-</td>
<td>-</td>
<td>#trees</td>
<td>✓, ✓</td>
</tr>
<tr>
<td>XGBoost</td>
<td>20</td>
<td>3(2)</td>
<td>1(1)</td>
<td>-</td>
<td>-</td>
<td>#trees</td>
<td>✓, ✓</td>
</tr>
<tr>
<td>MLP</td>
<td>8</td>
<td>2(2)</td>
<td>3(2)</td>
<td>-</td>
<td>-</td>
<td>epochs</td>
<td>✓, ✓</td>
</tr>
</tbody>
</table>
We ran >10 optimization methods on all benchmarks and studied the following:

1. Do advanced methods improve over random baselines?

2. Do multi-fidelity methods improve over single-fidelity methods

→ Short answer: Yes
Conclusion

→ **HPOBench** provides >100 containerized benchmarks for multi-fidelity HPO

What else you can do with HPOBench:

- multi-objective optimization and transfer-HPO across datasets
- compare raw, tabular and surrogate benchmarks
- ...

Thank you!

Twitter: /AutoML_org

/automl/HPOBench

1 Albert-Ludwigs-Universität Freiburg
2 Leibniz Universität Hannover
3 Amazon (work done prior to joining Amazon)
4 Bosch Center for Artificial Intelligence