Derandomized novelty detection with FDR control via conformal e-values

Kyungseon Lee

February 25, 2025

Seoul National University

Outline

1. Introduction
2. Experiment
3. Conclusion

Introduction

Derandomized novelty detection with FDR control via conformal e-values

- Novelty detection = Anomaly detection = Outlier detection
- Conformal inference = Conformal prediction
- Conformity score = Anomaly score

Derandomized novelty detection with FDR control via conformal e-values

1. Problem - The results of conformal inference change because it has randomness in train-calibration set splitting.
   - Solution : A weighted sum of p-values.

2. Problem - The traditional method has difficulty controlling the False Discovery Rate (FDR).
   - Solution : Instead of the traditional p-value that depends on the rank of the calibration set, a conformal e-value is used.

Traditional conformal anomaly detection

H_{0k} : The point x_k of test set is also an inlier

- If $p_k < \alpha$, then H_{0k} is rejected. So, x_k is outlier.

Conformal p-value

$$p_i = \frac{1 + \sum_{j \in \mathcal{C}} 1\left(s_j \geq s_i\right)}{|\mathcal{C}| + 1}$$

- $\mathcal{C}$: Calibration set
- s_j and s_i : The conformity scores for callibration and test points
- $f(\cdot)$: A conformity score function trained so that inlier data has a label of 0 and outlier data has a label of 1, $s_i = f(x_i)$

Main contributions

A weighted sum of conformal e-values

$$\bar{e}_j = \sum_{k=1}^{K} w^{(k)} e_j^{(k)}, \quad \sum_{k=1}^{K} w^{(k)} = 1$$

- $e_j^{(k)}$: The k-th conformal e-value
- $w^{(k)}$: The k-th weight
- k : The k-th split repetition

Main contributions

Conformal e-value

$$e_j^{(k)} = (1 + n_{\mathsf{cal}}) \cdot \frac{\mathbb{I}\left\{\hat{S}_j^{(k)} \geq \hat{t}^{(k)}\right\}}{1 + \sum_{i \in \mathcal{D}_{\mathrm{cal}}^{(k)}} \mathbb{I}\left\{\hat{S}_i^{(k)} \geq \hat{t}^{(k)}\right\}}$$

$$\hat{t}^{(k)} = \min\left\{t \in \tilde{\mathcal{D}}_{\textsf{cal-test}}^{(k)} : \widehat{\mathrm{FDP}}^{(k)}(t) \leq \alpha\right\} \text{ where } \tilde{\mathcal{D}}_{\textsf{cal-test}}^{(k)} = \left\{\hat{S}_i^{(k)}\right\}_{i \in \mathcal{D}_{\textbf{test}} \cup \mathcal{D}_{\textbf{cal}}^{(k)}}$$

$$\widehat{\mathrm{FDP}}^{(k)}(t) = \frac{n_{\mathsf{test}}}{n_{\mathsf{cal}}} \cdot \frac{\sum_{i \in \mathcal{D}_{\textbf{cal}}^{(k)}} \mathbb{I}\left\{\hat{S}_i^{(k)} \geq t\right\}}{\sum_{j \in \mathcal{D}_{\textbf{test}}} \mathbb{I}\left\{\hat{S}_j^{(k)} \geq t\right\}}$$

- $\mathcal{D} = \{1, \ldots, n\}$: The index set of the observed samples.
- k : The k-th split repetition

Main contributions

Conformal e-value - FDP : The empirical FDR

$$\mathrm{FDR} = \mathbb{E}\left[\left(\sum_{j \in \mathcal{D}_{\text{test}}^{\text{null}}} R_j\right) / \max\left\{1, \sum_{j \in \mathcal{D}_{\text{test}}} R_j\right\}\right]$$

$$\widehat{\mathrm{FDP}}(t) = \frac{n_{\text{test}}}{n_{\text{cal}}} \cdot \frac{\sum_{i \in \mathcal{D}_{\text{cal}}} \mathbb{I}\left\{\hat{S}_i \geq t\right\}}{\sum_{j \in \mathcal{D}_{\text{test}}} \mathbb{I}\left\{\hat{S}_j \geq t\right\}}$$

- $\mathcal{D} = \{1, \ldots, n\}$: The index set of the observed samples.
- $\mathcal{D}_{\text{test}}^{\text{null}}$: The set of true inlier test points.
- R_j : Whether it was predicted as an outlier.
- k : The k-th split repetition.

Main contributions

Conformal e-value

$$e_{(1)} \geq e_{(2)} \geq \cdots \geq e_{(n)}$$

$$x_{(i)} \text{ is outlier when } \bar{e}_{(i)} \geq \frac{1}{\alpha} \cdot \frac{i}{n_{\text{test}}}$$

- $\bar{e}_{(i)}$: The i-th value in the weighted sum of e-values sorted in descending order.
- n_{test} : The number of points in the test set.
- $\mathcal{D} = \{1, \ldots, n\}$: The index set of the observed samples.
- α : The significance level.

Main contributions

Theorem

If the inliers in $\mathcal{D}$ and the null test points are exchangeable conditional on the non-null test points, then weighted e-value method guarantees FDR $\leq \alpha$.

- $\mathcal{D} = \{1, \ldots, n\}$: The index set of the observed samples.
- α : The significance level.

Experiment

Experiment

- Baseline : AdaDetect
- Our method : E-AdaDetect
- $X_{inlier} \sim \mathcal{N}(0, I_{100})$
- $X_{outlier} \sim \mathcal{N}(\mu, I_{100})$

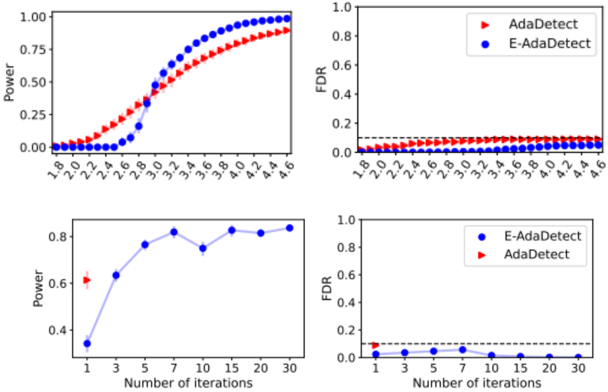

Figure 1: The two graphs above show how power and FDR change with variations in μ. The two graphs below show how power and FDR change with the number of train-calibration split repetitions.

Conclusion

Conclusion

- While e-values are less powerful than p-values for single tests, they allow FDR control in multiple testing and efficiently combine non-independent tests.
- This study remove randomness in split-conformal inference and AdaDetect.