Review: Improving UQ of Deep Classifiers via Neighborhood Conformal Prediction

Subhankar Gosh et. al.

Reviewer: Jihu Lee

IDEA lab
Department of Statistics
Seoul National University

February 26, 2025

Outline

Conformal Prediction

- Prediction step) trained model $\rightarrow$ conformity scores
- Calibration step) find threshold to construct prediction set using conformity scores

UQ from CP

- *coverage*: the probability that the true output is contained in prediction set
- *efficiency*: size of the prediction set (smaller is better)
- Tradeoff: *coverage* $\leftrightarrow$ *efficiency*

Main Question

- How can we improce CP to achieve provably higher efficiency by satisfying the marginal coverage constraint for pre-trained deep classifiers?
- Propose an algorithm named Neighborhood Conformal Prediction

Notations

- $x \in \mathcal{X}$: input
- $y^* \in \mathcal{Y}$: true output
- $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$
- F_θ: neural network with parameters θ
- $\Phi(X)$: representation of X from F_θ
- $\mathcal{C}(x)$: prediction set for input x
- $V(x, y)$: non-conformity score
- $\mathcal{B}(x) \triangleq \{x' \in \mathcal{X} : \|x - x'\| \leq B\}$: neighborhood size of B

Split Conformal Prediction (CP)

$$\hat{Q}^{\mathsf{CP}}(\alpha, V_{1:n}) = \min\left\{t : \sum_{i=1}^{n} \frac{1}{n}\mathbb{1}[V_i \le t] \ge 1-\alpha\right\} \tag{1}$$

Algorithm 1: Split Conformal Prediction (CP)

1: **Input**: Significance level $\alpha \in (0,1)$; Randomly split data into training set $\mathcal{D}_{\text{tr}}$ and calibration set $\mathcal{D}_{\text{cal}} = \{Z_1, \cdots, Z_n\}$.
2: If predictor F_θ is not given, train a prediction model F_θ on the training set $\mathcal{D}_{\text{tr}}$
3: Compute non-conformity score V_i for each example $Z_i \in \mathcal{D}_{\text{cal}}$
4: Compute $\hat{Q}^{\text{CP}}(\alpha, V_{1:n})$ as the $\lceil (1-\alpha)(1+|\mathcal{D}_{cal}|) \rceil$th smallest value in $\{V_i\}_{i \in \mathcal{D}_{\text{cal}}}$ as in (1).
5: $\hat{\mathcal{C}}(x_{n+1}) = \{y : V(x_{n+1}, y) \le \hat{Q}^{\text{CP}}(\alpha, V_{1:n})\}$ is the prediction set for a testing input x_{n+1}

Neighborhood Conformal Prediction (NCP)

NCP quantile

$$\alpha^{\mathsf{NCP}}(\alpha) = \max\left\{\tilde{\alpha} : \sum_{i=1}^{n} \frac{1}{n} \mathbb{1}[V_i \le \hat{Q}^{\mathsf{NCP}}(\tilde{\alpha}; V_{1:n}; p_{i,1:n})] \ge 1 - \alpha\right\} \tag{2}$$

$$\hat{Q}^{\mathsf{NCP}}(\tilde{\alpha}, V_{1:n}; p_{i,1:n}) = \min\left\{t : \sum_{j=1}^{n} p_{i,j} \mathbb{1}[V_j \le t] \ge 1 - \tilde{\alpha}\right\}$$

Basic case

$$p_{i,j} = \frac{\mathbb{1}\left[\Phi(X_j) \in \mathcal{B}(\Phi(X_i))\right]}{\sum_{k \in \mathcal{D}_{cal}} \mathbb{1}\left[\Phi(X_k) \in \mathcal{B}(\Phi(X_i))\right]} \tag{3}$$

Special case

$$IW(x_i, x_j) = \exp\left(-\frac{dist(\Phi(x_i), \Phi(x_j))}{\lambda_L}\right) \tag{4}$$

$$p_{i,j}^{exp} = \frac{IW(x_i, x_j)}{\sum_{k \in \mathcal{D}_{cal(KNN)}} IW(x_i, x_k)} \tag{5}$$

Neighborhood Conformal Prediction (NCP)

Algorithm 2: Neighborhood Conformal Prediction (NCP)

1: **Input**: Significance level $\alpha \in (0, 1)$. Randomly split data into training set $\mathcal{D}_{\text{tr}}$ and calibration set $\mathcal{D}_{\text{cal}} = \{Z_1, \cdots, Z_n\}$.

2: If predictor F_θ is not given, train a prediction model F_θ on the training set $\mathcal{D}_{\text{tr}}$.

3: Compute non-conformity score V_i for $Z \in \mathcal{D}_{\text{cal}}$

4: Compute importance weights $p_{i,j}$ for $X_i, X_j \in \mathcal{D}_{\text{cal}}$ needed to identify *neighborhood* according to (3 or 5)

5: Find α^{NCP} in (2) on $\mathcal{D}_{\text{cal}}$ and set $\hat{\mathcal{C}}(X_{n+1}) = \{y : V(X_{n+1}, y) \leq \hat{Q}(\alpha^{\text{NCP}}, V_{1:n+1}, p_{n+1,1:n+1})\}$ as the prediction set for the new data sample X_{n+1}

Theory

Theorem 1

Let $Q^{CP}(\alpha) \triangleq \min\{t : \mathbb{P}_X\{V(X, F^*(X)) \leq t\} \geq 1-\alpha\}$,
$\alpha^{NCP}(\alpha) \triangleq \max\{\tilde{\alpha} : \mathbb{P}_X\{X \leq Q^{NCP}(\tilde{\alpha}; X)\} \geq 1-\alpha\}$, *where*
$Q^{NCP}(\tilde{\alpha}; X) \triangleq \min\{t : \mathbb{P}_{X'}\{V(X'; F^*(X)) \leq t, X' \in \mathcal{N}_{\mathcal{B}}(X)\} \geq 1-\tilde{\alpha}\}$
in population.
Then to achieve the same α, `NCP` *gives smaller expected quantile and can be more efficient than* `CP`*, under some conditions:*

$$\mathbb{E}_X[Q^{NCP}(\alpha; X)] \leq Q^{CP}(\alpha)$$

Experimental Settings

Regression

- Dataset: BlogFeedback, Gas turbine, Facebook_i, MEPS_i
- Model: MLP with 3 hidden layers with 15, 20, 30 nodes
- Non-conformity score: $V(x, y^*) = |y^* - F_\theta(x)|$

Classification

- Dataset: CIFAR10, CIFAR100, ImageNet
- Model: ResNet architectures
- Conformity score: `APS`, `RAPS`

Main Results

Model	Naive	APS	RAPS	NCP(APS)	NCP(RAPS)
ImageNet($1-\alpha=0.900$)					
ResNet152	10.55(0.548)	11.34(0.843)	2.53(0.054)	5.24(0.312)	**2.12(0.007)**
ResNet101	10.85(0.588)	11.36(0.615)	2.63(0.08)	5.60(0.278)	**2.25(0.098)**
ResNet50	12.00(0.530)	13.24(0.892)	2.94(0.089)	6.44(0.381)	**2.54(0.091)**
ResNet18	16.13(0.642)	17.05(1.100)	5.00(0.220)	11.08(0.598)	**4.71(0.251)**
ResNeXt101	18.57(1.05)	20.57(1.10)	2.36(0.069)	8.24(0.388)	**2.01(0.060)**
DenseNet161	11.70(0.730)	12.86(1.21)	2.67(0.074)	5.89(0.447)	**2.29(0.103)**
VGG16	13.91(0.867)	14.10(0.486)	3.97(0.098)	8.56(0.382)	**3.62(0.106)**
Inception	77.51(3.78)	90.28(4.16)	5.96(0.373)	66.29(2.83)	**5.67(0.178)**
ShuffleNet	30.33(1.64)	34.31(2.10)	5.53(0.070)	22.36(1.37)	**5.49(0.131)**
CIFAR100 ($1-\alpha=0.900$)					
ResNet18	5.07(0.327)	5.57(0.224)	3.17(0.076)	3.61(0.164)	**2.77(0.100)**
ResNet50	8.21(0.746)	8.02(0.405)	3.37(0.189)	4.26(0.233)	**2.96(0.213)**
ResNet101	4.44(0.348)	4.64(0.184)	2.60(0.066)	2.80(0.188)	**2.19(0.073)**
CIFAR10($1-\alpha=0.960$)					
ResNet18	1.20(0.016)	1.25(0.021)	1.24(0.014)	1.07(0.013)	**1.06(0.012)**
ResNet50	1.15(0.013)	1.19(0.021)	1.17(0.011)	1.05(0.010)	**1.04(0.005)**
ResNet101	1.17(0.017)	1.20(0.018)	1.19(0.017)	1.04(0.008)	**1.03(0.007)**

(a) Classification

Dataset	CP	NCP
BlogFeedback	44.79(2.78)	18.97(2.42)
Gas turbine	16.69(1.47)	12.94(1.11)
Facebook_1	50.96(5.72)	21.89(7.27)
Facebook_2	44.12(3.85)	21.28(6.55)
Facebook_3	46.98(4.01)	39.73(9.59)
Facebook_4	46.78(5.35)	34.01(11.76)
Facebook_5	45.35(0.65)	43.73(5.55)
MEPS_19	54.99(2.64)	35.63(2.54)
MEPS_20	48.85(2.22)	33.85(3.53)
MEPS_21	61.46(7.49)	36.01(5.10)
Concrete	59.27(0.23)	50.42(0.62)

(b) Regression

Figure 1: Mean prediction set sizes, with the mean and s.d. over 5 (a) and 10 (b) runs.

References

[1] Ghosh, Subhankar, et al. "Improving uncertainty quantification of deep classifiers via neighborhood conformal prediction: Novel algorithm and theoretical analysis." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. No. 6. 2023.


THANK
you!