Conformalized Semi-supervised Random Forest for Classification and Abnormality Detection

THE UNIVERSITY OF HONG KONG 香 港 大 學
NORTHWESTERN UNIVERSITY
Yale University

Yujin Han[1] *, Mingwenchan Xu[2] *, Leying Guan[3]

AISTATS

Conformal Prediction

Definition

Based on the previous n observations, the conformal prediction creates a prediction set $\hat{C}_n(x_{n+1})$ for the new instance x_{n+1} and guarantee,

$$\mathbb{P}(y_{n+1} \in \hat{C}_n(x_{n+1})) \geq 1 - \alpha,$$

where $\alpha \in (0,1)$ is the allowed miscoverage level.

Assumption & Challenge

Classification uncertainty quantification assumes training and test samples are i.i.d.

However, in scenarios with data distribution shifts, like healthcare and network attacks, the above assumption fails.

Motivation

How to

- Assess uncertainty under distribution changes?
- Identify test samples (i.e., outliers) where predictions should not rely solely on the model trained with the training data?

Distribution Shift

For class k, its mixture proportion is π_k, and feature density is $f_k(x)$, with $\sum_{k=1}^{K} \pi_k = 1$. The test distribution $\mu(x)$ is

$$\mu(x) := \sum_{k=1}^{K} \tilde{\pi}_k f_k(x) + \delta \cdot f_R(x)$$

Traditional label shift: $\{\pi_k\} \neq \{\tilde{\pi}_k\}$; **Outliers**: $\delta \neq 0$

CSForest

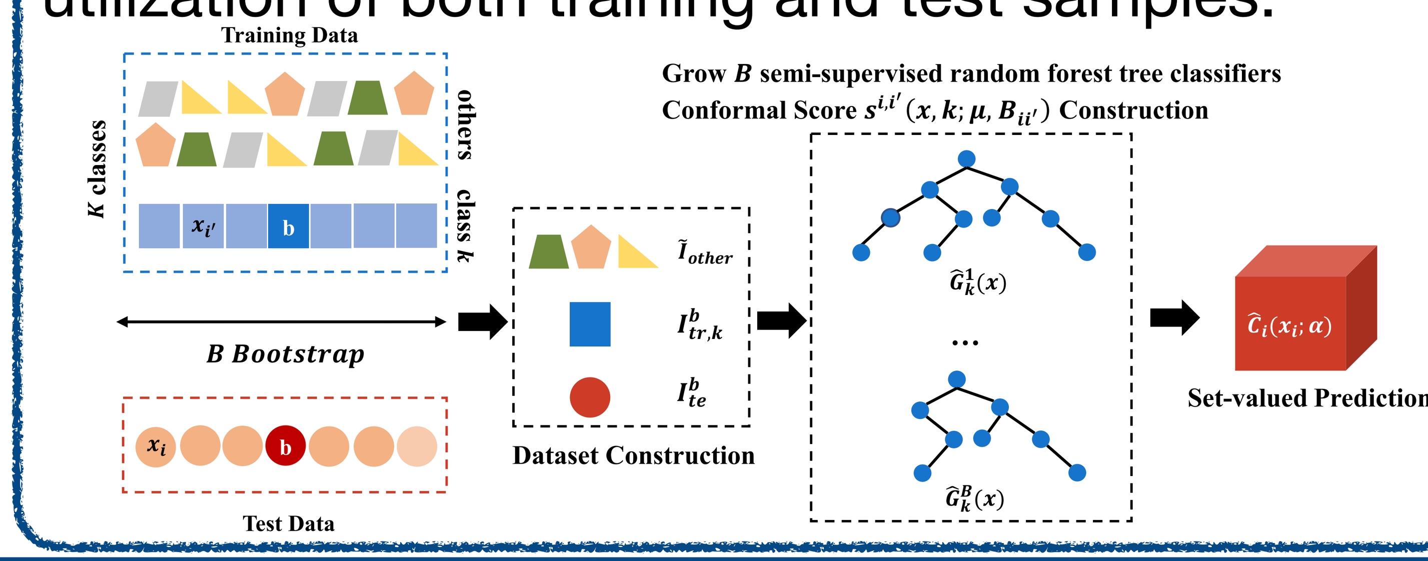
conformalized semi-supervised random forest

- Semi-supervised Random Forest Structure

Differentiates between observed training classes and unlabeled test data.

- with Jackknife+aB Technique

Handles the case of joint and asymmetric utilization of both training and test samples.



Method and Experiments

Algorithm Details

Input : Training Data $\{(x_i, y_i), i \in \mathcal{I}_{tr}\}$, Test Data $\{x_i, i \in \mathcal{I}_{te}\}$, $\tilde{B}$ and w (1 by default.)
Output : Prediction sets $\hat{C}_i(x_i)$ for $i \in \mathcal{I}_{te}$.

for $k = 1, \ldots, K$ **do**
  Sample B from Binomial$(\tilde{B}; (1 - \frac{1}{n_k+1})^{n_k})$. **for** $b = 1, \ldots, B$ **do**
    Let $\mathcal{I}_k^b$, $\mathcal{I}_{te}^b$ be the Bootstraps of $\mathcal{I}_k$ (index of training class k) and $\mathcal{I}_{te}$. Let $\tilde{\mathcal{I}}_{other}$ be the Bootstrap of size $\min(\lceil n_{te} w \rceil, n - n_k)$ from the remaining training sample indices $\mathcal{I} \setminus \mathcal{I}_k$.
    Grow a single random forest tree classifier $\hat{G}^b(x)$ separating different labeled classes and the test samples using $\mathcal{I}_k^b \cup \mathcal{I}_{te}^b \cup \tilde{\mathcal{I}}_{other}$.
  end
  For sample pair $i \in \mathcal{I}_{te}, i' \in \mathcal{I}_k$, set $\mathcal{B}_{ii'} = \{b : i \notin \mathcal{I}_{te}^b, i' \notin \mathcal{I}_k^b\}$ and construct the conformal score function $\hat{s}^{ii'}(x, k; \mu) = \left(\sum_{b \in \mathcal{B}_{ii'}} \hat{G}_k^b(x)\right) / |\mathcal{B}_{ii'}|$.
end
for $i \in \mathcal{I}_{te}$ **do**
  for $k = 1, \ldots, K$ **do**
    Construct $\hat{s}_{ik}$ for sample i and class k via eq. (3).
  end
  Construct $\hat{C}(x_i) = \{k : \hat{s}_{ik} \geq \alpha\}$.
end

Coverage Guarantee for True Labels

Suppose the generalized label shift model holds where features from class k are i.i.d generated from a distribution P_k. For any fixed integers $\tilde{B} \geq 1$, the constructed $\hat{C}_i(x)$ from CSForest satisfies:

$$P\left[k \in \hat{C}(x_i) \,|\, y_i = k\right] \geq 1 - 2\alpha,$$

$$\textit{for all } i \in \mathcal{I}_{te} \textit{ and } k = 1, \ldots, K.$$

Experimental Settings

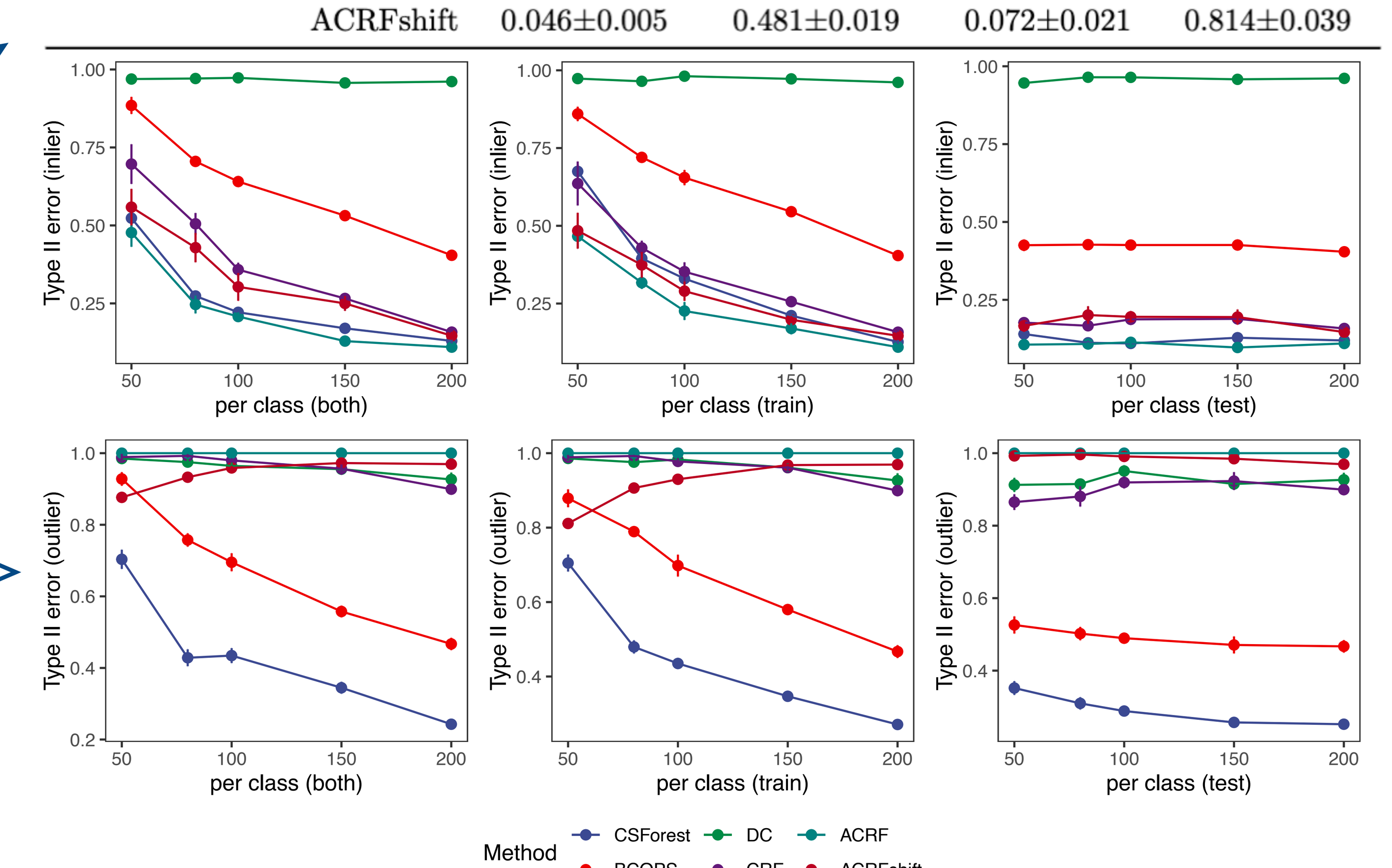
- **Q1 (outliers w/o shift):** Can CSForest efficiently identify outliers and accurately predict inliers without traditional label shift?
- **Q2 (shift w/o outliers):** With no outliers but traditional label shift, can CSForest match or outperform other classifiers?
- **Q3**: Does CSForest is stable as the training and test sample sizes vary?

Q1&Q2

Q3

Experimental Results

Dataset	Method	outliers w/o shift		shift w/o outliers	
		Type I Error	Type II Error	Type I Error	Type II Error
MNIST	CSForest	0.049±0.006	**0.091 ± 0.008**	0.048±0.016	**0.291±0.038**
	BCOPS	0.048±0.004	0.237±0.019	0.042±0.007	0.556±0.040
	DC	0.049±0.008	0.890±0.021	0.046±0.016	0.968±0.022
	CRF	0.048±0.007	0.338±0.035	0.046±0.018	0.428±0.082
	ACRF	0.046±0.006	0.430±0.003	0.025±0.011	0.884±0.012
	ACRFshift	0.046±0.006	0.432±0.009	0.055±0.013	0.828±0.015
CIFAR-10	CSForest	0.051±0.008	**0.000±0.000**	0.049±0.013	0.009±0.035
	BCOPS	0.049±0.006	0.001±0.000	0.042±0.009	0.029±0.006
	DC	0.046±0.007	0.048±0.091	0.039±0.010	0.071±0.115
	CRF	0.049±0.008	0.003±0.000	0.047±0.015	**0.000±0.000**
	ACRF	0.003±0.001	0.402±0.001	0.040±0.009	0.221±0.023
	ACRFshift	0.003±0.001	0.069±0.003	0.046±0.007	0.230±0.035
FashionMNIST	CSForest	0.050±0.005	**0.266±0.018**	0.038±0.009	**0.311±0.040**
	BCOPS	0.050±0.007	0.381±0.020	0.038±0.009	**0.311±0.040**
	DC	0.051±0.007	0.666±0.033	0.038±0.013	0.584±0.066
	CRF	0.051±0.006	0.514±0.021	0.038±0.014	0.804±0.080
	ACRF	0.051±0.006	0.537±0.013	0.054±0.009	0.835±0.020
	ACRFshift	0.046±0.005	0.481±0.019	0.072±0.021	0.814±0.039



Conclusions

- Q1&Q2: CSForest shows the **strongest capability to detect outliers** (smaller type II errors) in both scenarios.
- Q3: CSForest **detects outliers** while maintaining **lower inlier type II errors** across **various sample sizes.**

arXiv