Master Thesis

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

Cheng Chen
(matriculation number 1662473)

July 31, 2025

Submitted to

Data and Web Science Group
Prof. Dr. Margret Keuper
University of Mannheim

Abstract

Some contents

Contents

List of Algorithms

List of Figures

List of Tables

1. Introduction

1.1. Motivation

1.2. Related Work

1.3. Contributions and Limitations

1.4. Outline

2. Literature Review

2.1. Neural Architecture Search (NAS)

2.1.1. Background

NAS is a subfield of AutoML.

Search Space Details will be introduced in 3

Search Strategy

Performance Evaluation

2.1.2. BANANAS

BANANAS is an Bayesian Optimization based search strategy. Give an introduction how Bayesian Optimization works. We list the five engineering decisions and review each field's related works. Maybe briefly cite Gaussian Process.

Architecture Encoding

Neural Predictor

Uncertainty Calibration

Acquisition Function

Acquisition Optimization

2.2. Uncertainty Quantification

Understanding uncertainty is important for real-world application of artificial intelligence, e.g., in autonomous driving, medical diagnosis.

2.2.1. Types of Uncertainty

- aleatory uncertainty (data uncertainty): uncertainty that arises due to inherent variations and randomness, and cannot be reduced by collecting more information
- epistemic uncertainty (model uncertainty): uncertainty that arises due to lack of knowledge, and can be reduced by collecting more information.

2.2.2. Alternative Uncertainty Estimation Methods

- Bayesian-based: e.g., Bayesian Neural Network
- Ensemble-based: e.g., Monte-Carlo dropout
- Bootstrapping

But these techniques are limited in several perspectives. First, quantifying uncertainty requires training models for several times, which means that the models cannot be applied for real-time prediction or in an online-learning setup. Second, some models are pre-trained and are only accessible via API. Besides, models (pre-)trained on certain datasets may struggle to generalize across different domains or contexts.

2.3. Conformal Prediction

2.3.1. Theoretical Background

Starting from i.i.d data, and provide an intuitive demonstration how the prediction interval is constructed (can add a figure illustrating why conformal prediction works, i.e., symmetry). From the most intuitive expression to the finite-sample adjusted expression.

Terminology Then, relax the i.i.d assumption to exchangeability, and lay a formal definition of the conformal prediction. And list the most importance three ingredients of the conformal predictions.

- A trained predictor f
- A conformity score function s. The conformity score is an important engineering decision and has an impact on the size on the prediction set, i.e., the efficiency. The conformity score function can be either a negatively- or positively oriented, in which ... And it can be a random variable as well.
- A target coverage alpha

Marginal coverage is guaranteed regardless of the choices in dataset and black box model. Only the model predictions are required to apply the technique.

A Link to Statistical Testing (clarify the relationship between conformal prediction and hypothesis testing) In this video (22:21), it is explained the intuition why conformal prediction guarantees the coverage, which is quite similar to the spirit of hypothesis testing.

The coverage parameters which should be pre-set plays a similar role as the confidence interval in hypothesis testing. Conformal prediction is like hypothesis testing with hypotheses:

H0: test instance i conforms to the training instances.

H1: test instance i does not conform to the training instances.

Table 2.1.: Comparison between CP and Hypothesis Testing

CP	**Hypothesis Testing**
(desired) Coverage level	Confidence level
Nominal error level (1 - Coverage level)	Significance level
The conformity score of the new instance	p-value (is an empirical term)

2.3.2. Online Conformal Prediction

2.3.3. Split Conformal Prediction

2.3.4. Limitations and Extensions

Limitations of split conformal predictions: - Distribution shift. The conformal prediction is built on the core assumption of exchangeability, which means the data points are identically distributed. However, this assumption is hard to meet in real-world application. For example, with time-series data this assumption is generally violated due to the temporal relationships. - Adaptivity. Once the conformity scores are computed on the calibration set, the decision threshold is settled and is applied to all test datapoints, regardless of the intrinsic complexity of the exact example. It is desirable that the threshold can adapts to the difficulty of the problem and produce a larger prediction interval/set on hard-to-solve example and smaller prediction interval/set on easy-to-solve example. This limitation echoes with the characteristic of Conformal Prediction that the guaranteed coverage is only marginal over all datapoints but not conditional on a specific data points..

Variations of Conformal predictions have been proposed to overcome the limitations. There are three main streams: - find an empirical coverage rate which leads to the desired coverage level. For example, if the desired coverage rate is 90- find an efficient conformity score: Alternatively, [. . .] apply the conformal prediction in an online setting to dynamically incorporate the conformity score of new data points. - find suitable predictor: The trained predictor can be just a poor approximation of the real data generation process. Cross-validation / Jackknife +, Conformal Quantile Prediction

Besides, [. . .] proposes a CP algorithm that samples datapoints using Monte-Carlo sampling to approach the real distribution of labels in case the ground-truth is ambiguous and consequently cause a biased distribution in manually-annotated labels.

3. Datasets Description

This thesis will primarily focus on NAS-Bench-201 due to comparison concerns.

3.1. NAS-Bench-201

3.2. Robust dataset

3.3. Others

NAS-Benchmark-101 / 301 and DARST

4. Methodology

4.1. Baseline

BANANAS + Split Conformal Prediction

4.2. Enhancement

4.2.1. Conformity scores

4.2.2. Conformal Quantile Regression

4.2.3. Conformal Prediction with Cross-validation / Jackknife+

4.2.4. Conformal Prediction with Boosting

4.2.5. ...

....

4.3. Evaluation Metrics

5. Experiments and Results

6. Conclusion

Summary + Future Work

A. Additional Experimental Results

Ehrenwörtliche Erklärung

Ich versichere, dass ich die beiliegende Bachelor-, Master-, Seminar-, oder Projektarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und in der untenstehenden Tabelle angegebenen Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Declaration of Used AI Tools

Tool	Purpose	Where?	Useful?
ChatGPT	Rephrasing	Throughout	+

Unterschrift
Mannheim, den 31.07.2025