

# コストを考慮した ハイパーパラメーター 最適化の重要性について

実際に重要なコストメトリックを把握する

Eric Hans Lee SigOpt リサーチエンジニア  
Michael McCourt SigOpt エンジニアリング責任者

この記事では、[Conference on Uncertainty in Artificial Intelligence](#)（英語）で発表した弊社のコスト制約のあるベイズ最適化への非短絡的アプローチ（英語）に関する研究を紹介します。この研究は、Facebook や Amazon のサイエンティストと共同で行われました。この要約では、この研究の動機となった要因について説明します。

## 通常のハイパーパラメーター最適化：反復による進捗状況の測定

ほとんどの実用的なハイパーパラメーター最適化パッケージは、一定の反復回数を実行して最適なハイパーパラメーターを決定しようとします。例えば、HyperOpt\*、Optuna\*、SKOpt、SigOpt を 100 回ずつ反復する例について考えてみます。

<p><b>HyperOpt* :</b></p> <pre>hyperopt.fmin(     objective,     space = search_space,     max_evals = num_iterations)</pre>	<p><b>Optuna* :</b></p> <pre>study = optuna.create_study() study.optimize(     objective,     n_trials = num_iterations)</pre>
<p><b>SKOpt :</b></p> <pre>skopt.gp_minimize(     func=objective,     n_calls = num_iterations)</pre>	<p><b>SigOpt :</b></p> <pre>sigopt_conn = Connection() for _ in range(num_iterations):     suggestion =         conn.experiments().         suggestions().         create()      value = objective(         suggestion.assignments)     conn.experiments().     observations().     create(         suggestion = suggestion.id,         value = value)</pre>

ほとんどのオプティマイザーのインターフェイスは基本的に同じで、最大化 / 最小化したいもの（最適化の目的）と実行する反復回数を入力します。多くの開発者はこのインターフェイスを当たり前のように使っていますが、これは本当に目的を最適化する最良の方法なのでしょうか？ 累積トレーニング時間などが重要である場合に反復回数を求めることは、パフォーマンスを重視していないと言えます。

## 課題：ハイパーパラメーターの評価コストのばらつき

最適化の進捗状況を反復回数で測定することは、各評価にかかる時間が同じであれば合理的ですが、ハイパーパラメーター最適化（HPO）では、ハイパーパラメーターにより設定のトレーニング時間が大きく異なる場合があります。前述の研究では、最も一般的に使用されている 5 つのマシンラーニング・モデルでこれを確認しています。

- K 近傍法 (KNN)
- 多層パーセプトロン (MLP)
- サポート・ベクトル・マシン (SVM)
- 決定木 (DT)
- ランダムフォレスト (RF)

これらのモデルは、データサイエンティストが使用する一般的なモデルの大部分を構成しています。ここでは、ディープラーニング・モデルは省略していますが、これらのモデルについても結果は同じです。

5つのモデルは、一般的なベンチマークである OpenML\* w2a (英語) データセットで、標準の探索空間からランダムに選択した 5,000 個のハイパーパラメーター設定を使用してトレーニングしました。そして、各モデルのトレーニング時間の分布を図にしました (図 1)。各モデルのトレーニング時間は大きく異なり、1 桁以上の差があることも珍しくないことがわかります。これは、各モデルにおいて、いくつかのハイパーパラメーターが、モデルのパフォーマンスだけでなく、トレーニング時間にも大きく影響するためです (ニューラル・ネットワークの層のサイズや、フォレストのツリーの数など)。実際に、ほとんどすべての実用アプリケーションにおいて、評価コストは探索空間の異なる領域で大きく変化することが分かっています。

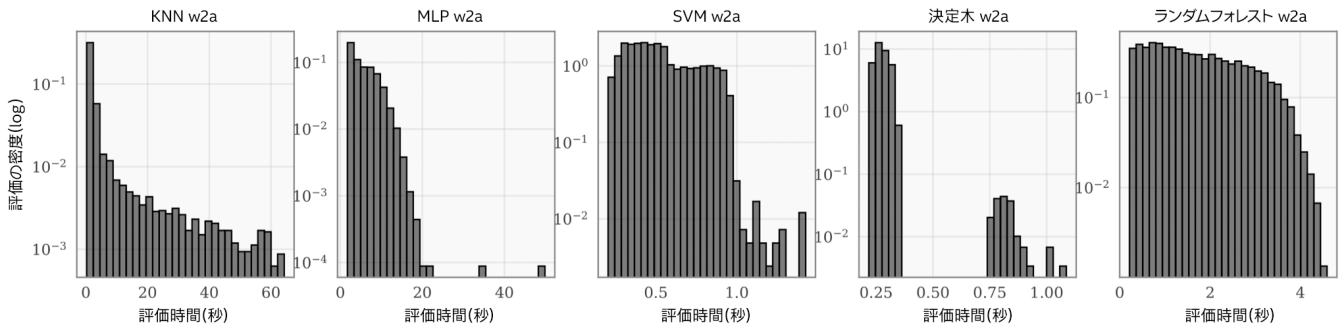


図 1. KNN、MLP、SVM、DT、RF の各ハイパーパラメーター最適化問題について、ランダムに選択した 5,000 点のランタイム分布 (ログスケール)。各ヒストグラムの X 軸はランタイム、Y 軸は密度 (全評価数に対する割合) を示します。

そのため、これらのハイパーパラメーターのチューニングにかかる累積トレーニング時間は、反復回数に正比例しません。実際、図 1 のヒストグラムから、ある最適マイザーが 1 つのハイパーパラメーター設定を評価し、別の最適マイザーが 100 個のハイパーパラメーター設定を評価した場合に、両者の時間が同じになる可能性は十分にあります。

## 課題 : 多様なコストへの対応

SigOpt では、目的関数の多様なコストを考慮し、次にどこを評価するかを判断する最適マイザーを開発しました。これは、コストを考慮した最適化ルーチンとして知られています。この最適化ルーチンは、反復回数ではなく、時間などのコストメトリックを考慮して最適化を判断します。例えば、最適化ルーチンに 100 回の反復で最適なハイパーパラメーターを計算するように指示する代わりに、100 分のトレーニング時間で最適なハイパーパラメーターを計算するように指示することができます。この最適化ルーチンは、次のように呼び出します。

```
optimizer.minimize(func=objective, num_minutes=100)
```

ハイパーパラメーター最適化を制約するものは、費用や時間などユーザーにより異なるため、ハイパーパラメーター最適化コミュニティではこの重要な研究課題に積極的に取り組んでいます。

## コストを考慮するメリット

SigOpt では、「コストを考慮したベイズ最適化 (CA-BO)」アルゴリズムを開発しました。簡単に説明するため、当社の CA-BO 手法と Preferred Networks 社の優れたオープンソース・ツールである [Optuna\\*](#) (英語) を比較する例について考えてみましょう。どちらのオプティマイザーも、100 回の最適化反復で XGBoost モデルの 2 項分類精度を最大化します。反復回数が増えるにつれて、Optuna\* が CA-BO を徐々に上回ることが分かります (図 2)。パフォーマンスの差は明らかです。何が起きているのでしょうか？ CA-BO が少なくとも Optuna\* と同等のパフォーマンスを発揮することを期待しましょう。

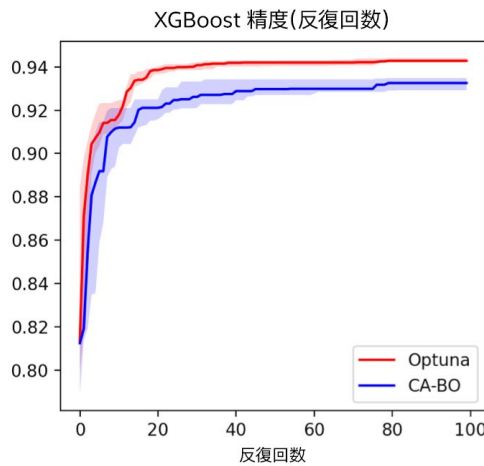


図 2. 最適化の反復回数を固定した場合の CA-BO と Optuna\* の比較

この差は、CA-BO がコストを考慮しているためです。評価コストが変化することを認識し、それに応じた判断をしています。X 軸を反復回数ではなく、累積トレーニング時間に置き換えて、評価コストの変化を考慮した場合の最適化パフォーマンスを見てみましょう (図 3)。CA-BO のほうが、はるかに良いパフォーマンスであることが分かります。Optuna\* は最終的に CA-BO を上回りますが、それは CA-BO が目的の時間内で実行し終わった後のことです (目的の時間内では、CA-BO が Optuna\* を常に上回っています)。

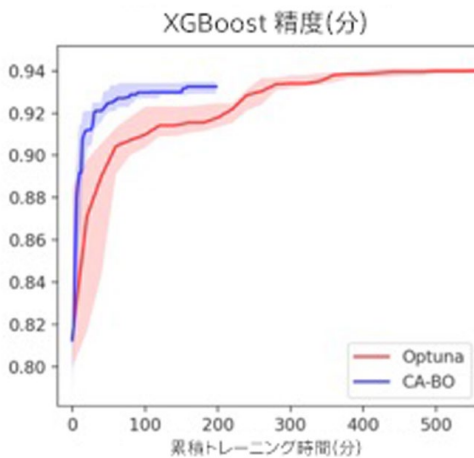


図 3. トレーニング時間に基づく CA-BO と Optuna\* の比較

対照比較では、コストを考慮した最適化ルーチンとコストを考慮しない最適化ルーチンの差が顕著に表れています (図 4)。右のコストを考慮したアプローチの図では、CA-BO が 200 分以内に優れたハイパーパラメーターを見つけていることが分かります。左のコストを考慮しないアプローチの図では、トレーニングに 500 分以上かかる精度の低いハイパーパラメーターを 100 回の反復で見つけていることが分かります。

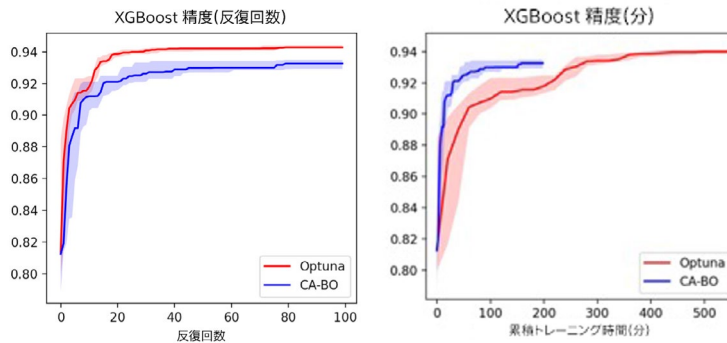


図 4. コストを考慮した最適化ルーチンとコストを考慮しない最適化ルーチンの対照比較

左の図は HPO の現状を示しています。右の図は、モデルをプロダクション環境に移行する際に重要なこと、つまり、考慮すべきメトリックに基づいてインテリジェントな決定を行う手法を示しています。

## まとめ

CA-BO は、急速に進化している HPO アルゴリズムです。CA-BO の詳細は、SigOpt のアプローチの技術的な詳細を含む論文 (英語) を参照してください。ほかの最適化手法やアプリケーションについては、SigOpt の Research ページ (英語) を参照してください。このような手法は、SigOpt Intelligent Experimentation Platform (英語) の柱の 1 つである SigOpt オプティマイザーに組み込まれています。今すぐサインアップ (英語) して、オプティマイザーとその他のプラットフォームを無料でお試しいただけます。