Uncertainty Estimation with Online Conformal Prediction in Neural Architecture Search: An Evaluation under the BANANAS Framework

Cheng Chen

February 24th, 2025

1 Introduction

Over the last decade, deep learning models have emerged in an unprecedented speed and have proven to be successful in a wide range of machine learning tasks. Consequently, Neural architecture search (NAS), the process of automating the neural architecture design, is a natural next step of building high-performance neural networks. Recently, the framework "Bayesian Optimization (BO) + neural predictor" is proved to be efficient in NAS. In particular, [5] identifies five components of this framework and performs a thorough analysis on each components importance towards the performance. Based on both theoretical and empirical findings, [5] proposes BANANAS, an NAS algorithm that jointly leverages architecture path encoding and Bayesian optimization and has achieved state-of-the-art performance on NAS search spaces.

Despite of its provable strength, BANANAS still assumes a Gaussian distribution in uncertain calibration, one of the five identified components in the framework. However, this assumption does not necessarily hold and might limit the application of NAS in real world. In order to attain higher-quality uncertainty estimate, this thesis focuses on applying Conformal Prediction (CP), a distribution-free framework for uncertainty calibration, under the framework of BANANAS.

2 Background

The BO side in the "Bayesian Optimization+neural predictor" framework requires two components: a surrogate model to approximate the unknown objective function, which predicts the performance of neural architectures that have not been evaluated yet; and an acquisition function, which balances between exploitation and exploration to pick up the next data point for evaluation. In order to evaluate the acquisition function, uncertainty estimate for each input datapoint is required as well. [5] assumes that the probability distribution of each datapoint x is Gaussian. Therefore, the problem of uncertainty estimation is reduced to variance estimation, though at a cost of estimate precision. To mitigate this constraint, this thesis explores the potential of applying CP-based calibration to bypass the Gaussian assumption. Despite of zero assumption on the underlying data distribution, CP-based calibration comes with other virtues:

- **Balancing exploration-exploitation:** Bayesian optimization uses a probabilistic model to make exploration-exploitation decisions. In less explored regions, the confidence interval around the ground truth should be large to promote exploration. Calibration helps mitigate overconfidence and promotes accurate confidence intervals that encourage exploration [1].

- **Simple and lightweight:** Two popular ways to quantify uncertainties are by using a Bayesian neural network (BNN), or by using an ensemble of neural predictors. For example, [5]

Algorithm 1 Estimate uncertainty using Conformal Prediction in BANANAS

Input: NAS parameters: Search space $\mathcal{A}$, dataset $\mathcal{D}$, exploration budget T, a neural predictor $\mathcal{M}$, acquisition function ϕ, surrogate model $f(\cdot)$ returning validation error of an architecture after training; CP parameters: t_0, p_{train}, an array of desired coverage levels $\boldsymbol{\alpha}$, and a conformity score function $s(\cdot)$.

1: Draw t_0 architectures $\{a_0, a_1, ..., a_{t_0}\}$ uniformly at random from $\mathcal{A}$ and train them on $\mathcal{D}$.
2: $\mathcal{A}_{t_0} \leftarrow \{a_0, a_1, ..., a_{t_0}\}$,
3: **for** t in $t_0, ..., T$ **do**

    i   Randomly select a proportion p_{train} from the trained t_0 architectures as the training set $\mathcal{A}_{t,train}$, use the remaining architectures as the calibration set $\mathcal{A}_{t,cal}$.

    ii   Train a neural predictor $\mathcal{M}$ on $\{a, f(a)\}, a \in \mathcal{A}_{t,train}$ using the path encoding to represent each architecture.

    iii   Compute the respective quantile of the conformity scores s on $\mathcal{A}_{t,cal}$ for each coverage level in $\boldsymbol{\alpha}$.

    iv   Fit a distribution d using the quantile values computed in the previous step.

    v   Generate a set of candidate architectures from $\mathcal{A}$.

    vi   For each candidate architecture a, evaluate the acquisition function $\phi(a)$.

    vii   Denote a_{t+1} as the candidate architecture with maximum $\phi(a)$, and evaluate $f(a_{t+1})$.

    viii $\mathcal{A}_{t+1} \leftarrow \mathcal{A}_t \cup \{a_{t+1}\}$

4: **end for**
5: **Output:** $a^* = \operatorname{argmin}_{t=0,...,T} f(a_t)$

---

adopts an ensemble of five Feedforward neural networks (FNN). However, either approach is computationally expensive. Meanwhile, CP requires only training a single predictor in one search, which should benefit the entire search process especially under a given budget.

- **Model-agnostic:** CP is a general framework and can be applied to any dataset and any black-box model, which grants high flexibility in making engineering decision on other components.

3 Goals and Work Plan

The entire thesis writing period of 6 months can be roughly seen as a two-stage experiments. The first stage is to setup preliminary experiments and should last for approximately 2 weeks. The second stage is to enhance the results that are acquired in the previous stage by exploring variations of the Conformal Prediction algorithms and extending to alternative search spaces and datasets. The second experiment stage will last for approximately 4 months. Report drafting should take place over the entire course of the thesis writing period and the remaining one month is for final consolidation. The detailed experiment plan and timeline are outlined as follows:

- **Experiment Phase 1: Design and Initial Setup (2 weeks)** Upon the submission of this proposal, preliminary literature review on uncertainty quantification techniques and CP algorithms is expected to be completed. Therefore, the focuses of this stage are:

  - Design the experiment framework and conduct baseline experiments. Algorithm 1 depicts how CP can be incorporated into the BANANAS framework. The baseline

experiments begins with NAS-Bench-201 as the search space and is evaluated on CIFAR-10.

- Design ablation study, e.g., evaluating if the performance of CP-based NAS algorithm is sensitive to different parameterization, such as the size of set of neural architectures to be initially evaluated.

- **Experiment Phase 2: Enhancement (4 months)** The focuses of this phase are on the enhancement of the performance and robustness of the approach.

  - Incorporate different CP algorithms into the existing experiment framework and evaluate NAS performance. Candidates include applying Conformal Quantile Prediction [4], Boosted CP [6], weighting CP scores based on architecture similarity, or combining CP with Cross-validation and/or Jackknife+ [3].

  - Extend to different search spaces and evaluation datasets. For example, one direction is multi-objective NAS and evaluate the impact of uncertainty estimation on the robust accuracy of the best-found architecture [2].

References

[1] Shachi Deshpande, Charles Marx, and Volodymyr Kuleshov. Online calibrated and conformal prediction improves bayesian optimization. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 238, pages 7262–7273. PMLR, 2024.

[2] Steffen Jung, Jovita Lukasik, and Margret Keuper. Neural architecture design and robustness: A dataset. In *ICLR*, 2023.

[3] Jos M. Romano, Sol Minsker, Hampus Sill, and Maximilian Kasy. Classification with valid and adaptive coverage. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 8064–8074, 2020.

[4] Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32, pages 3543–3553, 2019.

[5] Colin White, Willie Neiswanger, and Yash Savani. Bananas: Bayesian optimization with neural architectures for neural architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[6] Ran Xie, Rina Foygel Barber, and Emmanuel J. Cands. Boosted conformal prediction intervals, 2024.